Prediction models in growth research using advanced statistical methods

Maria de Ridder

Prediction models in growth research using advanced statistical methods

Predictiemodellen in groeionderzoek met geavanceerde statistische methoden

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus prof.dr. S.W.J. Lamberts

en volgens besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op
woensdag 16 januari 2008 om 15.45 uur

door

Maria Antonia Johanna de Ridder

geboren te Gouda

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotoren: Prof.dr. A.C.S. Hokken-Koelega
                 Prof.dr. Th. Stijnen

Overige leden: Prof.dr. E.W. Steyerberg
                 Prof.dr. C.G.M. Moons
                 Prof.dr. M. Ranke

Table of contents

Manuscripts based on studies described in this thesis

Chapter 2
de Ridder MAJ, Stijnen Th, Hokken-Koelega ACS.
Validation and calibration of the Kabi Pharmacia International Growth Study prediction model for children with idiopathic growth hormone deficiency.
J Clin Endocrinol Metab. 2003 Mar;88(3):1223-7.

Chapter 3
de Ridder MAJ, Stijnen Th, Drop SLS, Blum WF, Hokken-Koelega ACS.
Validation of a calibrated prediction model for response to growth hormone treatment in an independent cohort.
Horm Res. 2006;66(1):13-6.

Chapter 4
de Ridder MAJ, Stijnen Th, Hokken-Koelega ACS.
Prediction of adult height in growth-hormone-treated children with growth hormone deficiency.
J Clin Endocrinol Metab. 2007 Mar;92(3):925-31.

Chapter 5
de Ridder MAJ, Stijnen Th, Hokken-Koelega ACS.
Prediction model for adult height of SGA children at start of growth hormone treatment.
J Clin Endocrinol Metab. 2008, *in press*.

Chapter 6
de Ridder MAJ, Stijnen Th, Hokken-Koelega ACS.
SGA children without early catch-up growth: spontaneous growth and prediction of height at 8 years.
Horm Res. 2007, *in press*.

Chapter 7
de Ridder MAJ, Stijnen Th, Hokken-Koelega ACS.
A new method to determine mean adult height from incomplete follow-up data.
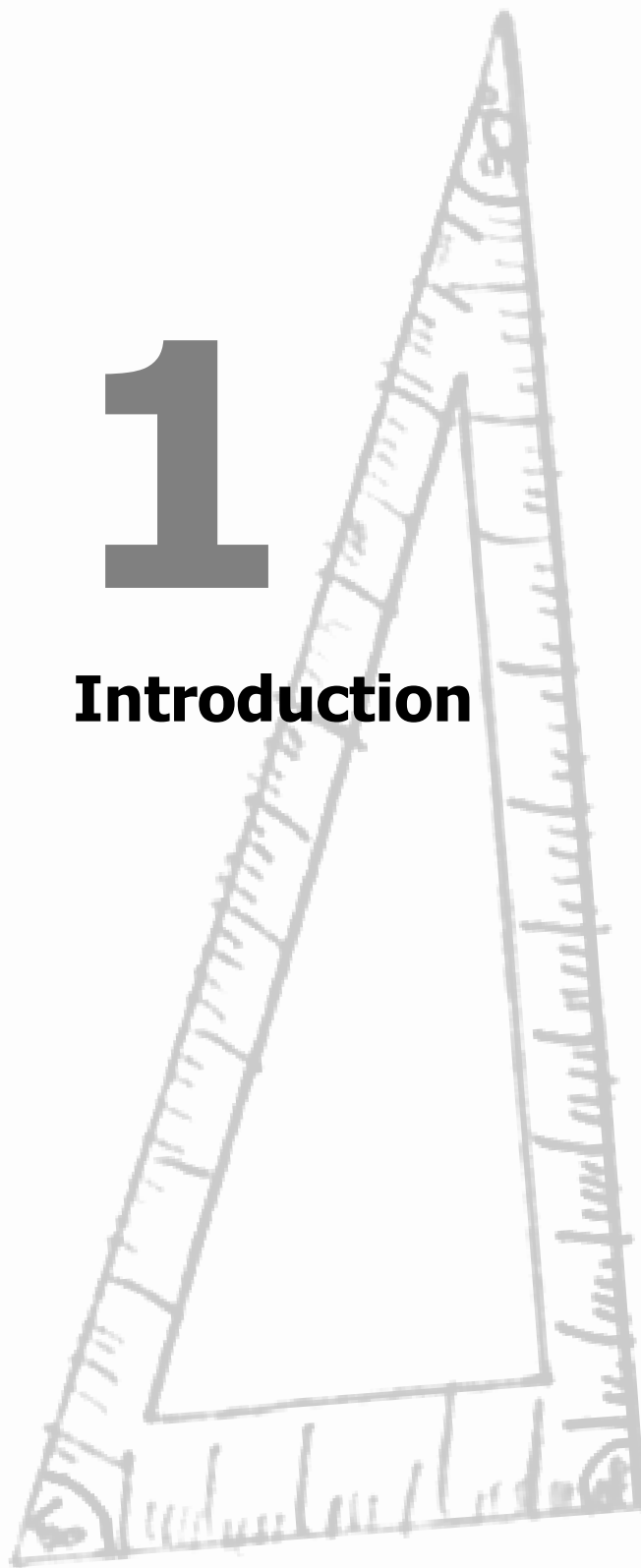Horm Res. 2007;67(4):205-10.

Chapter 8
de Ridder MAJ, Verburg BO, Steegers EAP, Mackenbach JP, Moll HA, Hofman A, Witteman JCM, Jaddoe VWV.
Individually customised fetal weight charts derived from ultrasound measurements in a population-based cohort. The Generation R Study.
Submitted.

1

Introduction

In this chapter the background will be presented of the topics of this thesis. First, some aspects of growth in general are described. Secondly, the current state of art with regard to prediction modelling for growth analyses is given. Finally an outline of this thesis is presented.

Growth

Prenatal growth

Human growth is most intense prenatally. In 40 weeks, the fetus gains on average 3200 grams of weight and 50 cm in length. Fetal growth is the result of very complex metabolic and endocrine processes. The insulin-like growth factors (IGF's) I and II and insulin are major determinants of fetal growth (1). The influence of growth hormone (GH) on prenatal growth is very moderate, particularly due to low sensitivity of the GH receptors (2, 3). Epidemiological and clinical studies reported the following determinants of fetal growth: height of parents, ethnicity, age of mother, gender of the fetus but also environmental factors such as social economic status of the parents, condition of the mother, smoking during pregnancy (4, 5).

During the third trimester of gestation, the fetus grows very fast, with a maximum growth velocity of 100 cm/year (Figure 1). At the end of gestation, height velocity decreases to approximately 25 cm/year.

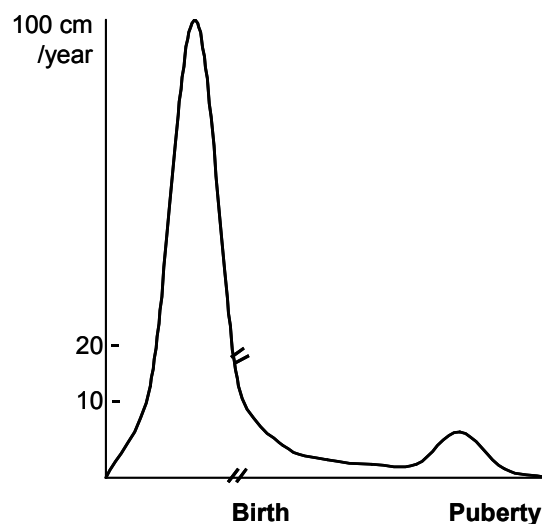
Figure 1. Pre- and postnatal growth; adapted from Widdowson

Postnatal growth

Postnatal longitudinal growth is also a complex biological process resulting from multiple interactions between endogenous factors, such as genetics, hormones,

metabolic processes, target tissue responsiveness, and exogenous factors, such as nutrition, physical activity, and psychosocial influences.

Normal growth requires the cooperation of several hormones, amongst which GH, thyroid hormones, sex steroids and corticosteroids. While fetal growth and early postnatal growth until the age of 3-6 months is mostly GH independent, there is a progressive increment in GH dependency after the first months when GH becomes the most important hormone in controlling longitudinal growth (6, 7).

GH is secreted in a pulsatile pattern by the anterior pituitary gland (8). It stimulates various processes in the body (Figure 2). GH has a stimulatory effect on the production of IGF-I, which is considered as one of the main growth factors and is involved in a large number of cellular processes, such as cell proliferation, metabolism, differentiation, motility and migration and cell survival (9).
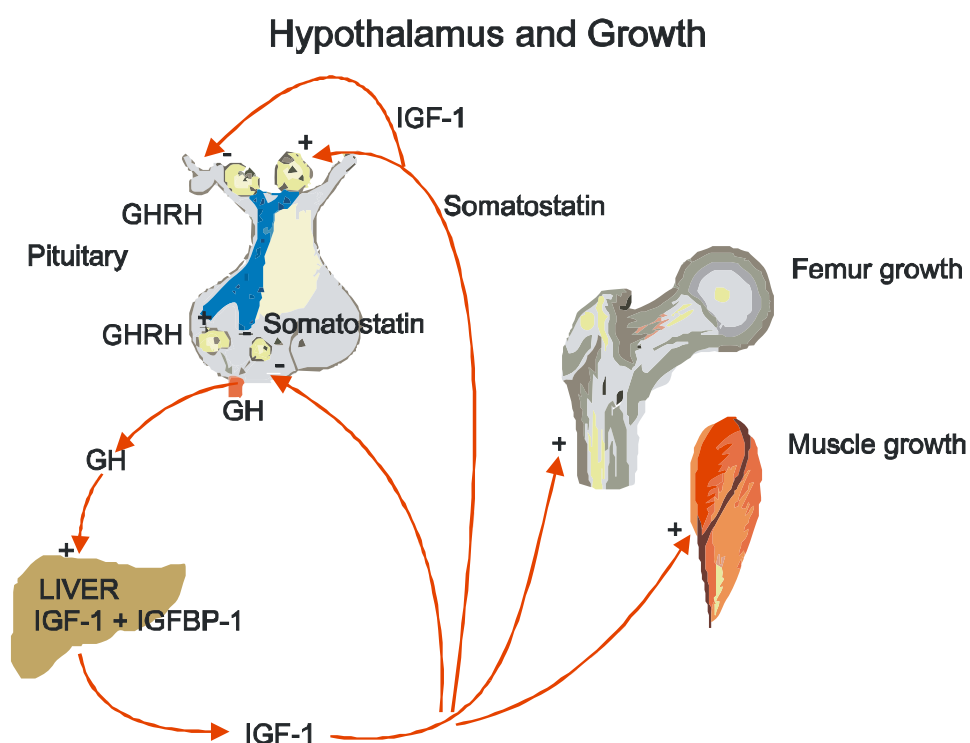
Figure 2. The GH/IGF-I axis

The growth velocity decreases during the first two years of life to 7 cm/year, and then slowly to less than 5 cm/year at start of puberty (Figure 1). During puberty, starting in girls at a mean age of 10.7 years and in boys at 11.5 years (10) there is a growth spurt with mean peak height velocity 8.3 cm/year in boys and 7.0 cm/year in girls (11). The total pubertal height gain is 29 cm in boys and 24 cm in girls (11).

A child's height is usually evaluated relative to that of the reference population. A height measurement is expressed in standard deviation score (SDS), i.e. the

difference between the observed height and the mean height of children of the same age and sex in the reference population, divided by the standard deviation (Figure 3).
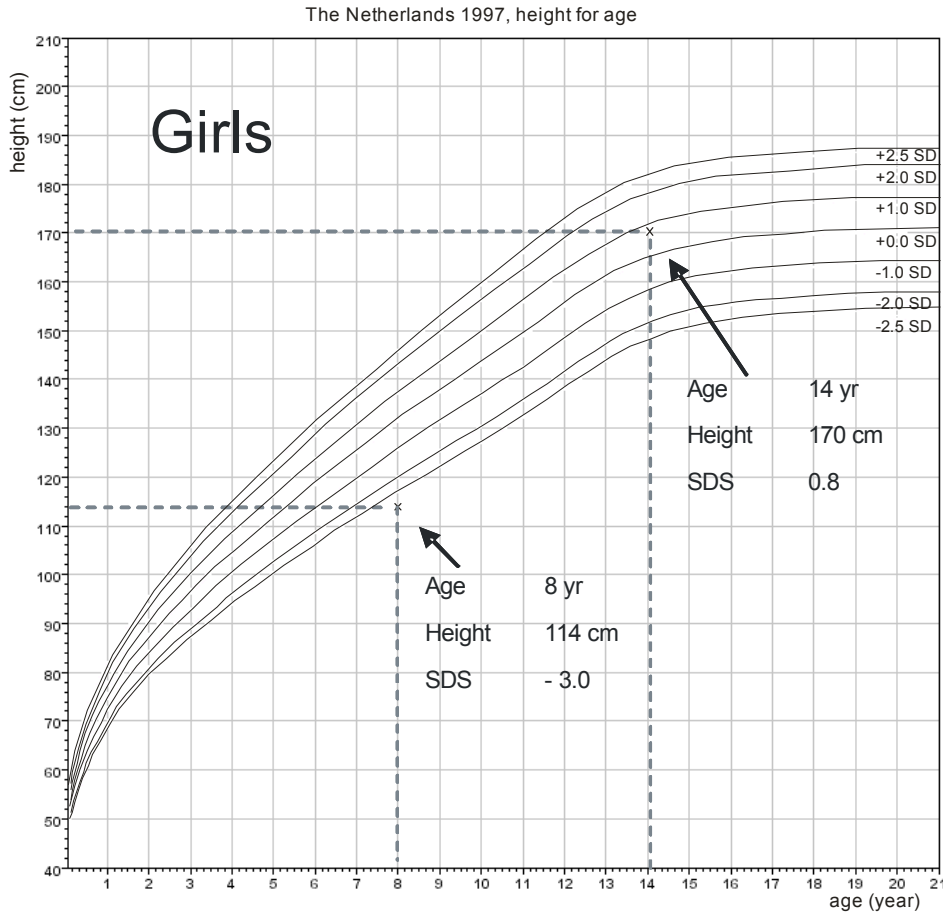
Figure 3. Reference chart for Dutch girls, with examples of height observations

The endpoint of postnatal growth is adult height. In healthy persons, important determinants of adult height are gender (Dutch male adults are on average 13.4 cm taller than females(10)) and parental height, explaining 22 – 25% of the variance in adult height (12-14). In almost all industrialized countries, the average adult height showed an important increase during the last century, caused by improvement of nutritional, hygienic and health status (15-17). Between 1955 and 1997, this increase in adult height (called secular trend) was in the Netherlands 8.0 cm for males and 7.75 cm for females (10) (Figure 4).
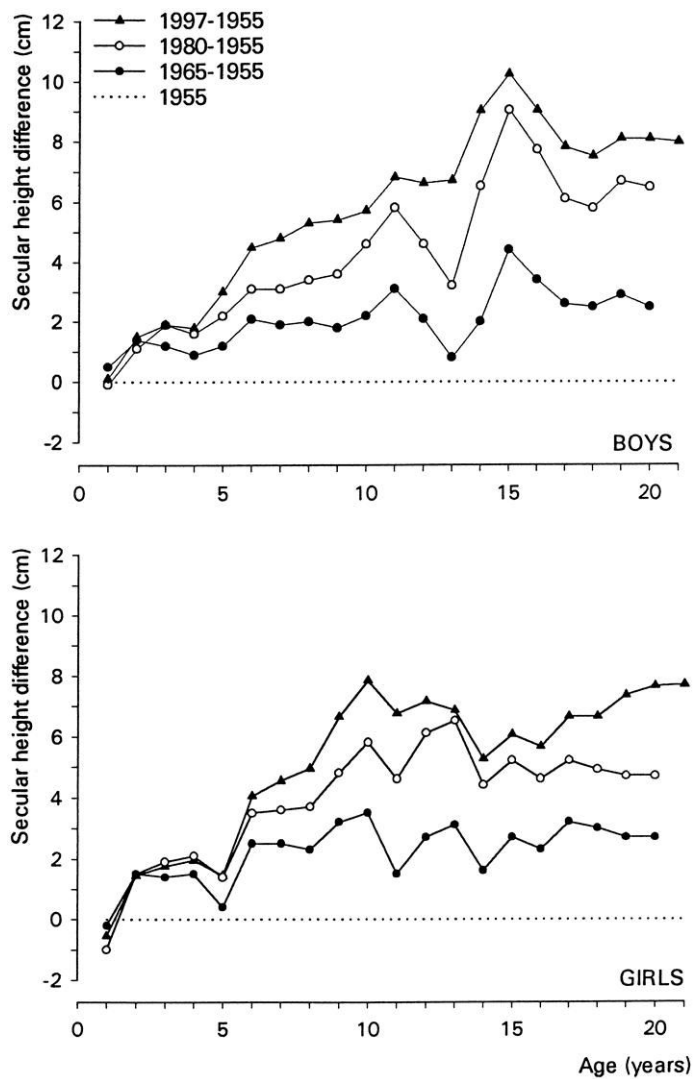
Figure 4. Secular differences in mean height between the 1997, 1980, and 1965 growth studies in comparison to data from the 1955 growth study. From: FREDRIKS: Pediatr Res, Volume 47(3).March 2000.316-323

Growth hormone deficiency

Growth retardation in children might be due to a subnormal level of growth hormone (GH). Estimates of the prevalence of children with growth hormone deficiency (GHD) cover a broad range from 18 to 287 per million (18), depending on the definition used. The prevalence in the Netherlands is estimated between 500 per million (19) and 200 per million [data of the Dutch Growth Foundation: treated children with GH max < 20 mU/l and IGF-I below the mean for age and sex, compared to the number of life born infants per year (20)]. GH secretion is a continuum between normality and abnormality (21). The diagnosis of severe GHD is usually straightforward, but diagnosing moderate GHD is difficult and is sometimes made on arbitrary grounds (22).

In approximately 65% of the children diagnosed as GHD, the cause is idiopathic (unknown) [data from the Dutch National Registry of Growth Hormone Treatment in Children]. In the other 35%, a variety of identifiable causes has been shown, like genetic forms and organic lesions in the pituitary or the hypothalamic area (congenital defects, hydrocephalus, tumours like craniopharyngiomas).

When GH deficiency is accompanied by one or more other pituitary hormone deficiencies, hypoglycaemia or prolonged neonatal jaundice, diagnosis is often made during the neonatal period. At an older age, children with GHD usually present with growth retardation and short stature compared to healthy children of the same age, sex and pubertal stage. Age at presentation can vary from a few days after birth until adolescence. The variability in height and age at presentation are highly influenced by the time of onset and the severity of GHD (23).

Due to the pulsatile nature of normal GH-secretion (24), measurement of random serum GH concentrations is of little value in establishing a diagnosis of GHD. Usually, a GH provocation test is performed in order to assess the maximum GH release in response to a secretagogue, mainly a pharmacological agent (25). During the test, serum GH-levels are measured for two hours after the stimulus. If the maximum GH level exceeds a level of 7.7 µg/L (20 mU/L), a normal GH production is assumed. A GH response that remains below 7.7 µg/L, or a combination of GH response below 11.5 µg/L and subnormal levels of IGF-I and IGFBP-3, is considered as too low (26). Unfortunately, it is has been reported that children without GHD might also "fail" one single GH provocation test (27). Therefore, two provocation tests are required to accurately diagnose GHD (22, 28, 29).

If one would aim to determine spontaneous GH secretion, a 24- or 12-hours GH profile should be made, consisting of plasma GH measurements every 20-30

minutes. However, this procedure is rarely performed, as it is invasive, costly and time-consuming. It is necessary in case of suspicion of neurosecretory dysfunction (NSD), since children with this diagnose have normal response to GH provocation tests but impaired spontaneous GH secretion (30, 31).

GH-treatment

Children (and adults) with GHD can be treated by subcutaneous injections with GH. Since the 1950's, GH was isolated from human pituitaries obtained at obductions. Several pituitaries were required to obtain substantial quantities of the hormone for one injection. In 1985, treatment with human pituitary GH-injection was related to Creutzfeld-Jakob disease in some patients (32-34), with onset many years after the use of GH. Treatment with human pituitary GH was immediately stopped in nearly all countries. In 1986, biosynthetic GH, which is identical to the GH produced by healthy children, became available. Besides for children with GHD, treatment with GH also proved to be effective for other children with growth retardation, such as girls with Turner syndrome and short children born short for gestational age (35-37).

In Europe, the accepted daily dose for treatment of children with GHD varies between 0.025 and 0.035 mg/kg·day. In the United States, doses between 0.025 and 0.05 mg/kg·day are commonly used, in Japan 0.025 mg/kg·day or less (38, 39). For many children, growth response to the lower recommended doses is good (40, 41). However, there is variation in response. Several studies found a GH dose-effect on the growth response during the first treatment years (40-42). For the long-term growth response, the dose-effect was smaller and less established (43, 44).

GH therapy induces a significant increase in insulin like growth factor I (IGF-I) and insulin like growth factor binding protein-3 (IGFBP-3) levels and this increase is dose-dependent (40, 45). The majority of children with GHD have IGF-I and IGFBP-3 levels in the lower range prior to GH treatment. During GH treatment, some children display IGF-I values above the normal range for age and sex (SD scores > 2). In retrospective studies on healthy adults, serum levels of IGF-I and IGFBP-3 have been correlated with the risk of developing prostate and premenopausal breast cancers later in life (46, 47). However, until now no direct causative effect of IGF-I on cancer development has been demonstrated. A study on disease recurrence or death in survivors of childhood cancer revealed no increased risk for GH-treated subjects (48). Another study reported an increased incidence of cancer in patients treated with pituitary GH, but this was based on small numbers (49). Regular measurement of IGF-I and IGFBP-3 levels during GH treatment is recommended (29). In case of IGF-I levels above the normal range, lowering of GH dose should be considered (37).

Many children experience important benefits when treated with GH. However, it should be pointed out that, in order to be efficacious, GH therapy should be daily administered by subcutaneous injection, for several years (50). GH treatment is expensive, the costs being on average € 15.000 – 25.000 per child per year, depending on the dosing and on the body surface of the child.

Small for gestational age (SGA)

Weight and length at birth are usually evaluated using reference charts taking into account gestational age at birth and sometimes gender (51, 52). Neonates with a birth weight and/or length at 2 standard deviations (SD) or more below the mean are classified as small for gestational age (SGA). By definition, 2.3% of the life-born neonates is SGA.

Several factors have been mentioned as possible causes for the development of SGA. These include maternal factors as nutrition, smoking, or several diseases, fetal factors as chromosomal or congenital abnormalities, and placental factors such as infarctions or placental development aberrations. However, for the majority of children no underlying pathology can be identified.

Most children born SGA show catch-up growth during the first two years of life and reach a height in the normal range (> -2 SDS) (53, 54). At the age of two years, 15% still has a height below -2 SD scores, and the percentage of adults with a height below the normal range is only slightly smaller (54, 55). The underlying mechanism of inadequate catch-up growth is still not fully understood. Disturbances in the GH/IGF-I axis (Figure 2) may play a role. Sixty percent of SGA children with persistent short stature showed subnormal GH secretion measured over 24 hours, whereas 25% showed low GH peaks during GH provocation tests (56, 57).

Short SGA children show a significant reduces lean body mass, fat mass skin fold measurements and body mass index (58, 59). They have a lower food intake than the recommended daily intake of children of the same age (58). Children born SGA might be at higher risk of a number of adult diseases like diabetes mellitus type II and cardiovascular diseases as birth weight is inversely associated with risks for these disorders (60). Several studies suggest that short children born SGA are psychosocially disadvantaged (61-65). They also have lower total IQ, performal IQ and verbal IQ compared to their peers (65). Another study showed subnormal intelligence and psychological performance in adult men born SGA, especially in those who remained short (66).

GH-treatment

After the introduction of biosynthetic GH, in 1986, trials on GH treatment for short SGA children were started. It appeared that daily GH treatment, at a higher dosage than for children with GHD, caused a significant catch-up growth in height, at least on the short term (67-69). Randomised trials showed that during the first treatment years the growth response is related to the dose of GH treatment (68, 70-73). For the long-term growth response, the dose-effect was smaller, but still significant (38, 39). The recommended GH dose for short SGA children is in the range 0.035 – 0.070 mg/kg·day (35, 74).

Since low birth weight has been associated with the development of diabetes mellitus type 2 and cardiovascular disease in later life, it is important to monitor the effect of GH treatment on risk factors for these diseases in children born SGA. A study on the effect on carbohydrate metabolism during 6 years of GH treatment in short SGA children showed no adverse effects on glucose levels but higher fasting insulin levels and glucose-stimulated insulin levels (75). The effects on body compostion, blood pressure and lipi metabolism were positive.(76)

In a randomised trial, it was found that during the first year of GH treatment, caloric, carbohydrate and protein intake increased significantly, in contrast to untreated controls. In the group of treated children, lean body mass SDS and BMI SDS also showed a significant increase (58). Another study showed an increase in muscle tissue mass during 3 year of GH treatment (77).

Furthermore, improvement in IQ, behavior and self-perception during GH treatment was reported (65).

Prediction models for growth response to growth hormone treatment

For as long as GH treatment is given, it has been investigated which factors influence the growth response and whether it is possible to predict the growth response before the start or in an early phase of treatment. These predictions aim to identify which children will benefit from GH treatment and to determine what would be the optimal dose for a child.

Most prediction models are developed for prediction of short-term response (71, 78-80). These models have the advantage that, because of the short follow-up period, collection of data is relatively easy and the predictive performance of the model can be substantial. Other models aimed to predict growth response over

several years (42) or during puberty (81). Prediction of long-term response to GH treatment is of major importance, as the final aim is to reach a satisfactory adult height.

Most prediction models are build using linear relations between the determinants and the outcome and do not consider interaction terms. An exception are the models developed by Kriström et al (78, 82). They used a nonlinear method described as empirical curve fitting. Their algorithms for the predictions of first-year response to GH treatment, or response during the first two years, contain complex nonlinear terms and interactions.

Modelling of continuous predictors

When building a model, often only straight-line relations between predictors and outcome are considered. However, a non-linear relation between a predictor and the outcome is very well possible. Previously, quadratic or cubic polynomials have been used when linearity was not satisfactory. But the range of curve shapes afforded by conventional low order polynomials is limited. Royston and Altman (83) proposed a more general family of parametric models, called fractional polynomials. Here, one, two or more terms of the form X^p are fitted, the exponents p being chosen from a small pre-selected set of integer and non-integer values (Figure 5). Another method to fit more flexible relations is the use of spline functions (84-86). The idea is that the relation between predictor and outcome is best modelled with different polynomials within intervals of the predictor. The polynomials are connected in the endpoints of the intervals, called knots. A popular choice for the splines are the restricted cubic splines, build up by cubic polynomials, smoothly connected with each other, and linear before the first knot and after the last knot (Figure 6). Although both fractional polynomials and restricted cubic splines can be fitted with standard statistical software, employing them in model building will require extra efforts, especially when flexible modelling has to be considered for several continuous predictors (87). Furthermore, to conclude that another relation is significantly better than the linear relation will require a sufficiently large sample size. The complexity of the prediction rules when more flexible relations are allowed should not be a reason to avoid them. The relation between predictor and outcome can be plotted to check the plausibility. Computations to obtain a predicted value can nowadays easily be performed on a PC or applications can even be made available on websites.

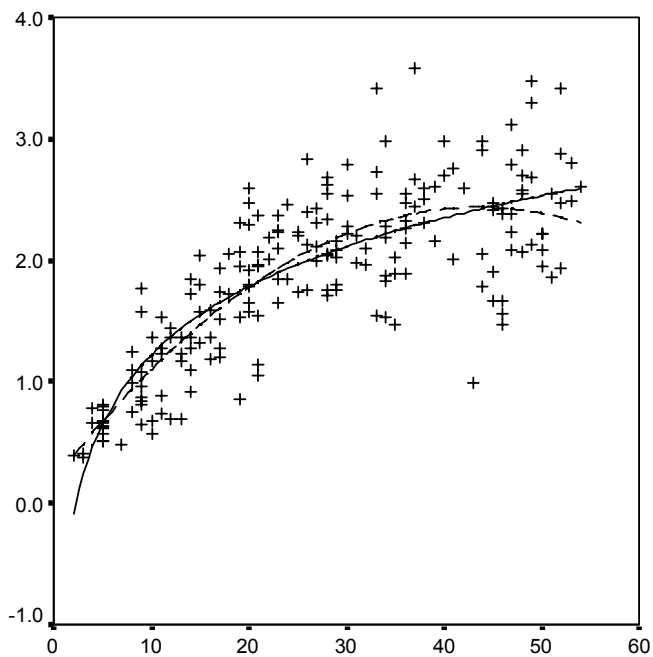**Figure 5.** Example of a relation modelled by a fractional polynomial. The dashed line gives the fitted quadratic curve. The solid line is the fractional polynomial y = a + b*LN(x).
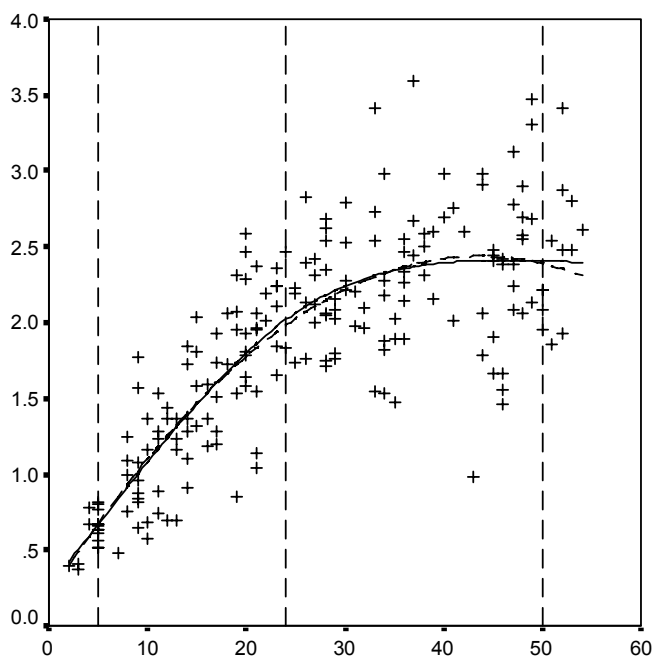

Figure 6. Example of a relation modelled by a restricted cubic spline. The dashed line gives the fitted quadratic curve. The solid line is the fitted restricted cubic spline, with knots at x = 5, x = 24 and x = 50.

Interaction terms

Another aspect not often considered in developing prediction models, is the use of interaction terms. Interaction means that the effects of two predictors in the model cannot be separated; that is, the effect of one predictor on the outcome depends on the level of another predictor. In a model with several predictors, many interaction terms could be tested. Therefore, the interactions to be considered should be pre-specified and kept limited. Harrell (88) described types of interactions that can be important in biostatistics and (clinical) epidemiology. For models predicting growth response, especially the following will apply: Firstly, interactions between treatment and the severity of the disease (growth restriction). Generally, patients with mild disease may receive little benefit from treatment. The severity of growth restriction can be reflected by the height SD score at start, the difference between height SD score and target height SD score, the severity of growth hormone deficiency (maximum GH response to a GH provocation test), or possibly other parameters such as IGF-I and IGFBP-3. Other interactions to be considered are interactions of age at start of treatment with other predictors.

Correction for overfitting

Development of a prediction model is not finished when a well-fitting model is found. It is well known that a prediction model developed by selection of determinants and fitting of the coefficients, both in the same data set, suffers from *overfitting* (89). The degree of overfitting will be larger if the model is more complex. The predictive performance of a prediction model in other data sets than the set on which it is developed will be lower. The estimated regression coefficients of the model will be too extreme and so will be the predicted values when the model is applied to other data sets. This is illustrated in Figure 7, showing two *calibration plots*, where the observed values are plotted against the values predicted by a prediction model. Figure 7A shows a calibration plot of a model suffering from overfitting: for observed low values, the prediction is on average too low, while for observed high values, the prediction is on average too high. The plot in Figure 7B illustrates a model without overfitting, with the data points scattered around the line where the predicted value is equal to the observed value.

A

B
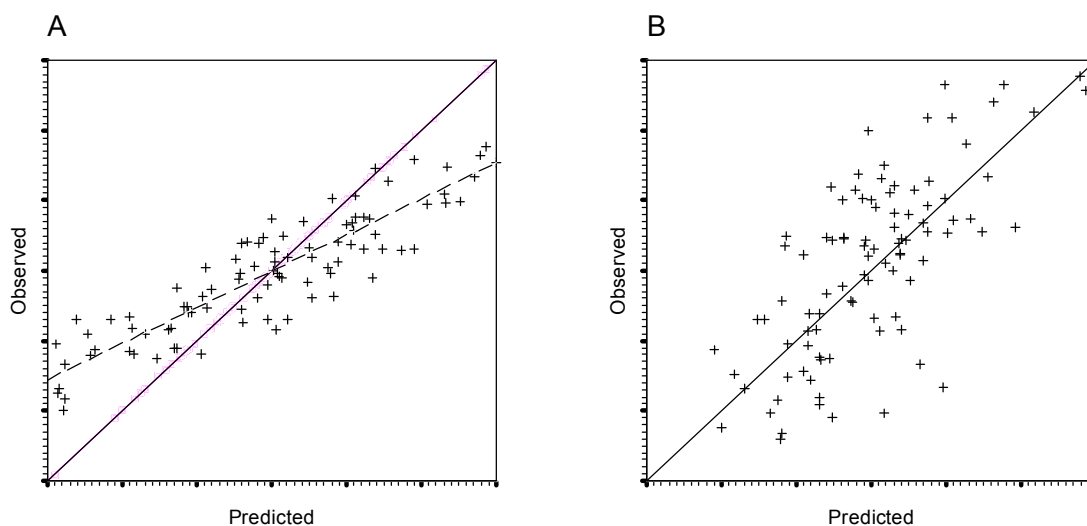
Figure 7. Calibration plots of (A) a prediction model suffering from overfitting and (B) a prediction model without overfitting

Bootstrapping (89, 90) is a recently established method to determine the degree of overfitting of a model and to correct for it. This requires a strict protocol for the development of the model, because the same procedure must be repeated many times, so it must be possible to do this automatically.

Software for bootstrapping of regression models is available in some statistical packages (Stata, SAS, S-Plus, R). However, for complex model-building strategies, including interaction terms and polynomials, additional programming is required.

Investigators might be reluctant to correct a prediction model for overfitting, because, by definition, it will decrease the performance (91). A lower percentage of explained variability (R^2) must be presented and the confidence intervals for predicted values will be wider. However, these corrected results give a more realistic presentation of the predictive ability of the model.

Validation of prediction models

To determine the usefulness of a prediction model, it should be established whether it works satisfactorily for other patient groups. This is called validation (91).

Validation of prediction models for growth response is regularly performed, but mostly it only comprises testing whether the mean of the residuals (observed value minus predicted value) is not significantly different from zero, sometimes accompanied by a graphical inspection of the relation between residuals and observed outcomes. Testing the mean of the prediction errors is, however, only a check for a systematic prediction error (overall too low or too high predictions). Constructing a calibration plot, in which observed values are plotted versus predicted

values (Figure 7), gives information about the relation between the observations and their predictions. Fitting a regression line for this relation enables testing for deviations from the line of identity.

Dealing with missing values

Despite great efforts to achieve a complete dataset, most studies will face the fact of missing values. For analysis of prenatal growth, missings might for example occur in paternal characteristics, last menstruational period but also in variables dealing with social-economic status. For postnatal growth, data for birth length are often missing, because it was thought that briefly stretching the legs and knees in order to measure the length after birth, could be harmful for the development of the hip joints (92), or because of lack of medical staff. Parental height is often not available for adopted children. Also laboratory parameters and measurements of bone age are frequently not available for all subjects.

The disadvantages of selecting only the cases with complete data for all the potential determinants (complete case analysis) are well known (93). This results not only in a smaller sample size, often the selected group is not representative for the population aimed for analysis. However, in many analyses of study data, including growth analyses, this is still how missings are handled. As crude guideline, this can only be sufficient if the proportion of cases with missing data is below 5% (88). If the percentage is larger, imputation of the missing values should be done. Preferably, the imputed value is obtained using relationships among determinants as well as between determinants and the outcome. Even single imputation and then performing an ordinary analysis as if the imputed values were real measurements is usually better than excluding subjects with incomplete data. Better is to perform *multiple imputation* (94, 95). This means that several data sets are created, with different imputed values. These values are random draws from the conditional distribution of the target variable, given the other variables. This reflects the uncertainty of the imputed value. Subsequently, each completed data set is analysed using standard methods and the results are combined, providing valid estimations and standard errors.

Software for multiple imputation and analysis of imputed data sets is nowadays available in many statistical packages, although the possibilities and quality of the imputation methods vary (96-99).

Missing outcome measurements in growth studies

Longitudinal studies can be classified into unbalanced and balanced studies (100). In an unbalanced study, the measurement times vary across subjects. In such a study it is usually not possible to identify non-response, unless scheduled measurement times are recorded, even when no measurement was actually taken. Nonetheless, presence or absence of measurements might not be random. In a balanced study, the number of measurements per subjects is fixed and the measurements are taken at specific time points. In this situation, missing observations can be identified without ambiguity.

A specific case of missing data is dropout, i.e. after a certain point in time no measurements are available. This can be caused by loss to follow-up. It might also be due to censoring, i.e. at the time of analysis some cases have not yet reached the endpoint of interest, e.g. adult height or onset of puberty.

If an analysis of repeated measurements (in growth studies often height observations) is performed, the results will be valid if the missing data are *missing at random.* This condition allows that the unobserved measurements depend on the observed measurements and observed determinants, but it requires that they do not depend on the value of the unobserved measurement itself.

If a data analysis at a certain endpoint is required, like at adult height, but some cases do not have complete follow-up data, common practice is to perform the analysis on the subgroup of patients with observed adult height. However, adult height and the age at which this is attained are usually correlated. Patients reaching adult height at a relatively older age are more likely to have incomplete follow-up data until adult height. Therefore, the group of patients with complete follow-up usually is not representative for the total group. This selection not only occurs in clinical trials, in which it is not always possible to wait until all participants have reached adult height. It also accurs if all available adult heights are selected from a continuous longitudinal growth registration data base. To obtain unbiased results in this situation, a possibility is to carefully select the cases that should be used. Preferably, an analysis should be performed that uses all available data, including the incomplete follow-up of some cases, in a correct and efficient way. Such a method has been developed for medical cost-effectiveness analyses (101) but not yet for growth research.

Outline of this thesis

Chapter 2 describes the validation of a prediction model used in clinical practice for first-year response to GH-treatment in children with GHD. For this validation, data from the Dutch National Registry of Growth Hormone Treatment in Children were used. Children were selected that fulfilled the inclusion criteria used for the development of the original model. Special attention was given to the problem of overfitting and the analysis results in a adjustment of the original model.

This modified, *calibrated* model was validated in a third data set, described in chapter 3. The performances of both the original and the modified model, applied to these new data, were compared.

In chapter 4, prediction models for the prediction of adult height of children with GHD treated with GH are presented. Data were used of a well-defined selection of children registered in the Dutch National Registry of Growth Hormone Treatment in Children. Because of clinical relevance, a model was developed that uses only information available at start of GH treatment, as well as a model using information available after one year of treatment. Separate models were developed for prepubertal as well as pubertal children.

In chapter 5, a prediction model is presented for long-term growth response to GH treatment of children born short for gestational age without spontaneous catch-up growth. With this model, at start of GH treatment a prediction can be obtained for height at onset of puberty and for adult height. Data were used from two clinical trials, with patients randomised to two different doses for GH. This enables to fit a GH dosage effect. Predictions given by this model will give information about the expected adult height for an individual child and might be useful to decide on the GH dose to prescribe.

In chapter 6, prepubertal growth is presented of children born short for gestational age and with a height still below the normal range at the age of 2 years. Aim was to describe which percentage of these children had catch-up growth before the age of 8 years and the factors related to this catch-up growth. A prediction model is presented for height SDS at the age of 8 years, which can be helpful for decisions about treatment with GH.

Chapter 7 describes a method to get unbiased estimations on adult height, in case part of the study group does not yet have complete growth follow-up until adult height. The performance of this method was tested in data from a clinical trial on

response to GH treatment and in data from the Dutch National Registry of Growth Hormone Treatment in Children.

Chapter 8 describes a study of intra-uterine growth. A model was fitted for the effects of physiological determinants, as gender, parity and parental height, on estimated fetal weights, derived from ultra-sound measurements, during pregnancy. Such a model enables to construct individually customised intra-uterine growth charts. Using these charts, fetal growth can be assessed, taking into account the characteristics of the fetus that determine its growth potential. Large deviations from the mean growth, given the characteristics, indicate pathological growth restriction or macrosomia.

In chapter 9, the studies presented in this thesis are discussed and some directions are given for future research.

References

1. Daughaday WH, Rotwein P. Insulin-like growth factors I and II. Peptide, messenger ribonucleic acid and gene structures, serum, and tissue concentrations. Endocr Rev 1989 Feb;10(1):68-91.
2. Hill DJ, Riley SC, Bassett NS, Waters MJ. Localization of the growth hormone receptor, identified by immunocytochemistry, in second trimester human fetal tissues and in placenta throughout gestation. J Clin Endocrinol Metab 1992 Aug;75(2):646-50.
3. Gluckman PD, Harding JE. Fetal growth retardation: underlying endocrine mechanisms and postnatal consequences. Acta Paediatr Suppl 1997 Jul;422:69-72.
4. Gardosi J, Chang A, Kalyan B, Sahota D, Symonds EM. Customised antenatal growth charts. Lancet 1992 Feb 1;339(8788):283-7.
5. Mongelli M, Gardosi J. Longitudinal study of fetal growth in subgroups of a low-risk population. Ultrasound Obstet Gynecol 1995 Nov;6(5):340-4.
6. Isaksson OG, Lindahl A, Nilsson A, Isgaard J. Mechanism of the stimulatory effect of growth hormone on longitudinal bone growth. Endocr Rev 1987 Nov;8(4):426-38.
7. Hindmarsh PC, Brook CG. Normal growth and its endocrine control. In: Brook CG, editor. Clinical Paediatric Endocrinology. Oxford: Blackwell Scientific Publications; 1989. p. 57-73.
8. Van Cauter E, Plat L. Physiology of growth hormone secretion during sleep. J Pediatr 1996 May;128(5 Pt 2):S32-7.
9. Rosenfeld RG. Insulin-like Growth Factors. In: Brook CG, editor. Clinical Paediatric Endocrinoloty. Oxford: Blackwell Science; 1995. p. 107-22.
10. Fredriks AM, van Buuren S, Burgmeijer RJ, Meulmeester JF, Beuker RJ, Brugman E, et al. Continuing positive secular growth change in The Netherlands 1955-1997. Pediatr Res 2000 Mar;47(3):316-23.
11. Gasser T, Muller HG, Kohler W, Prader A, Largo R, Molinari L. An analysis of the mid-growth and adolescent spurts of height based on acceleration. Ann Hum Biol 1985 Mar-Apr;12(2):129-48.
12. Preece MA. The genetic contribution to stature. Horm Res 1996;45 Suppl 2:56-8.
13. Wright CM, Cheetham TD. The strengths and limitations of parental heights as a predictor of attained height. Arch Dis Child 1999 Sep;81(3):257-60.
14. Tanner JM, Goldstein H, Whitehouse RH. Standards for children's height at ages 2-9 years allowing for heights of parents. Arch Dis Child 1970 Dec;45(244):755-62.
15. van Wieringen JC. Secular growth changes. In: Falkner F, Tanner JM, editors. Human Growth. London: Bailliere Tindage; 1978. p. 445-73.
16. Hughes JM, Li L, Chinn S, Rona RJ. Trends in growth in England and Scotland, 1972 to 1994. Arch Dis Child 1997 Mar;76(3):182-9.
17. Hauspie RC, Vercauteren M, Susanne C. Secular changes in growth. Horm Res 1996;45 Suppl 2:8-17.
18. Guyda HJ. Four decades of growth hormone therapy for short children: what have we achieved? J Clin Endocrinol Metab 1999 Dec;84(12):4307-16.
19. Wit JM, De Muinck Keizer-Schrama SMPF, Delemarre-Van de Waal HA. Groeistoornissen. Maarssen: Elsevier/Bungen; 1999.
20. Statistics Netherlands. http://statline.cbs.nl. 2007

21. Hindmarsh PC, Swift PG. An assessment of growth hormone provocation tests. Arch Dis Child 1995 Apr;72(4):362-7; discussion 7-8.

22. Rosenfeld RG, Albertsson-Wikland K, Cassorla F, Frasier SD, Hasegawa Y, Hintz RL, et al. Diagnostic controversy: the diagnosis of childhood growth hormone deficiency revisited. J Clin Endocrinol Metab 1995 May;80(5):1532-40.

23. Adan L, Souberbielle JC, Brauner R. Diagnostic markers of permanent idiopathic growth hormone deficiency. J Clin Endocrinol Metab 1994 Feb;78(2):353-8.

24. Albertsson-Wikland K, Rosberg S. Analyses of 24-hour growth hormone profiles in children: relation to growth. J Clin Endocrinol Metab 1988 Sep;67(3):493-500.

25. Rose SR, Ross JL, Uriarte M, Barnes KM, Cassorla FG, Cutler GB, Jr. The advantage of measuring stimulated as compared with spontaneous growth hormone levels in the diagnosis of growth hormone deficiency. N Engl J Med 1988 Jul 28;319(4):201-7.

26. de Muinck Keizer-Schrama SMPF, Oostdijk W, Rikken B, Waelkens JJJ. Consensus Diagnostiek Kleine Lichaamslengte. Leiden: Bureau van de Groeistichting; 1995.

27. Rochiccioli P, Pienkowski C, Tauber MT, Uboldi F, Enjaume C. Association of pharmacological tests and study of 24-hour growth hormone secretion in the investigation of growth retardation in children: analysis of 257 cases. Horm Res 1991;35(2):70-5.

28. Carel JC, Coste J, Gendrel C, Chaussain JL. Pharmacological testing for the diagnosis of growth hormone deficiency. Growth Horm IGF Res 1998 Feb;8 Suppl A:1-8.

29. Consensus guidelines for the diagnosis and treatment of growth hormone (GH) deficiency in childhood and adolescence: summary statement of the GH Research Society. GH Research Society. J Clin Endocrinol Metab 2000 Nov;85(11):3990-3.

30. Spiliotis BE, August GP, Hung W, Sonis W, Mendelson W, Bercu BB. Growth hormone neurosecretory dysfunction. A treatable cause of short stature. Jama 1984 May 4;251(17):2223-30.

31. Bercu BB, Shulman D, Root AW, Spiliotis BE. Growth hormone (GH) provocative testing frequently does not reflect endogenous GH secretion. J Clin Endocrinol Metab 1986 Sep;63(3):709-16.

32. Koch TK, Berg BO, De Armond SJ, Gravina RF. Creutzfeldt-Jakob disease in a young adult with idiopathic hypopituitarism. Possible relation to the administration of cadaveric human growth hormone. N Engl J Med 1985 Sep 19;313(12):731-3.

33. Powell-Jackson J, Weller RO, Kennedy P, Preece MA, Whitcombe EM, Newsom-Davis J. Creutzfeldt-Jakob disease after administration of human growth hormone. Lancet 1985 Aug 3;2(8449):244-6.

34. Steendijk R, Stoelinga GB. [Growth hormone and Creutzfeldt-Jakob disease] Groeihormoon en de ziekte van Creutzfeldt-Jakob. Ned Tijdschr Geneeskd 1985 Jun 8;129(23):1101-2.

35. Chernausek SD. Treatment of short children born small for gestational age: US perspective, 2005. Horm Res 2005;64 Suppl 2:63-6.

36. Fujieda K, Hanew K, Hirano T, Igarashi Y, Nishi Y, Tachibana K, et al. Growth response to growth hormone therapy in patients with different degrees of growth hormone deficiency. Endocr J 1996 Oct;43 Suppl:S19-25.

37. Wilson TA, Rose SR, Cohen P, Rogol AD, Backeljauw P, Brown R, et al. Update of guidelines for the use of growth hormone in children: the Lawson Wilkins Pediatric Endocrinology Society Drug and Therapeutics Committee. J Pediatr 2003 Oct;143(4):415-21.

38. de Zegher F, Hokken-Koelega A. Growth hormone therapy for children born small for gestational age: height gain is less dose dependent over the long term than over the short term. Pediatrics 2005 Apr;115(4):e458-62.

39. van Pareren Y, Mulder P, Houdijk M, Jansen M, Reeser M, Hokken-Koelega A. Adult height after long-term, continuous growth hormone (GH) treatment in short children born small for gestational age: results of a randomized, double-blind, dose-response GH trial. J Clin Endocrinol Metab 2003 Aug;88(8):3584-90.

40. Cohen P, Bright GM, Rogol AD, Kappelgaard AM, Rosenfeld RG. Effects of dose and gender on the growth and growth factor response to GH in GH-deficient children: implications for efficacy and safety. J Clin Endocrinol Metab 2002 Jan;87(1):90-8.

41. de Muinck Keizer-Schrama S, Rikken B, Hokken-Koelega A, Wit JM, Drop S. Comparative effect of two doses of growth hormone for growth hormone deficiency. The Dutch Growth Hormone Working Group. Arch Dis Child 1994 Jul;71(1):12-8.

42. Ranke MB, Lindberg A, Chatelain P, Wilton P, Cutfield W, Albertsson-Wikland K, et al. 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. KIGS International Board. Kabi Pharmacia International Growth Study. J Clin Endocrinol Metab 1999;84(4):1174-83.

43. Cutfield W, Lindberg A, Albertsson Wikland K, Chatelain P, Ranke MB, Wilton P. Final height in idiopathic growth hormone deficiency: the KIGS experience. KIGS International Board. Acta Paediatr Suppl 1999 Feb;88(428):72-5.

44. Wit JM, Kamp GA, Rikken B. Spontaneous growth and response to growth hormone treatment in children with growth hormone deficiency and idiopathic short stature. Pediatr Res 1996 Feb;39(2):295-302.

45. de Muinck Keizer-Schrama SM, Rikken B, Wynne HJ, Hokken-Koelega AC, Wit JM, Bot A, et al. Dose-response study of biosynthetic human growth hormone (GH) in GH-deficient children: effects on auxological and biochemical parameters. Dutch Growth Hormone Working Group. J Clin Endocrinol Metab 1992 Apr;74(4):898-905.

46. Chan JM, Stampfer MJ, Giovannucci E, Gann PH, Ma J, Wilkinson P, et al. Plasma insulin-like growth factor-I and prostate cancer risk: a prospective study. Science 1998 Jan 23;279(5350):563-6.

47. Hankinson SE, Willett WC, Colditz GA, Hunter DJ, Michaud DS, Deroo B, et al. Circulating concentrations of insulin-like growth factor-I and risk of breast cancer. Lancet 1998 May 9;351(9113):1393-6.

48. Sklar CA, Mertens AC, Mitby P, Occhiogrosso G, Qin J, Heller G, et al. Risk of disease recurrence and second neoplasms in survivors of childhood cancer treated with growth hormone: a report from the Childhood Cancer Survivor Study. J Clin Endocrinol Metab 2002 Jul;87(7):3136-41.

49. Swerdlow AJ, Higgins CD, Adlard P, Preece MA. Risk of cancer in patients treated with human pituitary growth hormone in the UK, 1959-85: a cohort study. Lancet 2002 Jul 27;360(9329):273-7.

50. Lippe B, Frasier SD. How should we test for growth hormone deficiency, and whom should we treat? J Pediatr 1989 Oct;115(4):585-7.

51. Niklasson A, Ericson A, Fryer JG, Karlberg J, Lawrence C, Karlberg P. An update of the Swedish reference standards for weight, length and head circumference at birth for given gestational age (1977-1981). Acta Paediatr Scand 1991;80(8-9):756-62.

52. Usher R, McLean F. Intrauterine growth of live-born Caucasian infants at sea level: standards obtained from measurements in 7 dimensions of infants born between 25 and 44 weeks of gestation. J Pediatr 1969 Jun;74(6):901-10.

53. Hokken-Koelega AC, de Ridder MA, Lemmen RJ, den Hartog H, de Muinck Keizer-Schrama SM, Drop SL. Children born small for gestational age: do they catch up? Pediatr Res 1995 Aug;38(2):267-71.

54. Karlberg J, Albertsson-Wikland K. Growth in full-term small-for-gestational-age infants: from birth to final height. Pediatr Res 1995 Nov;38(5):733-9.

55. Paz I, Seidman DS, Danon YL, Laor A, Stevenson DK, Gale R. Are children born small for gestational age at increased risk of short stature? Am J Dis Child 1993 Mar;147(3):337-9.

56. Boguszewski M, Rosberg S, Albertsson-Wikland K. Spontaneous 24-hour growth hormone profiles in prepubertal small for gestational age children. J Clin Endocrinol Metab 1995 Sep;80(9):2599-606.

57. de Waal WJ, Hokken-Koelega AC, Stijnen T, de Muinck Keizer-Schrama SM, Drop SL. Endogenous and stimulated GH secretion, urinary GH excretion, and plasma IGF-I and IGF-II levels in prepubertal children with short stature after intrauterine growth retardation. The Dutch Working Group on Growth Hormone. Clin Endocrinol (Oxf) 1994 Nov;41(5):621-30.

58. Boonstra VH, Arends NJ, Stijnen T, Blum WF, Akkerman O, Hokken-Koelega AC. Food intake of children with short stature born small for gestational age before and during a randomized GH trial. Horm Res 2006;65(1):23-30.

59. Leger J, Carel C, Legrand I, Paulsen A, Hassan M, Czernichow P. Magnetic resonance imaging evaluation of adipose tissue and muscle tissue mass in children with growth hormone (GH) deficiency, Turner's syndrome, and intrauterine growth retardation during the first year of treatment with GH. J Clin Endocrinol Metab 1994 Apr;78(4):904-9.

60. Barker DJ, Winter PD, Osmond C, Margetts B, Simmonds SJ. Weight in infancy and death from ischaemic heart disease. Lancet 1989 Sep 9;2(8663):577-80.

61. Low JA, Handley-Derry MH, Burke SO, Peters RD, Pater EA, Killen HL, et al. Association of intrauterine fetal growth retardation and learning deficits at age 9 to 11 years. Am J Obstet Gynecol 1992 Dec;167(6):1499-505.

62. Pryor J, Silva PA, Brooke M. Growth, development and behaviour in adolescents born small-for-gestational-age. J Paediatr Child Health 1995 Oct;31(5):403-7.

63. van der Reijden-Lakeman I, Slijper FM, van Dongen-Melman JE, de Waal WJ, Verhulst FC, Rosenfeld RG, et al. Self-concept before and after two years of growth hormone treatment in intrauterine growth-retarded children. Horm Res 1996 Oct;46(2):88-94.

64. van der Reijden-Lakeman IE, de Sonneville LM, Swaab-Barneveld HJ, Slijper FM, Verhulst FC. Evaluation of attention before and after 2 years of growth hormone treatment in intrauterine growth retarded children. J Clin Exp Neuropsychol 1997 Feb;19(1):101-18.

65. van Pareren YK, Duivenvoorden HJ, Slijper FS, Koot HM, Hokken-Koelega AC. Intelligence and psychosocial functioning during long-term growth hormone therapy in children born small for gestational age. J Clin Endocrinol Metab 2004 Nov;89(11):5295-302.

66. Lundgren EM, Cnattingius S, Jonsson B, Tuvemo T. Intellectual and psychological performance in males born small for gestational age with and without catch-up growth. Pediatr Res 2001 Jul;50(1):91-6.

67. Albertsson-Wikland K. Growth hormone secretion and growth hormone treatment in children with intrauterine growth retardation. Swedish Paediatric Study Group for Growth Hormone Treatment. Acta Paediatr Scand Suppl 1989;349:35-41; discussion 53-4.

68. Chatelain P, Job JC, Blanchard J, Ducret JP, Oliver M, Sagnard L, et al. Dose-dependent catch-up growth after 2 years of growth hormone treatment in intrauterine growth-retarded children. Belgian and French Pediatric Clinics and Sanofi-Choay (France). J Clin Endocrinol Metab 1994 Jun;78(6):1454-60.

69. Stanhope R, Preece MA, Hamill G. Does growth hormone treatment improve final height attainment of children with intrauterine growth retardation? Arch Dis Child 1991 Oct;66(10):1180-3.

70. Boguszewski M, Albertsson-Wikland K, Aronsson S, Gustafsson J, Hagenas L, Westgren U, et al. Growth hormone treatment of short children born small-for-gestational-age: the Nordic Multicentre Trial. Acta Paediatr 1998 Mar;87(3):257-63.

71. Ranke MB, Lindberg A, Cowell CT, Wikland KA, Reiter EO, 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). J Clin Endocrinol Metab 2003 Jan;88(1):125-31.

72. Sas T, de Waal W, Mulder P, Houdijk M, Jansen M, Reeser M, et al. Growth hormone treatment in children with short stature born small for gestational age: 5-year results of a randomized, double-blind, dose-response trial. J Clin Endocrinol Metab 1999 Sep;84(9):3064-70.

73. de Zegher F, Albertsson-Wikland K, Wilton P, Chatelain P, Jonsson B, Lofstrom A, et al. Growth hormone treatment of short children born small for gestational age: metanalysis of four independent, randomized, controlled, multicentre studies. Acta Paediatr Suppl 1996 Oct;417:27-31.

74. Clayton PE, Cianfarani S, Czernichow P, Johannsson G, Rapaport R, Rogol A. Management of the Child Born Small for Gestational Age Child (SGA) through to Adulthood: A Consensus Statement of the International Societies of Paediatric Endocrinology and the Growth Hormone Research Society. J Clin Endocrinol Metab 2007 Jan 2.

75. Sas T, Mulder P, Aanstoot HJ, Houdijk M, Jansen M, Reeser M, et al. Carbohydrate metabolism during long-term growth hormone treatment in children with short stature born small for gestational age. Clin Endocrinol (Oxf) 2001 Feb;54(2):243-51.

76. Sas T, Mulder P, Hokken-Koelega A. Body composition, blood pressure, and lipid metabolism before and during long-term growth hormone (GH) treatment in children with short stature born small for gestational age either with or without GH deficiency. J Clin Endocrinol Metab 2000 Oct;85(10):3786-92.

77. Leger J, Garel C, Fjellestad-Paulsen A, Hassan M, Czernichow P. Human growth hormone treatment of short-stature children born small for gestational age: effect on muscle and adipose tissue mass during a 3-year treatment period and after 1 year's withdrawal. J Clin Endocrinol Metab 1998 Oct;83(10):3512-6.

78. Kristrom B, Karlberg J, Albertsson-Wikland K. Prediction of the growth response of short prepubertal children treated with growth hormone. Swedish Paediatric Study Group for GH treatment. Acta Paediatr 1995 Jan;84(1):51-7.

79. Ranke MB, Guilbaud O, Lindberg A, Cole T. Prediction of the growth response in children with various growth disorders treated with growth hormone: analyses of data from the Kabi Pharmacia International Growth Study. International Board of the Kabi Pharmacia International Growth Study. Acta Paediatr Suppl 1993;82(Suppl 391):82-8; discussion 9.

80. Schonau E, Westermann F, Rauch F, Stabrey A, Wassmer G, Keller E, et al. A new and accurate prediction model for growth response to growth hormone treatment in children with growth hormone deficiency. Eur J Endocrinol 2001;144(1):13-20.

81. Ranke MB, Lindberg A, Martin DD, Bakker B, Wilton P, Albertsson-Wikland K, et al. The mathematical model for total pubertal growth in idiopathic growth hormone (GH) deficiency suggests a moderate role of GH dose. J Clin Endocrinol Metab 2003 Oct;88(10):4748-53.

82. Albertsson-Wikland K, Kristrom B, Rosberg S, Svensson B, Nierop AFM. Validated Multivariate Models Predicting the Growth Response to GH Treatment in Individual Short Children with a Broad Range in GH Secretion Capacities. Pediatric Research 2000;48(4):475-84.

83. Royston P, Altman DG. Regression using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling. Applied Statistics 1994;43(3):429-67.

84. De Boor C. A Practical Guide to Splines. New York: Springer-Verlag; 1978.

85. Harrell FE, Jr., Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. J Natl Cancer Inst 1988 Oct 5;80(15):1198-202.

86. Smith PL. Spines as a useful and convenient statistical tool. American Statistician 1979;33:57-62.

87. Sauerbrei W, Royston P. Building Multivariable Prognostic and Diagnostic Models: Transformation of the Predictors by Using Fractional Polynomials. Journal of the Royal Statistical Society, Series A 1999;162(1):71-94.

88. Harrell FE, Jr. Regression modeling strategies. New York: Springer-Verlag; 2001.

89. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15(4):361-87.

90. Efron B. TR. An introduction to the bootstrap. New York; 1993.

91. Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med 2000;19(4):453-73.

92. Engelberts AC, Koerts B, Waelkens JJ, Wit JM, Burger BJ. [Measuring the length of newborn infants] Lengtemeting bij de pasgeborene. Ned Tijdschr Geneeskd 2005 Mar 19;149(12):632-6.

93. Donner A. The Relative Effectiveness of Procedures Commonly Used in Multiple Regression Analysis for Dealing with Missing Values. The American Statistician 1982;36(4):378-81.

94. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York; 1987.

95. Schäfer JL. Analysis of Incomplete Multivariate Data. New York; 1997.

96. SAS Online Doc 9.1.3. Cary, NC: SAS Institute Inc.; 2006.

97. Royston P. ICE: Stata module for multiple imputation of missing values. S446602 ed: Boston College Department of Economics; 2006.

98. van Buuren S, Oudshoorn CGM. Multiple imputation by chained equations: MICE V1.0 User's Manual Leiden: TNO Preventie en Gezondheid; 2000.

99. van Buuren S. www.multiple-imputation.com. 2005

100. Verbeke G, Molenberghs G. Linear Mixed Models in Practice. New York: Springer-Verlag; 1997.

101. Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. Biometrics 1997 Jun;53(2):419-34.

2

Validation and calibration of the KIGS prediction model for children with idiopathic growth hormone deficiency
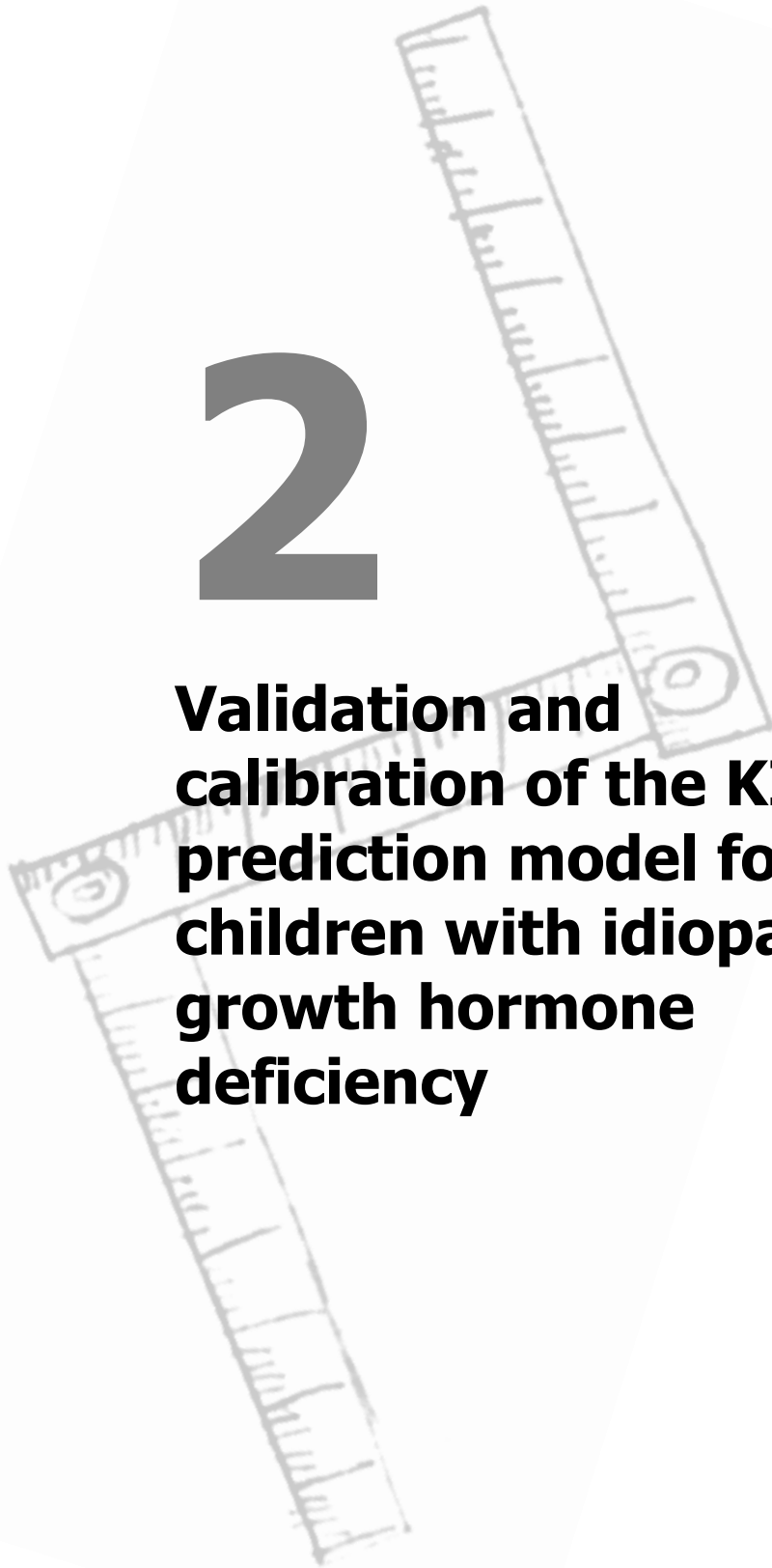
Abstract

In 1999 a model was published for prediction of growth in children with idiopathic growth hormone deficiency (IGHD) during GH therapy, derived using data from the KIGS database (Pharmacia & Upjohn, Inc., International Growth Database). We validated and calibrated this KIGS model for growth in the first year of GH therapy using data of 136 Dutch children with IGHD. Observed versus predicted outcomes were plotted and the fitted regression line was significantly different from the line of identity ($p = 0.03$). It appeared that the predictions were too extreme: relatively low predictions were too low, relatively high predictions too high. This is a well-known phenomenon in the context of prediction models, called over-optimism. For valid application to other data the KIGS predictions should be calibrated. Calibrated predictions are obtained using $Y_{cal} = Y_{orig} + (2.153 - 0.192 * Y_{orig})$, where Y_{cal} is the calibrated prediction and Y_{orig} the KIGS prediction. The calibrated prediction will be higher than the original KIGS prediction when the original prediction was below 11.2 cm/yr and lower otherwise. The variability of the prediction errors of the calibrated predictions was positively related to the value of the prediction ($p < 0.001$), described by the equation $SD_{pred\ err} = -1.017 + 0.286 * Y_{cal}$. Our calibrated model will give better predictions for children with IGHD fulfilling the same criteria.

Introduction

Because of the variability of response to growth hormone (GH) treatment during the last decade several prediction models have been developed (1-4). It is well-known that prediction models often perform less well than expected when applied to new patients, either of the same population or other populations. According to recent statistical insights this is in many instances due to the problem of over-optimism, which results in predictions that are too extreme (5, 6). This over-optimism, also called overfitting, is more pronounced if the prediction model is developed by selecting the predictor variables from a group of possible candidate predictors and this selection is determined by the data. If a representative data set is used, and a proper modelling method is applied, there is practically no doubt that in the final model the selected variables will be related to the outcome, also in other data sets fulfilling the same criteria. But the selection of exactly *these* predictor variables and exactly *these* values of the estimated coefficients of the predictors will partly be determined by the accidental characteristics of the data set. This will give a model that is too much data-driven and will differentiate the subjects in predicted outcome too extremely. Even apart from this effect of variable selection, overfitting is likely to occur in each prediction model with two or more prediction variables (7). For valid application of a prediction model to new patients the predictions should therefore be calibrated by shrinking them to less extreme values.

In this study we used data of Dutch children with idiopathic growth hormone deficiency (IGHD) to validate the prediction model derived on the database of the Kabi Pharmacia International Growth Study (KIGS; Pharmacia & Upjohn, Inc., International Growth Database) (8) for growth velocity of children with IGHD during the first year of GH therapy (2). This model, further referred to as the KIGS model, was developed on a data set containing data of 593 children and is widely used in clinical practice. A validation of the model, on data of subsequent children from the KIGS database (temporal validation) as well as on data from other studies (external validation) was described in the paper (2). Our study validated the model on a large external data set with special attention to the problem of overfitting.

Subjects and methods

Patients

For the validation of the model we used data from the database of the Dutch Growth Foundation, containing data of all Dutch children who have been or are being treated with GH. We applied the same in- and exclusion criteria as were used for the cohort from which the KIGS model for children with IGHD was derived, although other brands of biosynthetic GH than Genotropin ® were used too. Height measurements for calculation of first year height velocity (HV) were allowed to have been done between 0.8 - 1.25 year after the start of treatment, comparable with the validation described in the paper in which the KIGS model was presented (2).

Observed and predicted outcomes

HV was computed correcting for the time-interval between start of treatment and the actual date of the height measurement after one year of treatment. For the predictors used in the model the SD scores were computed with the same reference data as used for the KIGS cohort (9-11). The KIGS prediction was computed using the regression equation based on data at start of treatment:

predicted HV (cm/yr) = 14.55 - 1.37 * maximum GH response (ln;µg/L)
- 0.32 * age (yr) + 0.32 * birth weight SDS + 1.62 * GH dose (ln;IU/kg·wk)
- 0.40 * (height SDS - midparental height SDS) + 0.29 * weight SDS.

Statistical Analysis

For the validation we followed the method described by Van Houwelingen (12). Observed HV was plotted versus predicted HV in a calibration plot. A regression analysis was done with observed HV as outcome and predicted HV as determinant and it was tested whether the regression line was significantly different from the line of identity (where observed HV is equal to predicted HV). The estimated coefficients from the regression analysis were then used to determine the calibration correction. Using the calibrated predictions, we tested the homogeneity of variance of the prediction errors using the Spearman rank correlation between the absolute values of the prediction errors and the values of the predictions (13). In case of a significant correlation the relation was modelled (14). In searching for the regression model for absolute residuals depending on predicted values, we allowed for fractional polynomials (15).

Results

136 cases of the Dutch database fulfilled the in- and exclusion criteria. The characteristics of this Dutch cohort were well within the ranges of the KIGS cohort (Table 1).

Table 1. Characteristics of children with IGHD from the KIGS cohort and the Dutch cohort

	KIGS cohort			Dutch cohort		
	Mean	Min	Max	Mean	Min	Max
Age (yr)	7.3	2.1	11.9	6.8	2.0	11.9
Height SDS	-2.6	-5.4	-0.4	-2.1	-5.3	-0.3
Weight SDS	-2.2	-3.8	-0.7	-1.9	-4.6	0.8
MPH SDS	-0.6	-3.4	2.3	0.4	-2.1	3.4
Height SDS - MPH SDS	-1.9	-6.2	1.7	-2.5	-6.6	0.0
BW SDS	-0.5	-2.0	3.6	-0.3	-1.9	3.1
Max GH-peak (μg/L)	5.6	0.3	10.0	5.2	0.5	10.0
GH dose (IU/kg·wk)	0.6	0.2	1.3	0.6	0.3	1.3
HV in first year (cm/yr)	9.2	3.5	16.8	10.2	5.4	17.4

MPH = midparental height
BW = birth weight

The prediction errors (observed HV minus predicted HV) had a mean of 0.25 cm/yr and a standard deviation (SD) of 1.95 cm/yr. Observed versus predicted values were plotted in Figure 1, together with the line of identity (where observed HV is equal to predicted HV, dotted line) and the fitted regression line. The figure shows that relatively low predictions tended to be underestimated and relatively high predictions tended to be overestimated. The deviation of the regression line from the line of identity was statistically significant ($p = 0.03$) and the estimated slope was 0.808. To get a fitted regression line with slope 1 (line of identity), which is preferred, points at the left, so relatively low predictions, should be shifted a little to the right, and points at the right, relatively high predictions, should be shifted a little to the left. This calibration correction is described by the formula:

$$Y_{cal} = Y_{orig} + (2.153 - 0.192 * Y_{orig})$$

where Y_{cal} is the calibrated prediction and Y_{orig} is the original KIGS prediction.
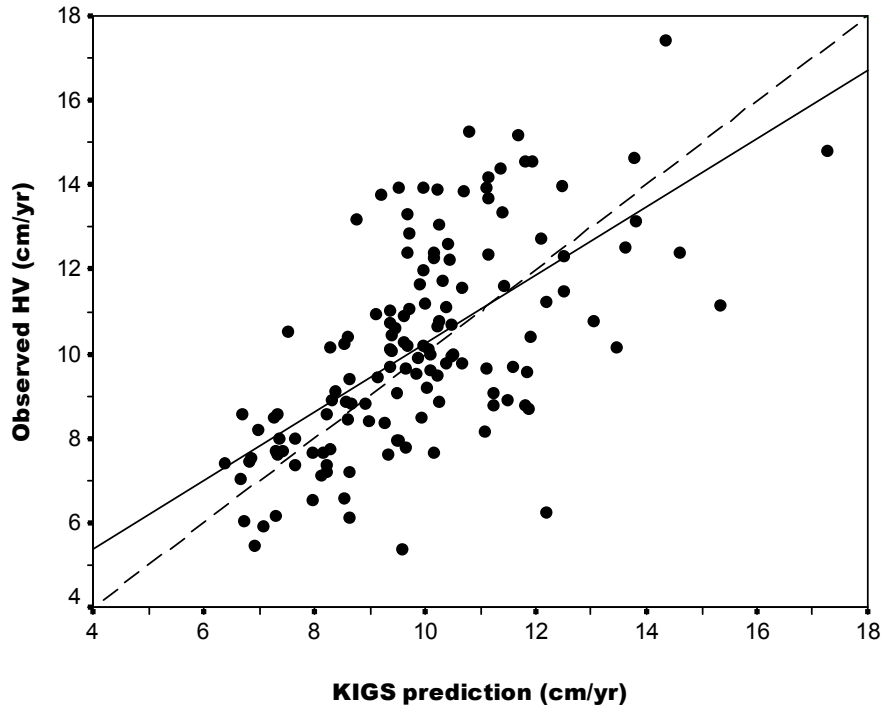
Figure 1. First year HV (cm/yr): observed versus KIGS predicted values, together with the line of identity (dotted) and the fitted regression line (solid)
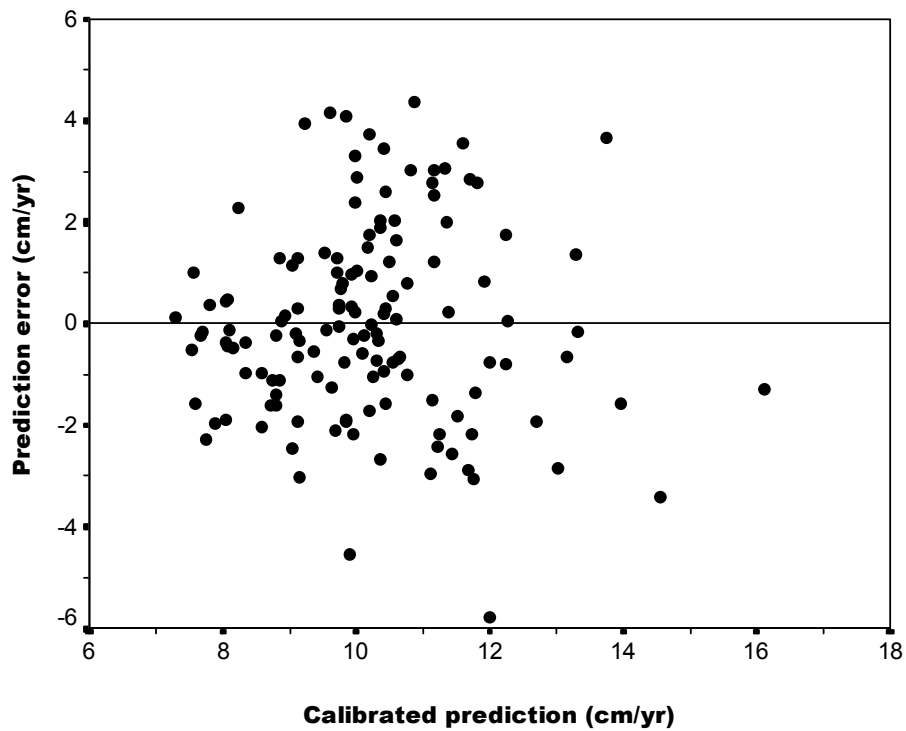
Figure 2. Prediction errors versus predicted value of the calibrated model

The correction term, (*2.153 - 0.192 * Y_orig*), has a positive value for predictions below 11.2 cm/yr (point of intersection of the regression line and the line of identity) and a negative value for predictions above 11.2 cm/yr.

Using the calibrated model the prediction errors had a mean of zero and an SD of 1.91 cm/yr. When prediction errors were plotted versus calibrated predictions (Figure 2) an increase in variance of the prediction errors was seen with increasing value of the prediction. The absolute values of the prediction errors turned out to be positively related to the value of the predictions ($p < 0.001$). The relation between the calibrated predictions and the variability of their prediction errors was described with the model:

$$SD_{pred\ error} = -1.017 + 0.286 * Y_{cal}$$

The differences between the predictions of the original KIGS model and the calibrated model are illustrated by two examples (Table 2). Example 1 shows that for a child with a relatively low prediction (predicted HV using the KIGS model is 7.0 cm/yr) the calibrated prediction is higher (7.8 cm/yr) whereas the 95% prediction interval (PI) is smaller than the PI using the KIGS model. Example 2 demonstrates that a relatively high prediction decreases after calibration (KIGS prediction 14.0 cm/yr, calibrated prediction 13.5 cm/yr) whereas the 95% PI becomes much wider.

Table 2. Examples of predictions and 95% prediction intervals using the KIGS model and using the calibrated model

	Original KIGS model			Calibrated model		
	Prediction (cm/yr)	95% PI	Width of PI	Prediction (cm/yr)	95% PI	Width of PI
Example 1	7.0	4.1 - 9.9	5.7	7.8	5.4 - 10.2	4.8
Example 2	14.0	11.1 - 16.9	5.7	13.5	7.9 - 19.0	11.1

For the calculation of the SD scores of the auxological characteristics used in the KIGS model, UK reference data (9,10) were used. When for our validation in the Dutch cohort national reference data were used, namely for height and weight at start of treatment the reference data of 1997 of Fredriks et. al. (16) and for height of the parents the reference data of 1965 of Van Wieringen (17), the mean of the prediction errors was 0.23 cm/yr . The overfitting was again significant ($p < 0.001$) and the estimated slope in the calibration plot was 0.791. If the 1997 reference data was also used for height of the parents the estimated slope was simular but the mean prediction error increased to 0.60 cm/yr and was significantly different from zero ($p = 0.001$).

Discussion

In this paper for the first time a validation according to modern statistical insights into the behaviour of prediction models was applied to a widely used prediction model for first year response to GH therapy. Special attention was given to the problem of overfitting. The analysis resulted in a calibrated model with a better fit when applied to new data.

The method of validation we have used is based on and illustrated by a calibration plot where observed values are plotted versus predicted values, in our case observed HV in the first year of GH treatment versus the prediction given by the KIGS model. Four examples of calibration plots are shown in Figure 3. In all figures the line of identity, where the observed outcome is equal to the predicted outcome, is drawn. Perfect predictions will all lay on the line of identity (Fig. 3a). If the predictions are not perfect, but unbiased, the plot will show points scattered randomly around the line of identity (Fig. 3b). Any other pattern in the calibration plot indicates that the prediction model does not fit well to the validation data. As an example Figure 3c shows a scatter plot with points too much on the right side of the line of identity. While the vertical position of the points is fixed, determined by the observed value, the horizontal position should be corrected by shifting all points a little to the left. This indicates that the original predictions are on average too high. Another, frequently occurring pattern is that the scatter plot shows a relation between observed and predicted outcomes with a lower slope than the line of identity (Fig. 3d). Now a correction should be done by shifting the points at the left side a little to the right and the points at the right side a little to the left. This indicates underestimation of the lowest predictions and overestimation of the highest predictions. The model classifies too extremely into good and bad responses and therefore this phenomenon is called overfitting or over-optimism. Overfitting is common when the selection of predictors and the estimation of their coefficients are both guided by the same data set (18, 19). The occurrence of overfitting is not the result of differences between the modelling group and the validation group. Suppose we start with one data set, and split this randomly up into a modelling group and a validation group, so without systematic differences between the two groups. If we develop a prediction model on the modelling group, applying the model to the validation group will very likely show overfitting too.
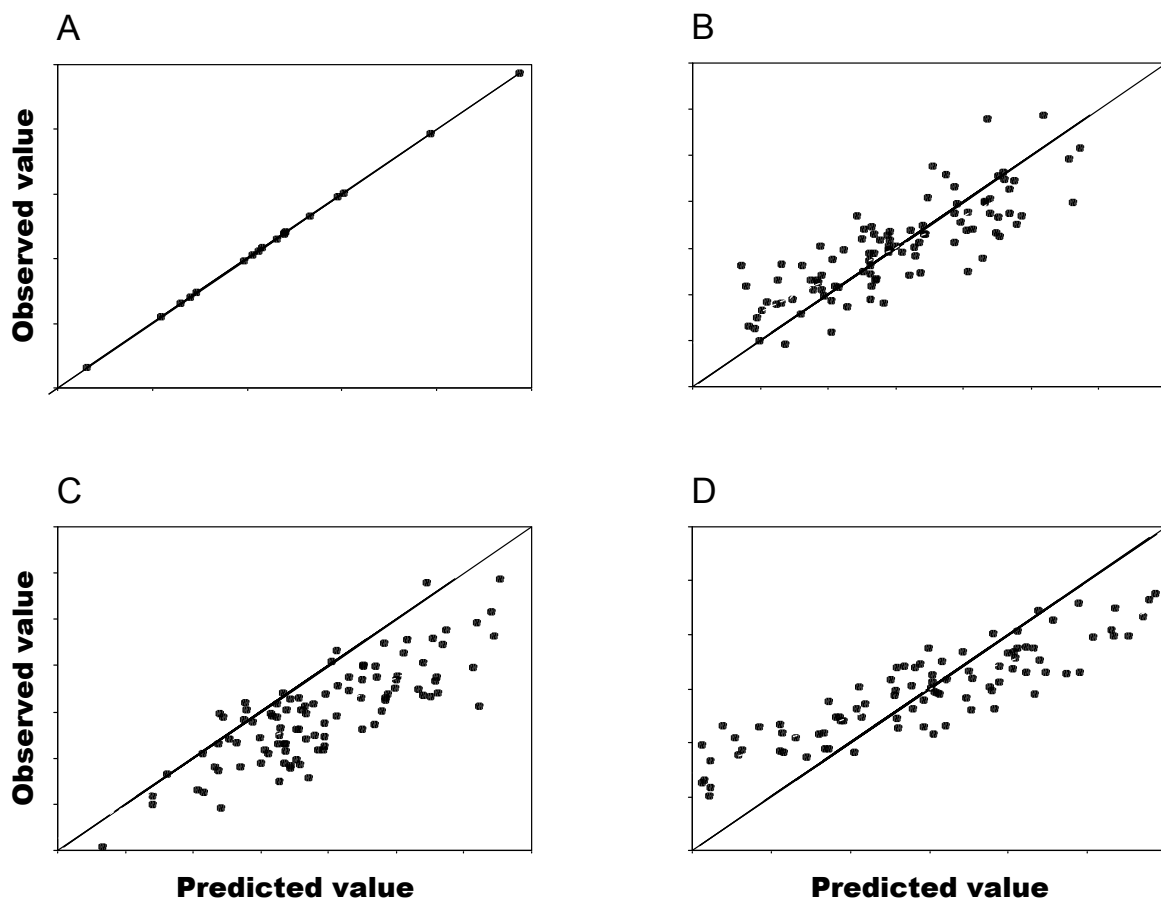
Figure 3. Examples of calibration plots of prediction models for which the predictions are perfect (A), unbiased (B), too high (C) and overfitted (D)

In our validation of the KIGS model the calibration plot showed a pattern like Figure 3d, although less extreme. The slope of the regression line was 0.808 and the line was significantly different from the line of identity. The overfitting could be corrected using the formula:

$$Y_{cal} = Y_{orig} + (2.153 - 0.192 * Y_{orig})$$

This leads to a correction of the predictions from the KIGS model ranging from + 0.92 cm/yr for a prediction of 6.4 cm/yr (the lowest KIGS prediction in our validation group) to - 1.2 cm/yr when 17.3 cm/yr was predicted (highest predicted value). Corrections for predictions of in-between values will be smaller, and for a predicted HV of 11.2 cm/yr the correction term is zero.

The validation performed on the KIGS model as described in the paper of Ranke et. al. (2), consisting of testing whether there was a significant difference between observed and predicted values, is only a check for a systematic prediction error (overall too low or too high predictions) and will not disclose overfitting.

In the Dutch cohort the SD of the prediction errors of the calibrated model was higher than the published SD of 1.46 cm/yr of the prediction errors found in the KIGS cohort (2). This was to be expected because a prediction model will be less precise for new data than for the data on which it was derived. Another important finding was the dependency of variance of the prediction error on the magnitude of the prediction. We found that the prediction error SD was ranging from 1.07 cm/yr for a prediction of 7.3 cm/yr (our lowest calibrated prediction) to 3.58 cm/yr when the predicted outcome was 16.1 cm/yr (highest calibrated prediction). This means that the accuracy of a high prediction of HV is less than the accuracy of a low prediction, as reflected in the 95% PI accompanying the predicted response.

When developing the KIGS model, the analysts also checked for a relation between the variance of the prediction error and the prediction, by plotting Studentized residuals versus predicted HV (2). In contrast to our findings, they did not find a relation. We cannot explain the discrepancy between their and our findings.

For the calculation of the SD scores for the KIGS cohort, UK reference data (9, 10) were used. Dutch children and adults are known to be among the tallest an earth (20). Moreover the secular trend in the Netherlands is relatively large (16, 10). If Dutch reference data were used, for parents dated back one generation, the predictions were not much different from the original predictions, but if we calculated the SD scores ignoring the secular trend, the predictions were signifantly too low. In both situations the overfitting remained. Thus, to apply the KIGS model correctly, one should use the UK reference data.

The KIGS model is a well-defined and easily applicable model, developed on a population of sufficient size. The percentage explained variability ($R^2 = 0.61$) might improve if more flexibility of the relations was allowed, or other patient characteristics could be used, but the aim might have been to restrict to simple modelling with widely available characteristics. However, the problem of overfitting, very likely to be present if model selection and estimation are done using the same data set, was not taken into account.

As a consequence of our calibration and of modelling the dependency of the prediction error SD, we propose that a KIGS prediction for the first year growth response in children with IGHD should be modified using the formula:

$$Y_{cal} = Y_{orig} + 2.153 - 0.192 * Y_{orig}$$

where Y_{cal} is the calibrated prediction and Y_{orig} is the original KIGS prediction. The 95% PI is given by:

$$Y_{low} = Y_{cal} - 1.96 * (-1.017 + 0.286*Y_{cal})$$

$$Y_{high} = Y_{cal} + 1.96 * (-1.017 + 0.286*Y_{cal})$$

where again Y_{cal} is the calibrated prediction, Y_{low} is the lower limit of the PI and Y_{high} is the upper limit of the PI. Figure 4 presents the calibrated predictions versus the original KIGS predictions, with 95% PI. It shows that for instance an original prediction of 14 cm/yr should be modified to 13.5 cm/yr, with 95% PI 8 - 19 cm/yr.
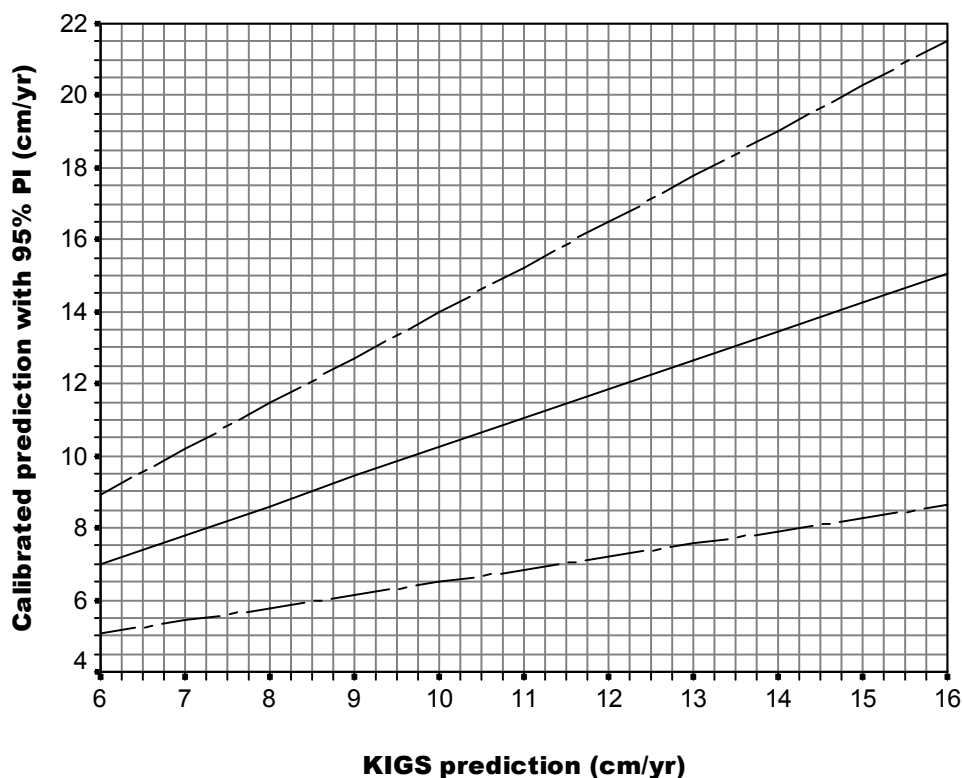
Figure 4. Relation between the calibrated prediction and the value of the KIGS prediction, with 95% prediction interval

Our validation and calibration of the KIGS model is quite different from developing a new model. In our calibrated model we maintained the relative contribution of the predictor variables as determined for the original model, but the outcomes were adjusted using a correction term which depends on the value of the original prediction. Clearly these calibrated predictions fitted better to the data of the Dutch cohort. Whether both the data from the KIGS cohort, used for modelling, as well as the data from the Dutch cohort, used for calibration, are representative for other cohorts of children with IGHD fulfilling the same in- and exclusion criteria is not yet known. We, however, postulate that our modification of the KIGS prediction rule will give better predictions for these children as well.

When new prediction models are developed in the future, they should be evaluated as to whether a calibration is required. The best way to examine this is by validation of the model on external data. This will improve the accuracy and benefit of prediction models for clinical practice.

References

1. Ranke MB, Guilbaud O, Lindberg A, Cole T. 1993 Prediction of the growth response in children with various growth disorders treated with growth hormone: analyses of data from the Kabi Pharmacia International Growth Study. International Board of the Kabi Pharmacia International Growth Study. *Acta Paediatr Suppl* 82(Suppl 391): 82-88

2. Ranke MB, Lindberg A, Chatelain P, Wilton P, Cutfield W, Albertsson-Wikland K, Price DA. 1999 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. KIGS International Board. Kabi Pharmacia International Growth Study. *J Clin Endocrinol Metab* 84(4): 1174-1183

3. Albertsson-Wikland K, Kriström B, Rosberg S, Svensson B, Nierop AFM. 2000 Validated Multivariate Models Predicting the Growth Response to GH Treatment in Individual Short Children with a Broad Range in GH Secretion Capacities. *Pediatr Res* 48(4): 475-484

4. Schönau E, Westermann F, Rauch F, Stabrey A, Wassmer G, Keller E, Brämswig J, Blum WF. 2001 A new and accurate prediction model for growth response to growth hormone treatment in children with growth hormone deficiency. *Eur J Endocrinol* 144(1): 13-20.

5. Van Houwelingen JC, Le Cessie S. 1990 Predictive value of statistical models. *Stat Med* 9(11): 1303-1325

6. Harrell FEJ, Lee KL, Mark DB. 1996 Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 15(4): 361-387

7. Harrell FEJ. 2001 *Regression modeling strategies*. New York: Springer-Verlag

8. Wallström A. 1999 KIGS: Structure and Organisation. In: Ranke MB, Wilton P, eds. Growth Hormone Therapy in KIGS - 10 Years' Experience. Heidelberg: Barth; 1-9

9. Tanner JM, Whitehouse RH, Takaishi M. 1966 Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children, 1965. II. *Arch Dis Child* 41(220): 613-635

10. Freeman JV, Cole TJ, Chinn S, Jones PR, White EM, Preece MA. 1995 Cross sectional stature and weight reference curves for the UK, 1990. *Arch Dis Child* 73(1): 17-24

11. Niklasson A, Ericson A, Fryer JG, Karlberg J, Lawrence C, Karlberg P. 1991 An update of the Swedish reference standards for weight, length and head circumference at birth for given gestational age (1977-1981). *Acta Paediatr Scand* 80(8-9): 756-762

12. Van Houwelingen JC. 2000 Validation, calibration, revision and combination of prognostic survival models. *Stat Med* 19(24): 3401-3415

13. Kleinbaum DG, Kupper LL, Muller KE, Nizam A. 1998 *Applied Regression Analysis and Other Multivariable Methods*. Pacific Grove: Duxbury Press

14. Altman DG. 1993 Construction of age-related reference centiles using absolute residuals. *Stat Med* 12(10): 917-924

15. Royston P, Altman DG. 1994 Regression using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling. *Appl Stat* 43(3): 429-467

16. Fredriks AM, Van Buuren S, Burgmeijer RJF, Meulmeester JF, Beuker RJ, Brugman E, Roede MJ, Verloove-Vanhorick SP, Wit JM. 2000 Continuing Positive Secular Growth Change in the Netherlands 1955-1997. Pediatr Res 47(3): 316-323

17. Van Wieringen JC, Wafelbakker F, Verbrugge HP, De Haas JH. 1971 Growth diagrams 1965 Netherlands. Leiden/Groningen: Nederlands Instituut voor Praeventieve Geneeskunde/Wolters-Noordhoff

18. Altman DG, Royston P. 2000 What do we mean by validating a prognostic model? Stat Med 19(4): 453-473

19. Chatfield C. 1995 Model uncertainty, data mining and statistical Inference. J R Stat Soc A 158(3): 419-466

20. Roede MJ, Van Wieringen JC. 1985 Growth diagrams 1980: Netherlands third nation-wide survey. Tijdschr Soc Gezondheidsz 63(suppl):1-34

3

Validation of a calibrated prediction model for response to growth hormone treatment in an independent cohort

Abstract

Background

Prediction models, e.g. for prediction of response to growth hormone treatment, need validation in appropriate independent cohorts, comparing predicted and observed outcomes. In a previous validation of a model for predicting the first-year response to growth hormone treatment in children with idiopathic growth hormone deficiency, overfitting was observed. We modified the prediction formula and now report validation of this modified model.

Patients and Methods

The modified and original prediction models were applied to a group of patients selected from Lilly's GeNeSIS database using the same inclusion and exclusion criteria as for the original model. For both prediction methods, observed first-year height velocity was plotted vs. predicted height velocity in a calibration plot. For a valid prediction, the regression line should correspond to the line of identity (observed outcome is equal to predicted outcome); the regression lines for each prediction model were tested for significant differences from this line of identity.

Results

The number of patients fulfilling the criteria was 226. The regression line in the calibration plot of the modified model was not significantly different from the line of identity ($p = 0.43$), in contrast to the original model ($p < 0.001$). For the modified model the mean (SD) prediction error was -0.11 (2.05) cm/yr and for the original model 0.28 (2.11) cm/yr.

Conclusion

The modified prediction method, obtained after calibration of the original model, performs well in an independent patient sample and gives more accurate predictions than the original model.

Introduction

It is widely recognized that the validity of a clinical prediction model should be evaluated by applying the model to an independent group of patients. This group should be selected using the same inclusion criteria as used for the development of the model. Observed and predicted outcomes of this validation group should be compared. Often, the main item of the validation is to test whether the mean of the prediction errors (differences between observed and predicted outcomes) is not significantly different from zero (1–3). However, this procedure only tests whether there is a systematic error in the predictions but does not evaluate the individual prediction errors, which are very relevant in clinical practice. To examine whether the observed outcomes are consistent with the predicted outcomes, a simple and clear approach is to produce a scatter plot of observed vs. predicted outcomes (4). If the data are close to the line of identity, the prediction model is considered valid, whereas if the regression line is significantly different from the line of identity, the model needs further calibration.

In 2003, we used this procedure to validate the model developed by Ranke et al. (1) to predict height velocity (HV) of prepubertal children with idiopathic growth hormone (GH) deficiency in the first year of GH treatment, based on data of the Kabi International Growth Study database. We observed overfitting in this validation, which resulted in modification of the model that should lead to less prediction errors (5). The aim of this study was to validate this modified prediction formula using a new group of patients and compare results using both the original and modified models.

Patients and Methods

Patients were selected from the Genetics and Neuroendocrinology of Short Stature International Study (GeNeSIS) database of Eli Lilly and Company, which contains data of children treated with Humatrope. The same inclusion and exclusion criteria were applied as used for development of the original model (1), namely: idiopathic GH deficiency, GH treatment throughout at least 1 year, 6 or 7 GH injections per week, age at start between 2 and 10 years for girls and between 2 and 12 years for boys, prepubertal during the first year of treatment, no use of other medication that may influence growth and normal size at birth.

The prediction model is based on data known at start of the GH treatment. The predictions using the original model were calculated with the formula:

$$predicted\ HV\ (cm/yr) = 14.55 - 1.37 * maximum\ GH\ response\ (ln;\ \mu g/L)$$
$$- 0.32 * age\ (yr) + 0.32 * birth\ weight\ SDS + 1.62 * GH\ dose\ (ln;\ IU/kg \cdot week)$$
$$- 0.40 * (height\ SDS - midparental\ height\ SDS) + 0.29 * weight\ SDS.$$

The predicted values using the modified prediction formula are calculated as:

$$modified\ prediction = 2.153 + 0.808 * original\ prediction$$

resulting in:

$$predicted\ HV\ (cm/yr) = 13.91 - 1.12 * maximum\ GH\ response\ (ln;\ \mu g/L)$$
$$- 0.26 * age\ (yr) + 0.26 * birth\ weight\ SDS + 1.31 * GH\ dose\ (ln;\ IU/kg \cdot week)$$
$$- 0.32 * (height\ SDS - midparental\ height\ SDS) + 0.23 * weight\ SDS.$$

Prediction errors with each method were calculated as observed HV - predicted HV. For both prediction methods, observed first-year HV was plotted vs. predicted HV in calibration plots. The resulting regression lines were tested for significant differences from the line of identity using an F-test.

Results

There were 226 patients fulfilling the criteria for inclusion in the study. Characteristics of this group are given in table 1. For the modified prediction method the mean (SD) of the prediction errors was –0.11 (2.05) cm/yr while for the original prediction formula it was 0.28 (2.11) cm/yr.

Figure 1 shows the calibration plots for the modified and the original models. In each plot, the dotted line is the line of identity and the solid line is the regression line. For the modified prediction method the regression line was not significantly different from the line of identity ($p = 0.43$). The equation of the regression line was

$$HV_{obs} = 0.693 + 0.916 * HV_{mod\ pred}$$

where $HV_{mod\ pred}$ represents the predictions using the modified model. For the original model, the equation was

$$HV_{obs} = 2.665 + 0.740 * HV_{orig\ pred}$$

where $HV_{orig\ pred}$ represents the predictions using the original model. This regression line was significantly different from the line of identity ($p < 0.001$).

Table 1. Characteristics of the validation group (n = 226) and of the original cohort, from which the model was derived (n = 593)

	Validation cohort		Modelling cohort	
	Mean	SD	Mean	SD
Age at start (yr)	7.6	2.6	7.3	2.4
Height SD score	-2.5	0.9	-2.6	0.8
Weight SD score	-2.3	1.3	-2.2	1.3
Mid-parental height (MPH) SD score	-0.6	1.3	-0.6	1.0
Height SD score - MPH SD score	-1.9	1.4	-1.9	1.4
Birth weight SD score	-0.5	1.0	-0.5	0.9
Maximum GH peak (μg/L) in stimulation tests	5.9	2.7	5.6	2.8
GH dose (IU/kg·week)	0.6	0.2	0.6	0.2
Height velocity in first year (cm/yr)	9.5	2.6	9.2	2.3
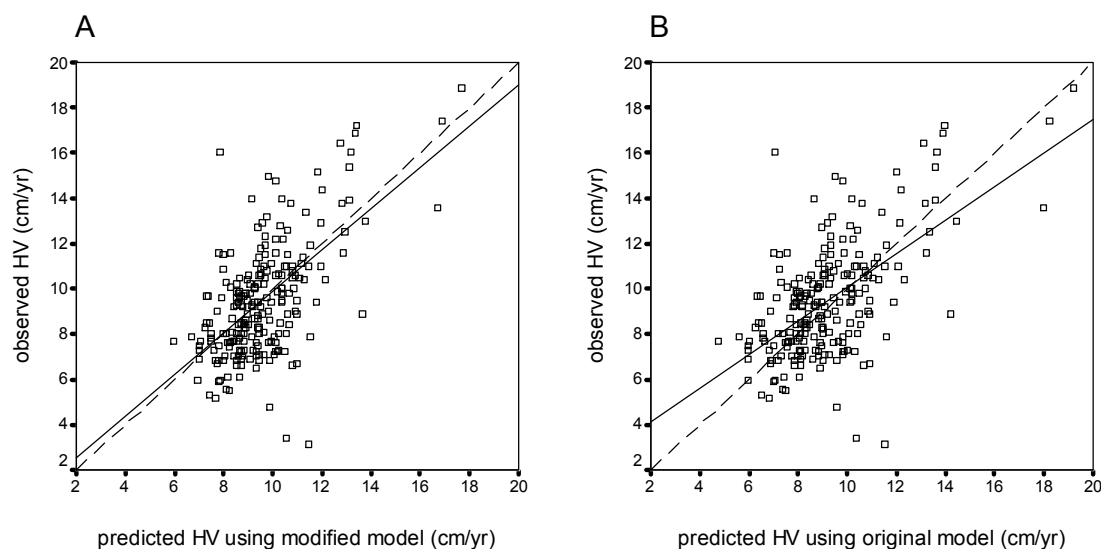
Figure 1. Calibration plot of observed and predicted height velocities using the modified model (A) and the original model (B). Solid lines represent the regression lines, dotted lines the line of identity.

Discussion

The study presented in this paper is a continuation to our earlier work of validation and calibration of a prediction model (5), modified from a previously published model (1), namely checking the validity of our modified model. The results showed that the predictions of this modified model are more accurate than the predictions of the original model.

Although a substantial number of prediction models have been developed (1, 3, 6–10), the procedure of validation of these models needs to be improved. If a sufficiently large number of patients is obtained that fulfil the inclusion criteria, and for whom all determinants used in the model are known and the outcome is observed, the best information can be obtained by producing a calibration plot. This plot shows the relation between observed and predicted values and, for a valid model, the points should be close to the line of identity.

Predictions often tend to be too extreme near the boundaries of the range of predicted values. Patients with low observed outcomes generally have predictions that are too low, whereas the opposite is found for patients with high observed outcomes; we observed this effect when validating the prediction model developed by Ranke et al. (1). This phenomenon was not due to an error in the selection of predictors or in the estimation of the coefficients, but is a common finding (11).

Using the results from a validation of the model with new data, one can adjust for this overfitting (4). Our present study showed that this approach is practical, simple and reliable, and leads to a model with high value for clinical practice.

References

1. Ranke MB, Lindberg A, Chatelain P, Wilton P, Cutfield W, Albertsson-Wikland K, 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. KIGS International Board. Kabi Pharmacia International Growth Study. J Clin Endocrinol Metab 1999;84:1174– 1183.

2. Vosahlo J, Zidek T, Lebl J, Riedl S, Frisch H: Validation of a mathematical model predicting the response to growth hormone treatment in prepubertal children with idiopathic growth hormone defi ciency. Horm Res 2004;61:143– 147.

3. Albertsson-Wikland K, Kristrom B, Rosberg S, Svensson B, Nierop AFM: Validated multivariate models predicting the growth response to GH treatment in individual short children with a broad range in GH secretion capacities. Pediatr Res 2000;48:475–484.

4. van Houwelingen JC: Validation, calibration, revision and combination of prognostic survival models. Stat Med 2000;19:3401–3415.

5. de Ridder MA, Stijnen T, Hokken-Koelega AC: Validation and calibration of the Kabi Pharmacia International Growth Study prediction model for children with idiopathic growth hormone deficiency. J Clin Endocrinol Metab 2003;88:1223–1227.

6. Ranke MB, Lindberg A, Chatelain P, Wilton P, Cutfield W, Albertsson-Wikland K, Price DA: Prediction of long-term response to recombinant human growth hormone in Turner syndrome: development and validation of mathematical models. KIGS International Board. Kabi International Growth Study. J Clin Endocrinol Metab 2000;85:4212–4218.

7. Ranke MB, Cutfield WS, Lindberg A, Cowell CT, Albertsson-Wikland K, Reiter EO, Wilton P, Price DA: A growth prediction model for short children born small for gestational age. J Pediatr Endocrinol Metab 2002;15(suppl 5):1273.

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

9. Schonau E, Westermann F, Rauch F, Stabrey A, Wassmer G, Keller E, Bramswig J, Blum WF: A new and accurate prediction model for growth response to growth hormone treatment in children with growth hormone defi ciency. Eur J Endocrinol 2001;144:13–20.

10. Spagnoli A, Spadoni GL, Boscherini B: Preliminary validation of a prediction model for the short-term growth response to growth hormone therapy in children with idiopathic short stature. Acta Paediatr 1996;417(suppl):66–68.

11. Harrell FEJ, Lee KL, Mark DB: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996; 15:361–387.

4

Prediction of adult height in growth hormone treated children with growth hormone deficiency

Abstract

Context

Several studies have searched for factors that significantly influence adult height (AH) of children with GH deficiency (GHD) who have been treated with biosynthetic GH, but a prediction model for AH has not yet been presented.

Objective

Our objective was to develop models for prediction of AH SDS, using information available at the start of GH treatment or after 1 year of treatment.

Design and Setting

For this retrospective study, data were collected from the National Registry of Growth Hormone Treatment in Children, which contained data of Dutch children treated with GH.

Patients/Intervention

Patients included males born before 1985 and females born before 1987 with either diagnosis of GHD (syndromes, tumors, and other diseases were excluded) or a maximal GH response during provocation tests of less than 11 ng/ml, treated with biosynthetic GH for at least 1 year. To be able to use the complete group of 342 children for the development of the models, multiple imputation was used for missing values.

Results

Each prediction model contained both target height SDS and current height SDS. The change in height SDS during the first year proved an important predictor for AH. In all models, addition of GH dose was not significant. The percent explained variance, after correction for overfitting, ranged from 37% (prepubertal children, prediction at start) to 60% (pubertal children, prediction after 1 year).

Conclusion

The presented prediction models give accurate predictions of AH for children with GHD at start and after 1 year of GH treatment. They are useful tools in the treatment of these children.

Introduction

Treatment with biosynthetic growth hormone (GH) is successful in improving adult height (AH) in children with GH deficiency (GHD). The reported mean AH SD scores range from -1.6 to -0.7, and the mean changes in height SDS during GH treatment from 1.1 to 2.0 (1-6).

Several prediction models have been developed for short-term growth response to GH treatment in children with GHD (7-9). For example Ranke et al. (8) developed models for prepubertal children predicting height velocity (HV) (in cm/yr) during the first, second, third and fourth years of treatment. In the models predicting HV during the second year or later, the HV in the previous year was the most prominent predictor (7, 8). Cole et al. (10) analysed first- and second-year growth response to GH treatment. They found that the maximum GH response during provocation tests was the most predictive factor for the first-year response, whereas the first-year response was much more important for the second-year response.

The following predictive factors for AH SDS were described: sex, birth weight SDS, age at start of GH treatment, height SDS at start for chronological age, height SDS at start for bone age (BA), weight SDS at start, target height (TH) SDS or mid-parental height SDS, maximum GH response during provocation tests, presence of multiple pituitary hormone deficiencies (MPHD), BA delay at start, pubertal stage at start, age at onset of puberty, height SDS at onset of puberty, HV in the first year of treatment, total GH dose, number of GH injections per week, duration of treatment and completion of treatment until AH (1-4, 6, 11-14).

In three studies, a regression model was developed for AH SDS or change in height SDS during GH treatment in a group of children with GHD treated with biosynthetic GH (1, 2, 6). These studies, however, also included patient characteristics during long-term follow-up, for example height SDS at onset of puberty, duration of treatment, and mean GH dose during treatment.

Therefore, until now, an accurate model for the prediction of AH SDS at start or after 1 year of GH treatment has not been developed.

In this study, we developed models for the prediction of AH SDS. We will first present a model using only information available at start of the GH treatment and then second a model using information available after 1 year of treatment, for prepubertal as well as pubertal children.

Patients and methods

Patients

We used data from the National Registry of Growth Hormone Treatment in Children by the Dutch Growth Foundation, which contains data of more than 2200 Dutch children treated with GH. Registration started in 1992 and has been obligatory since 1997. We selected males born before 1985 and females before 1987, to ensure a representative group with AH. Other selection criteria used were diagnosis of GHD or a maximum GH response during provocation tests of less than 11 ng/ml and treatment with biosynthetic GH for at least 1 year. Children with syndromes, tumors or other diseases were excluded (Figure 1).
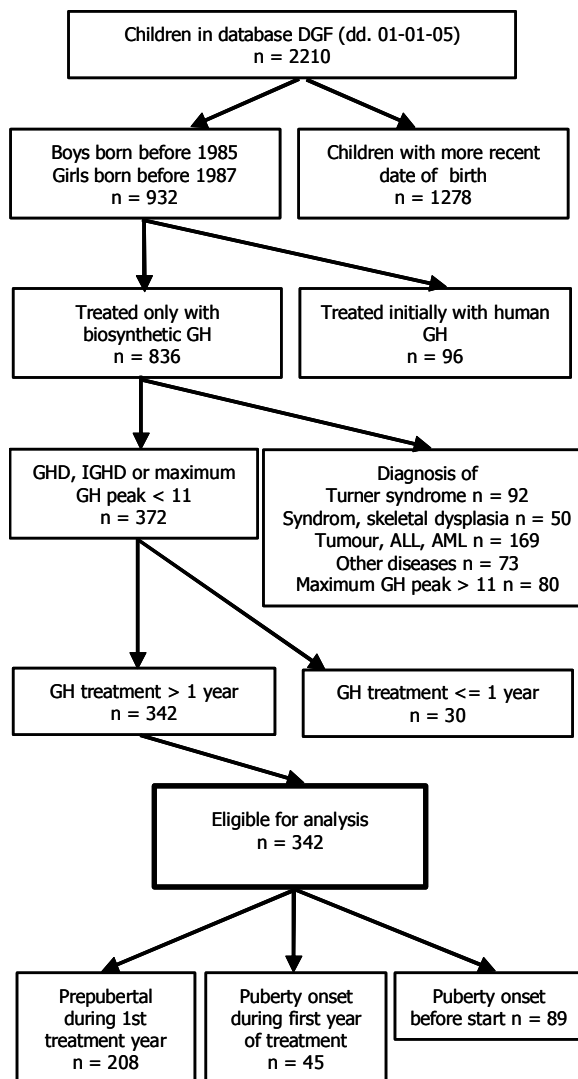
Figure 1. Description of cohort

For the development of the prediction models for application at start of GH treatment, the group was divided into two subgroups, one including prepubertal children and one including pubertal children. For the models to be applied after 1 year of GH treatment, the pubertal stage at that moment determined the prepubertal and pubertal subgroups.

Outcome and potential determinants

AH was defined as the height reached when growth velocity was less than 2 cm/yr and age above 14 years. AH SDS was calculated using references for Dutch adults, i.e. a mean (SD) of 184.0 (7.1) cm for males and 170.6 (6.5) cm for females (15).

The potential determinants were as follows 1. initial characteristics included sex, TH SDS and birth weight SDS (16), 2. characteristics at start of GH treatment included age, height SDS, weight SDS (15), body mass index (BMI) SDS (17), BA and BA delay, maximum GH response during provocation tests, diagnosis (idiopathic GHD or GHD with abnormalities on pituitary magnetic resonance imaging), presence of MPHD, IGF-I SDS and starting dose of GH; and 3. additional characteristics after 1 year of treatment included height SDS and change in height SDS during the first year, weight SDS and change, BMI SDS and change, change in BA during the first year and mean GH dose during the first year.

First-year changes in height and weight SDS were corrected for the actual time between the two measurements (range 0.7 - 1.3 year).

Eighty-seven percent of the BA were determined according to Greulich and Pyle. When the BA was determined according to the TW2 or RUS method, it was converted to Greulich and Pyle estimations (18). BA delay was computed as chronological age minus BA. Bone maturation during the first year was computed as the ratio of the change in BA and the exact difference between the chronological ages at time of the BA measurements.

In 80% of patients an arginine test was used to determine the maximum GH response, in 40% L-dopa and in 24% clonidine. Eighty-two percent of children had GHD defined as a failure to increase serum GH levels above 11 ng/ml in two or more tests, whereas 18% had only one test.

Eighty percent of the IGF-I measurements were centrally performed in a laboratory with published reference values (19). For the remaining 20% the laboratory-specific reference values for IGF-I were used to calculate the SDS. IGF-I measurements after 1 year of treatment were not used, because these were very scarce (11%).

Multiple imputation

It is unavoidable that in a registry database some data are missing for some patients. For example birth data and the height of parents can be missing for adopted children and BA data are not complete because these measurements were not always performed in former years. Using only cases with complete data for all the potential determinants would result in a sample much smaller and likely not a representative group of the population.

To develop the prediction models on the complete group, we used multiple imputation for missing values in the outcome or in the potential determinants (20, 21). For each missing variable, a value was imputed, using the relations between the variables in the data set. Because an imputed value does not have the same accuracy as an observed value, the imputation procedure was executed five times to generate five completed data sets. In each data set, a different value was imputed, thus reflecting the uncertainty of the imputed value. For the multiple imputation we used the procedure SAS Proc MI (22), which assumes that the variables have a multivariate normal distribution. Variables with a non-normal distribution were transformed to normality during the imputation procedure.

Each step in the data analyses was performed on each imputed data set separately, and the results were combined (20, 21).

Truncation of extreme values of potential determinants and outcome

Extreme values of the outcome or determinants can highly influence the estimation of regression coefficients in a model. To avoid this, the 1% lowest values of all continuous determinants were truncated to the first percentile and the 1% highest values to the 99th percentile.

Development of prediction model

For the development of the prediction model, we used forward selection with an inclusion criterion of p = 0.05. After the selection of the predictor variables, for each continuous predictor, it was tested whether adding the quadratic term or another transformation of the variable improved the model significantly (23).

We tested possible interactions of age or BA at start with TH SDS and with change in height SDS during first year, if both main terms were selected in the model.

A possible relation between the predicted outcomes and the residuals was examined by evaluation of the scatter plots and by fitting the linear regression with the absolute value of the residuals as determinant and the predictions as outcome.

Internal validation

It is well-known that a prediction model suffers from over-optimism (24, 25). The predictive performance of a prediction model in other data sets than the set on which it is developed will be lower. The predicted values generated by the model will tend to be too extreme.

To remove the over-optimism of the derived model, we used bootstrap techniques (24, 26), shortly explained as follows. From the data set used for the development of a prediction model, a random sample was drawn, representing another but comparable data set. This sampling was done with replacement, so each subject could be selected several times, and consisted of the same number of subjects as the original data set. On this data set, called a bootstrap sample, a prediction model was developed, using the same procedure as for the development of the model in the original data set. The predictive performance of this model developed in the bootstrap sample was evaluated by calculating R^2, which gives the percentage of variance explained by the model, and by calculating the mean of the squared residuals. This evaluation was done in both the bootstrap sample and the original data set. The predictive performance is always better in the bootstrap sample, on which this model is developed, than in the original data set (higher R^2 and smaller residuals). The difference between these predictive performances is called the optimism. We generated 200 bootstrap samples and used the average optimism to correct the predictive performance of the original model (26). Furthermore, a linear regression was performed in each bootstrap sample with the observed values as outcome and the predicted values as determinant. The coefficient for the slope will usually be below one. This procedure is the same as for the validation of a model on external data (25). The mean of the 200 slopes was used as shrinkage factor for the estimated regression coefficients in the original derived prediction model. The intercept was adjusted to get the same mean predicted value (27). This resulted in a calibrated model that provides predictions less extreme than the predictions from the original model. The predictions calculated with the calibrated model will be accurate in new patients with GHD.

Results

The study group consisted of 342 children, of whom 208 were prepubertal for at least 1 year after the start of GH treatment, and 89 were pubertal at start of treatment. In 45 children puberty started during the first year of treatment. Table 1

shows the characteristics of the study group and for each variable the percentage of patients with a missing value.

Table 1. Characteristics of the study group

	Prepubertal at start and during first year of treatment (n = 208)	Start of puberty During first year of treatment (n = 45)	Before start of treatment (n = 89)	Percentage with missing value
Male (%)	60	44	45	0
Birth weight SDS	-0.73 ± 1.31	-0.80 ± 1.13	-0.54 ± 1.37	14
TH SDS	-0.97 ± 0.96	-1.22 ± 0.91	-0.90 ± 0.87	5
At start				
Age (yr)	9.0 ± 3.3	12.8 ± 1.9	14.0 ± 2.0	0
Maximum GH peak (ng/ml)	4.5 ± 2.9	5.7 ± 3.0	5.2 ± 2.8	6
IGF-I SDS	-3.67 ± 2.76	-3.54 ± 3.44	-2.93 ± 3.18	37
Idiopathic GHD (%)	83	99	88	0
Presence of MPHD (%)	41	24	30	0
Height SDS	-3.40 ± 1.01	-3.32 ± 0.97	-2.92 ± 1.14	0
Weight SDS	-2.54 ± 1.48	-2.01 ± 1.44	-1.79 ± 1.68	0.3
BMI SDS	-0.35 ± 1.21	-0.06 ± 1.20	-0.10 ± 1.36	0.3
BA (yr)	6.5 ± 3.0	10.4 ± 2.0	12.0 ± 1.6	11
BA delay (yr)	2.6 ± 1.5	2.4 ± 1.3	2.0 ± 1.5	11
GH dose (mg/m^2·day)	0.71 ± 0.24	0.72 ± 0.26	0.73 ± 0.26	1
After first year				
Height SDS	-2.69 ± 0.95	-2.88 ± 1.05	-2.34 ± 1.13	0
Weight SDS	-2.08 ± 1.39	-1.74 ± 1.51	-1.44 ± 1.59	0.3
BMI SDS	-0.51 ± 1.24	-0.12 ± 1.28	-0.08 ± 1.34	0.3
Change in height SDS	0.71 ± 0.51	0.44 ± 0.36	0.58 ± 0.38	0
Change in weight SDS	0.45 ± 0.51	0.27 ± 0.49	0.35 ± 0.47	0.3
Change in BMI SDS	-0.17 ± 0.57	-0.06 ± 0.49	0.01 ± 0.46	0.3
Mean GH dose (mg/m^2·day)	0.71 ± 0.23	0.75 ± 0.27	0.75 ± 0.27	0.6
Adult height				
Duration GH treatment (yr)	7.9 ± 3.3	3.9 ± 1.3	3.3 ± 1.2	4
Mean GH dose (mg/m^2·day)	0.77 ± 0.21	0.80 ± 0.27	0.80 ± 0.24	8
AH SDS	-1.71 ± 0.91	-2.02 ± 1.08	-1.68 ± 0.94	20
Change in height SDS	1.72 ± 1.10	1.36 ± 0.81	1.18 ± 1.16	20
AH SDS – TH SDS	-0.74 ± 1.04	-0.68 ± 0.75	-0.85 ± 0.88	22

Results expressed as mean ± SD or percentage

Prediction models at start of GH treatment

For the prediction of AH SDS using the characteristics available at the start of treatment (Start model) the final model for prepubertal patients included six variables. Predictors with a positive effect were height SDS at start, TH SDS, female gender and presence of MPHD, whereas maximum GH response during provocation tests and BA at start had a negative effect. The relation between height SDS at start and AH SDS (corrected for the other predictors) was quadratic. Starting dose of GH was not a significant predictor. In Table 2, the estimated coefficients of the model are given. The percentage explained variance (R^2*100%) of the derived model was 43%. The optimism estimated by bootstrapping was 6%, so the corrected percentage is 37%. Corrected for optimism, the residual SD was 0.84. The estimated shrinkage factor for the correction of the regression coefficients was 0.94. The final prediction formula is given in Table 2.

Table 2. Results of the final models for prediction of AH at start of treatment (Start model)

Predictor variable	Prepubertal (n = 253)				Pubertal (n = 89)			
	Est. coef.	SE	P	Partial r^2	Est. coef.	SE	P	Partial r^2
Intercept	1.399	0.502	0.006		-0.645	0.226	0.006	
Height SDS at start	1.092	0.260	<0.0001	0.239	0.456	0.080	<0.0001	0.281
(Height SDS at start)^2	0.082	0.035	0.02					
TH SDS	0.282	0.060	<0.0001	0.086	0.428	0.100	<0.0001	0.188
Max GH peak (ng/ml; ln)	-0.158	0.064	0.02	0.022				
Gender[a]	0.278	0.108	0.01	0.022				
MPHD[b]	0.323	0.124	0.01	0.023				
BA at start (yr)	-0.051	0.017	0.003	0.031				
BA delay at start (yr)					0.265	0.058	<0.0001	0.213
R^2 corrected for optimism (original)	0.37 (0.43)				0.41 (0.53)			
Residual SD corrected (original)	0.84 (0.79)				0.83[c] (0.72[d])			
Shrinkage factor	0.94				0.91			

Prediction formulas after bootstrap-correction are as follows:
for the prepubertal group, AH SDS = 1.186 + 1.021*H SDS$_{start}$ + 0.077*H SDS$_{start}$^{2} + 0.264*TH SDS - 0.148*ln(max GH) + 0.260*gender + 0.302*MPHD - 0.047*BA;
for the pubertal group, AH SDS = -0.746 + 0.416*H SDS$_{start}$ + 0.391*TH SDS + 0.242*BA delay

[a]Male = 0, Female = 1
[b]No = 0, Yes = 1
[c]Accounting for relation with predicted value: corrected residual SD = $\sqrt{((0.38 - 0.19*\text{predicted value})^2 + 0.17)}$
[d]Idem: original residual SD = 0.38 - 0.19*predicted value

For patients already pubertal at start, age at onset of puberty is also a potential determinant. For boys, we reduced the age at onset by 0.8 year, according to the difference in mean age at onset of puberty between boys and girls in the Dutch population, as found by Fredriks et al. (15). For this group the Start model included three variables; height SDS at start, TH SDS and BA delay at start, all with positive effect (Table 2). Again, starting dose of GH was not in the final model. The percentage explained variance was 53% and after correction for optimism, 41%. For this prediction model, there was a significant relation between the predicted values and the residual SD. The residual SD was decreasing with increasing predicted value, according to the equation

$$residual\ SD = 0.38 - 0.19\ *predicted\ value.$$

This residual SD has to be adjusted using the optimism estimated by bootstrapping (see note at Table 2). The estimated shrinkage factor for the correction of the regression coefficients was 0.91.

Prediction models after 1 year of treatment

Using the characteristics available after 1 year of treatment (First-year model), the prediction model for prepubertal patients included height SDS after the first year (quadratic relation), TH SDS, female gender, presence of MPHD, BA delay at start and change in height SDS during the first year, all with a positive effect. The estimated coefficients are given in Table 3. As in the Start model, addition of starting dose of GH was not significant, nor was mean GH dose during the first year. If the latter was added to the final model, its influence on AH SDS was negative, the estimated coefficient being -0.35 (p = 0.11), whereas the coefficients of the other predictors did not change substantially (all changes < 10%). The percentage explained variance of the final model was 51% and after correction for optimism, 43%. The residual SD after correction for optimism was 0.76. The estimated shrinkage factor for the correction of the regression coefficients was 0.94.

The predictors in the First-year model for children who are pubertal after 1 year of GH treatment were height SDS after the first year, TH SDS, BA delay at start, and change in height SDS during the first year. The mean GH dose during the first year was not a significant predictor. The explained variance was 66% and after bootstrapping was reduced to 60%. Corrected for optimism, the residual SD was 0.69. The estimated shrinkage factor for the correction of the regression coefficients was 0.95.

Table 3. Results of the final models for prediction of AH after one 1 of GH treatment (First-year model)

Predictor variable	Prepubertal (n = 208)				Pubertal (n = 134)			
	Est. Coeff.	SE	P	Partial r^2	Est. Coeff.	SE	P	Partial r^2
Intercept	0.075	0.330	0.82		-0.866	0.189	<0.0001	
Height SDS after first year	1.250	0.213	<0.0001	0.336	0.527	0.064	<0.0001	0.385
(Height SDS after first year)^2	0.114	0.035	0.001					
TH SDS	0.200	0.057	0.0006	0.065	0.347	0.074	<0.0001	0.162
Gender[a]	0.348	0.106	0.001	0.054				
MPHD[b]	0.309	0.107	0.004	0.043				
BA delay at start (yr)	0.100	0.038	0.008	0.039	0.164	0.044	0.0003	0.103
Δ height SDS in first year	0.308	0.105	0.004	0.044	0.500	0.163	0.003	0.070
R^2 corrected for optimism (original)	0.43 (0.51)				0.60 (0.66)			
Residual SD corrected (original)	0.76 (0.69)				0.69 (0.62)			
Shrinkage factor	0.94				0.95			

Prediction formulas after bootstrap-correction are as follows:
for the prepubertal group, AH SDS = -0.049 + 1.169*H SDS$_{1yr}$ + 0.107*H SDS$_{1yr}$^{2} + 0.187*TH SDS + 0.325*gender + 0.289*MPHD + 0.094*BA delay$_{start}$ + 0.288* Δ H SDS$_{1yr}$;
for the pubertal group, AH SDS = -0.915 + 0.502*H SDS$_{1yr}$ + 0.331*TH SDS + 0.156*BA delay + 0.477*Δ H SDS$_{1yr}$

[a]Male = 0, Female = 1
[b]No = 0, Yes = 1

In Table 4 and Figure 2, examples are given of the use of the prediction models for prepubertal children. Without any model, we would predict the AH SDS of each individual prepubertal child with GHD and GH treatment as -1.77 (being the mean AH SDS of the children prepubertal at start) with a 95% prediction interval of -3.61 to 0.07. Table 4 gives the characteristics and the predictions of a (hypothetical) child with relatively positive prospects (child 1) and a child with less favourable characteristics (child 2). In Figure 2, the distributions of the predictions with the First-year model are plotted for the children in this example. For child 1, the predicted probability of reaching an AH SDS above -2 is 85%, and for child 2, this is only 11%.

Table 4. Examples of predictions for prepubertal children

	Child 1	Child 2
Characteristics		
Height SDS at start	-3.0	-4.0
TH SDS	-1.0	-2.0
Maximum GH peak (ng/ml)	2.0	7.0
Gender	Girl	Boy
MPHD	Yes	No
BA at start (yr)	6.0	8.0
BA delay at start (yr)	2.6	0.0
Height SDS after first year	-2.2	-3.3
Change in height SDS in first year	0.8	0.7
Predictions (95% prediction interval)[*]		
Without any model	-1.77 (-3.61 to 0.07)	-1.77 (-3.61 to 0.07)
Start model	-1.27 (-2.92 to 0.38)	-2.86 (-4.50 to -1.21)
First-year model	-1.20 (-2.69 to 0.28)	-2.92 (-4.41 to -1.43)

[*]The prediction intervals are calculated using the relevant SD, i.e. 0.94 without model, 0.84 for the Start model and 0.76 for the First-year model.

Example of calculation, child 1, start model: predicted AHSDS = $1.186 + 1.021*(-3) + 0.077*(-3)^2 + .264*(-1) - 0.148*\ln(2) + 0.260*1 + 0.302*1 - 0.047*6 = -1.27$

Child 1

probability AH SDS > -2 = 0.85

predicted value

95% prediction interval

Child 2

probability AH SDS > -2 = 0.11

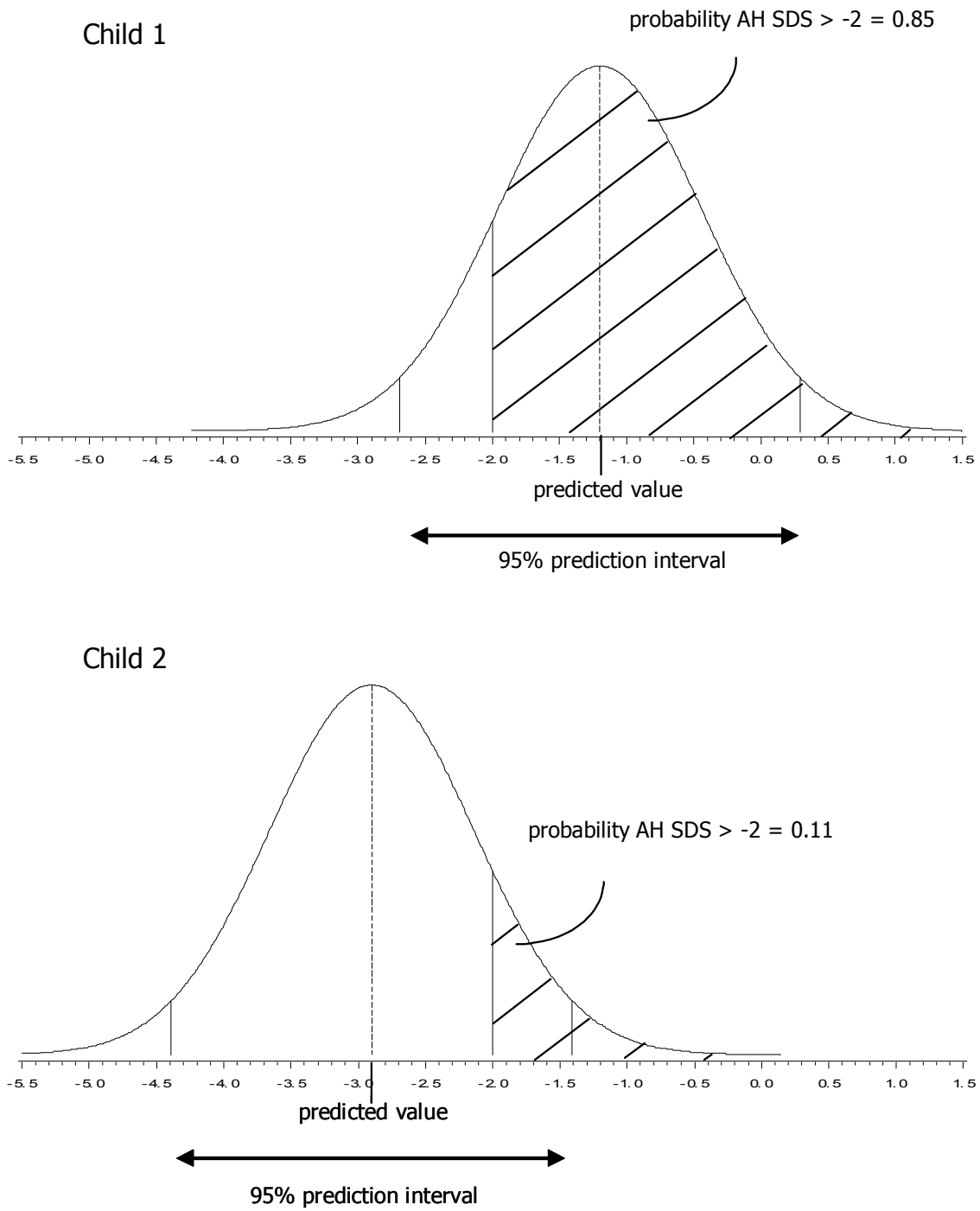predicted value

95% prediction interval

Figure 2. Examples of the distribution of the predicted value for AH SDS after 1 year of treatment. For the characteristics of the children used in these examples, see text.

Discussion

In the present study, we have developed prediction models for AH SDS for prepubertal and pubertal children with GHD treated with GH, according to state-of-the-art statistical methods. This includes dealing with missing data, investigation of more flexible relations between continuous determinants and the outcome and correction for over-optimism (24).

A prediction model for AH SDS is a useful tool for the clinician to become informed about the prospective long-term results of GH treatment. A prediction is desired before start of treatment. Application of the prediction model gives a patient his or her individual predicted value with prediction interval. It can identify patients with high or low chance of benefit from GH treatment. After 1 year of treatment, the expectations can be refined, which is useful for decisions about the continuation of GH treatment.

The prediction models presented in this paper explain percentages of variance in AH SDS from 37% (prepubertal group, Start model) to 60% (pubertal group, First-year model). The outcome is usually several years ahead from the moment of prediction. During childhood, many factors may influence the growth of a child, so we could expect that a substantial part of the variance remains unexplained.

As expected, all models show height SDS (at start or after 1 year of GH treatment) and TH SDS as most important predictive factors. A positive effect of female gender was found, in line with Carel et al. (1). They attributed this to sex-dependent differences in pubertal age. It is also possible that girls have a better compliance than boys. Children with MPHD had a more favourable outcome compared to children with isolated GHD, as previously reported by Reiter et al. (6). The negative coefficient for BA in the model for prepubertal children reflects that start of GH treatment at a younger age gives a higher growth response. It appeared that BA is a more informative predictor than chronological age. The positive coefficients for BA delay in the other models reflect that children with delayed BA have more growth potential.

Maximum response to GH provocation tests is included in the Start model for prepubertal children (negative effect) but is not significant anymore in the First-year model. A similar finding was reported by Cole et al.(10). The First-year models include change in height SDS during the first year, which is in line with several other studies (2, 3, 6, 11).

In none of the models GH dose was selected as significant predictor variable. Notably, this has also been reported by others (1, 6). For clinical practice, it might have been desirable if the prediction models could lend support in finding the optimal GH dose. A practical and logical question is what would be the predicted AH SDS if a higher or lower dose than the standard dose is prescribed? A significant positive effect of GH dose on short-term growth response is often found (8, 28, 29), but the dose-effect on long-term response is less established (3, 6, 30). For the patients in our data set, the GH dose at start and during treatment was assessed by the clinician based on unknown criteria. In 80% of the patients, the mean GH dose during treatment was in the range 0.5 - 1.0 mg/m^2·day (median 0.72 mg/m^2·day). It is possible that higher initial doses were prescribed to patients with supposed worse prospects, like children close to or entering puberty, children with only mild GHD or children with very small parents. During treatment, a change of GH dose might have been guided by the obtained growth response. Our data are therefore not suitable to estimate the effect of GH dose on AH SDS. This would only be possible if the dosages given were randomly assigned and had remained unchanged during treatment, as in randomised controlled trials.

One of the arguments for developing the First-year models was to investigate if such a model could provide a criterion for the first-year growth response needed for an AH in the desired (normal) range. Indeed, our study shows that change in height SDS during the first year of treatment is highly related to the AH attainment. Because a positive correlation between GH dose and first-year response is well established, one might tend to give a high GH dose in order to increase the first-year response, with the idea that this will subsequently increase AH. However, we found an inverse relation between first-year GH dose and AH, although this did not reach significance. This means that a first-year growth response obtained by a high GH dose gives a lower predicted AH than the same response obtained by a low GH dose.

We performed internal validation of the prediction models by bootstrapping. This results in models corrected for overfitting. It is not necessary to validate or calibrate the models with an external validation. Applying the models to data of an independent cohort is still interesting. In conclusion, the prediction models presented in this study can be a useful tool for decisions about GH treatment of children with GHD.

References

1. Carel JC, Ecosse E, Nicolino M, Tauber M, Leger J, Cabrol S, Bastie-Sigeac I, Chaussain JL, Coste J. Adult height after long term treatment with recombinant growth hormone for idiopathic isolated growth hormone deficiency: observational follow up study of the French population based registry. Bmj 2002;325(7355):70.

2. Thomas M, Massa G, Bourguignon JP, Craen M, De Schepper J, de Zegher F, Dooms L, Du Caju M, Francois I, Heinrichs C, Malvaux P, Rooman R, Thiry-Counson G, Vandeweghe M, Maes M. Final height in children with idiopathic growth hormone deficiency treated with recombinant human growth hormone: the Belgian experience. Horm Res 2001;55(2):88-94.

3. Cutfield W, Lindberg A, Albertsson Wikland K, Chatelain P, Ranke MB, Wilton P. Final height in idiopathic growth hormone deficiency: the KIGS experience. KIGS International Board. Acta Paediatr Suppl 1999;88(428):72-5.

4. August GP, Julius JR, Blethen SL. Adult height in children with growth hormone deficiency who are treated with biosynthetic growth hormone: the National Cooperative Growth Study experience. Pediatrics 1998;102(2 Pt 3):512-6.

5. Bramswig JH, Schlosser H, Kiese K. Final height in children with growth hormone deficiency. Horm Res 1995;43(4):126-8.

6. Reiter EO, Price DA, Wilton P, Albertsson-Wikland K, Ranke MB. Effect of Growth Hormone (GH) Treatment on the Final Height of 1258 Patients with Idiopathic GH Deficiency: Analysis of a Large International Database. J Clin Endocrinol Metab 2006.

7. Schonau E, Westermann F, Rauch F, Stabrey A, Wassmer G, Keller E, Bramswig J, Blum WF. A new and accurate prediction model for growth response to growth hormone treatment in children with growth hormone deficiency. Eur J Endocrinol 2001;144(1):13-20.

8. Ranke MB, Lindberg A, Chatelain P, Wilton P, Cutfield W, Albertsson-Wikland K, 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. KIGS International Board. Kabi Pharmacia International Growth Study. J Clin Endocrinol Metab 1999;84(4):1174-83.

9. Albertsson-Wikland K, Kristrom B, Rosberg S, Svensson B, Nierop AFM. Validated Multivariate Models Predicting the Growth Response to GH Treatment in Individual Short Children with a Broad Range in GH Secretion Capacities. Pediatric Research 2000;48(4):475-84.

10. Cole TJ, Hindmarsh PC, Dunger DB. Growth hormone (GH) provocation tests and the response to GH treatment in GH deficiency. Arch Dis Child 2004;89(11):1024-7.

11. Blethen SL, Baptista J, Kuntze J, Foley T, LaFranchi S, Johanson A. Adult height in growth hormone (GH)-deficient children treated with biosynthetic GH. The Genentech Growth Study Group. J Clin Endocrinol Metab 1997;82(2):418-20.

12. Cacciari E, Cicognani A, Pirazzoli P, Zucchini S, Salardi S, Balsamo A, Cassio A, Pasini A, Carla G, Tassinari D, Gualandi S. Final height of patients treated for isolated GH deficiency: examination of 83 patients. Eur J Endocrinol 1997;137(1):53-60.

13. Coste J, Letrait M, Carel JC, Tresca JP, Chatelain P, Rochiccioli P, Chaussain JL, Job JC. Long-term results of growth hormone treatment in France in children of short stature: population, register based study. Bmj 1997;315(7110):708-13.

14. Rikken B, Massa GG, Wit JM. Final height in a large cohort of Dutch patients with growth hormone deficiency treated with growth hormone. Dutch Growth Hormone Working Group. Horm Res 1995;43(4):135-7.

15. Fredriks AM, van Buuren S, Burgmeijer RJ, Meulmeester JF, Beuker RJ, Brugman E, Roede MJ, Verloove-Vanhorick SP, Wit JM. Continuing positive secular growth change in The Netherlands 1955-1997. Pediatr Res 2000;47(3):316-23.

16. Niklasson A, Ericson A, Fryer JG, Karlberg J, Lawrence C, Karlberg P. An update of the Swedish reference standards for weight, length and head circumference at birth for given gestational age (1977-1981). Acta Paediatr Scand 1991;80(8-9):756-62.

17. Fredriks AM, van Buuren S, Wit JM, Verloove-Vanhorick SP. Body index measurements in 1996-7 compared with 1980. Arch Dis Child 2000;82(2):107-12.

18. Wit JM, Rekers-Mombarg LT, Cutler GB, Crowe B, Beck TJ, Roberts K, Gill A, Chaussain JL, Frisch H, Yturriaga R, Attanasio AF. Growth hormone (GH) treatment to final height in children with idiopathic short stature: evidence for a dose effect. J Pediatr 2005;146(1):45-53.

19. Rikken B, van Doorn J, Ringeling A, Van den Brande JL, Massa G, Wit JM. Plasma levels of insulin-like growth factor (IGF)-I, IGF-II and IGF-binding protein-3 in the evaluation of childhood growth hormone deficiency. Horm Res 1998;50(3):166-76.

20. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York; 1987.

21. Schäfer JL. Analysis of Incomplete Multivariate Data. New York; 1997.

22. Inc. SI. SAS Online Doc 9.1.3. Cary, NC: SAS Institute Inc.; 2006.

23. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. Int J Epidemiol 1999;28(5):964-74.

24. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15(4):361-87.

25. de Ridder MA, Stijnen T, Hokken-Koelega AC. Validation and calibration of the Kabi Pharmacia International Growth Study prediction model for children with idiopathic growth hormone deficiency. J Clin Endocrinol Metab 2003;88(3):1223-7.

26. Efron B. TR. An introduction to the bootstrap. New York; 1993.

27. Van Houwelingen JC, Le Cessie S. Predictive value of statistical models. Stat Med 1990;9(11):1303-25.

28. Cohen P, Bright GM, Rogol AD, Kappelgaard AM, Rosenfeld RG. Effects of dose and gender on the growth and growth factor response to GH in GH-deficient children: implications for efficacy and safety. J Clin Endocrinol Metab 2002;87(1):90-8.

29. De Muinck Keizer-Schrama S, Rikken B, Hokken-Koelega A, Wit JM, Drop S. Comparative effect of two doses of growth hormone for growth hormone deficiency. The Dutch Growth Hormone Working Group. Arch Dis Child 1994;71(1):12-8.

30. Wit JM, Kamp GA, Rikken B. Spontaneous growth and response to growth hormone treatment in children with growth hormone deficiency and idiopathic short stature. Pediatr Res 1996;39(2):295-302.

5

Prediction model for adult height of SGA children at start of growth hormone treatment

Abstract

Context

GH treatment is approved for short children born SGA. The optimal dose is not yet established.

Objective

To develop a model for prediction of height at onset of puberty and of adult height (AH).

Design and Setting

Two GH studies in short SGA children.

Patients/Intervention

150 SGA children with height SDS < -2, age ≥ 3, no signs of catch-up growth, available height at onset of puberty and at least one year of GH treatment prior to onset of puberty. In one study, patients were randomly assigned to either 0.033 or 0.067 mg/kg·day, in the other study all received 0.033 mg/kg·day. In 71 children, AH was reached.

Results

Determinants positively related to height SDS at onset of puberty were: height SDS at start, target height (TH) SDS and GH dose, whereas age at start and female gender were negatively related. Positively related to AH SDS were: height SDS and chronological age minus bone age (CA-BA) at start, TH SDS and GH dose, whereas serum IGFBP-3 SDS at start was negatively related. There was a significant interaction between GH dose and IGFBP-3 SDS, indicating a smaller GH dose effect for higher levels of IGFBP-3. The final model explained 57% of the variance in height SDS at onset of puberty and 41% of AH SDS.

Conclusion

The prediction model for height SDS at onset of puberty and AH SDS of short SGA children treated with GH provides useful information about the expected long-term growth. Because GH dosage is one of the determinants, the model aids in determining the optimal GH dose for each child.

Introduction

Children born small for gestational age (SGA) with persistent short stature can be effectively and safely treated with growth hormone (GH) (1-6). Patient characteristics found to be related to short-term response were: chronological age (CA) or bone age (BA) at start, weight SDS at start and mid-parental height SDS (7, 8). For long-term response, TH SDS, height and weight SDS at start, CA-BA at start and maximum GH response to a stimulation test were identified as significant factors (4, 9).

In Europe, the recommended dose for the approved indication (EMEA, 2003) is 0.035 mg/kg·day. In the US, the FDA-approved indication describes a dose of 0.070 mg/kg·day. Several studies found a dose-effect on growth response during the first treatment years (7, 8, 10-15). For the long-term growth response, the dose-effect was smaller, but still significant (4, 5). The optimal dose for individual short SGA-children is not yet established. Some investigators stated that a dose of 0.033 mg/kg·day results in significant gains in long-term growth, with IGF-I levels in the normal range and at lower costs (16). Others argued that the lower dose might be sufficient for children without extremely short stature (above -3 SDS), but that shorter or older children might better start with a higher dose (≥ 0.050 mg/kg·day), with tapering of the dose per kg when the absolute GH dose (in milligrams) is maintained over the years (5).

In the present study, we developed a model to predict height at onset of puberty and AH for short children born SGA who will start GH treatment. Determinants were various baseline characteristics and GH dose. The predictions from this model can be used as information about the expected adult height for an individual child in order to decide on the prescribed GH dose.

Subjects and methods

Subjects

We used data from two GH trials in short SGA children (8, 17). These studies included children with birth length SDS for gestational age below -2.00 (18), height SDS for CA at start below -2.00 (19) and height velocity (HV) SDS for CA below zero (19, 20), to exclude children with spontaneous catch-up growth. In both studies, CA at the start had to be 3 years or older. Study I (8) included prepubertal children, defined as Tanner breast stage I for girls and a testicular volume <4 ml for boys (21), and age below 12 years in boys and below 10 years in girls. Study II (17)

comprised a prepubertal group, aged 3-8 years, and a group aged 8-14 years. Both studies excluded children who had had a complicated neonatal period, endocrine or metabolic disorders, chromosomal disorders, growth failure caused by other disorders (emotional deprivation, severe chronic illness, or chondrodysplasia) or syndromes (except for Silver-Russell syndrome), as well as children who used or had used drugs that could interfere with GH treatment.

In study I, children were randomly assigned to either 0.033 mg GH / kg·day (= 1 mg/m^2·day, n = 41) or 0.067 mg GH / kg·day (2 mg/m^2·day, n = 38). The prepubertal children in study II were randomised into either a GH group or a control group that remained untreated for three years and started treatment afterwards. In the older age group, all children were treated from start. The GH dose in study II was 0.033 mg/kg·day. The inclusion period for study I was between April 1991 and January 1993 and for study II between October 1996 and December 1998.

For the present analysis, we selected children who were prepubertal for at least one year after start of GH treatment and had started or completed puberty at the time of analysis (December 2006). Birth year for boys had to be prior to 1993 and for girls prior to 1994. AH was used of boys born prior to 1988 and girls born prior to 1989.

Assays

Serum IGF-I and IGFBP-3 levels were measured in one central laboratory, using a specific RIA (22). The intraassay coefficient of variation (CV) was 4% and the interassay CV was 6%. IGF-I and IGFBP-3 values were converted into SD scores (23).

Measurements

Standing height was measured every 3 months using a Harpenden stadiometer. The mean of four measurements was used for analysis. Heights were converted into height SD scores (19). AH SDS was computed using the reference values of adults (age > 20 years). Bone age (BA) at start was determined according to the Tanner and Whitehouse radius, ulna, short bones score (RUS TW-2) (24). At each visit, pubertal stages were assessed according to the method of Tanner (21). The onset of puberty was defined as a breast stage II for girls and a testicular volume ≥ 4 ml for boys.

AH was defined as the height reached when HV had decreased below 0.5 cm during the previous 6 months and a BA of 15 years or older for girls and 16.5 years or older for boys.

Imputation of missing values and truncation of extreme values

Missing values were imputed using multiple imputation (25, 26). We generated five imputed data sets using the procedure SAS Proc MI (27). Variables with a non-normal distribution were transformed to normality during the imputation procedure.

To restrict the influence of outlying values, the lowest values of the outcome and of the determinants were truncated to the first percentile and the highest values to the 99th percentile.

Development of prediction model

The potential determinants were 1) initial characteristics included sex, birth length SDS and birth weight SDS (18), gestational age, TH SDS (19), and 2) characteristics at start of GH treatment included CA, height SDS, weight SDS, BMI SDS (19), BA and CA-BA, maximum GH response to GH stimulation tests, serum IGF-I SDS and IGFBP-3 SDS and GH dose.

We first developed separate models for height SDS at onset of puberty and for AH SDS, using forward selection with an inclusion criterion of $p < 0.05$.

Next we constructed a model for both outcomes (height SDS at onset of puberty and AH SDS), using repeated measurements analysis. We started with all determinants selected in the two separate models as covariables, and the interaction terms for each determinant with time (0 = onset of puberty, 1 = AH), to allow for different effects on the two outcomes. Non-significant ($p > 0.05$) terms were excluded stepwise. After this selection, possible interactions between GH dose and each determinant in the model were tested.

Analysis of residuals

Using the model obtained by the previously described procedure, we calculated predicted outcomes and residuals (observed outcome minus predicted outcome). A possible relation between these values was assessed by examining the scatter plots and by fitting the linear regression with the absolute values of the residuals as dependent and the predictions as independent variable.

Internal validation

For internal validation of the derived model, we used bootstrap techniques (28, 29). This is an important step in the procedure of development of a prediction model, needed to make the model less dependent from the data set. The predictive performance of a model applied to other data sets than the set on which it is

derived, will be lower. Bootstrapping evaluates this difference by taking many random samples from the original data set. The models derived on these samples are consequently applied to the sample and to the original data set, and the performances are compared. The mean of the differences, called the optimism, is used to correct the residual standard deviation and the R^2 of the original model. Another result from bootstrapping is the shrinkage factor for the estimated coefficients of the model, which can be used to obtain the final prediction formula, corrected for over-optimism. For the internal validation of our model, we used 200 bootstrap samples.

Results

The study group consisted of 150 children of whom 115 (35 in study I, 80 in study II) received GH treatment in a dose of 0.033 mg/kg·day (dose 1) and 35 in a dose of 0.066 mg/kg·day (dose 2). Characteristics are listed in Table 1. At onset of puberty, 129 (86%) of the children had a height above -2.00 SDS. Seventy-one children had AH available (38 children with dose 1 and 33 children with dose 2). Fifty-five (77%) of them reached an AH above -2.00 SDS.

For height SDS at onset of puberty, the significant determinants were height SDS at start, TH SDS, GH dose (positive effects), age at start, gender and IGF-I SDS at start (negative effects). For AH SDS, the significant determinants were height SDS at start, TH SDS, GH dose, CA-BA at start (positive effects), and IGFBP-3 SDS at start (negative effect).

Table 1. Descriptives

	Total group (n = 150)			AH group (n = 71)		
	n	Median	10th-90th percentile	n	Median	10th-90th percentile
Male/female	150	87/63		71	47/24	
Gestational age (wk)	150	38.0	31.0 to 40.0	71	38.0	31.0 to 40.0
Birth length SDS	123	-3.4	-5.0 to -2.1	68	-3.2	-6.0 to -2.1
Birth weight SDS	150	-2.4	-3.8 to -1.0	71	-2.6	-3.8 to -1.2
Target height SDS	147	-0.5	-1.5 to 0.7	71	-0.7	-2.0 to 0.4
At start of GH treatment						
CA (yr)	150	7.5	4.9 to 10.3	71	7.9	4.1 to 10.7
Height SDS	150	-3.0	-3.8 to -2.1	71	-3.1	-4.0 to -2.2
Weight SDS	150	-2.7	-3.4 to -1.7	71	-2.8	-3.5 to -1.7
BMI SDS	150	-1.2	-2.7 to 0.3	71	-1.2	-2.8 to 0.2
IGF-I SDS	146	-0.8	-2.3 to 0.6	70	-1.2	-2.6 to 0.3
IGFBP-3 SDS	144	-1.3	-2.8 to 0.0	70	-1.3	-3.4 to -0.1
Maximal GH peak (ng/ml)	129	8.3	4.1 to 15.7	66	9.2	3.6 to 15.8
Bone age (yr)	139	6.5	3.5 to 9.1	71	7.0	3.1 to 9.8
CA-BA (yr)	139	0.9	-0.4 to 2.3	71	0.8	-0.9 to 1.9
GH dose (single/double)	150	115/35		71	38/33	
At start puberty						
CA (yr) - boys	87	12.1	10.9 to 13.3	47	11.9	10.7 to 13.2
- girls	63	10.8	9.5 to 12.2	24	11.2	9.7 to 12.9
Height SDS	150	-1.2	-2.2 to 0.1	71	-1.0	-2.0 to 0.2
Duration GH treatment (yr)	150	4.0	1.6 to 6.7	71	4.0	1.5 to 7.2
Δ height SDS since GH start	150	1.9	0.9 to 3.0	71	2.1	1.1 to 3.6
Adult height						
Duration GH treatment (yr)				71	8.0	6.0 to 11.9
AH SDS				71	-1.4	-2.5 to -0.3
Δ height SDS during puberty				71	-0.4	-1.8 to 0.5
Δ height SDS since GH start				71	1.6	0.7 to 2.7
AH SDS – TH SDS				71	-0.5	-2.1 to 0.4

So for the repeated measurements analysis, fitting one model for both outcomes, we started with candidate predictors TH SDS, GH dose, gender, age, CA-BA, height SDS, IGF-I SDS and IGFBP-3 SDS (all measured at start of treatment) and the interactions of these variables with time. In the backward selection, IGFBP-3 SDS was significant but IGF-I SDS was removed, because it had no significant contribution to the model. IGFBP-3 SDS was correlated with IGF-I SDS (r = 0.47, p < 0.0001) but proved to be a stronger determinant. The interactions of height SDS at start, TH SDS and GH dose effects with time (onset of puberty or AH) were not significant, indicating that the effects of these determinants are equal for both outcomes. The effects of age at start and gender were only significant for height SDS at onset of puberty, whereas IGFBP-3 SDS was only significant for AH SDS. For CA-BA, the effect on height at onset of puberty was much smaller than the effect on AH SDS. There was a significant interaction between GH dose and IGFBP-3 SDS. This indicated that there was not one constant dose effect, but the dose effect was depending on the value of IGFBP-3 SDS at start. A higher IGFBP-3 was related to a smaller effect of GH dose. The relation between the GH dose effect and the level of IGFBP-3 is illustrated in Figure 1. We plotted the outcomes (height SDS at onset of puberty in Figure 1A and AH SDS in Figure 1B), adjusted for gender, TH SDS, height SDS, age and CA-BA at start, against the IGFBP-3 SDS value. Through these scatter plots, separate splines were drawn for cases treated with dose 1 and cases treated with dose 2. The distance between the splines, which is decreasing with increasing value of IGFBP-3 SDS, represents the dose-effect on the outcome. The dose-effect is plotted at the bottom of each figure. The plots also show that there is only a minority (15%) of children with IGFBP-3 below -2.5 SDS, where the GH dose effect is substantial.
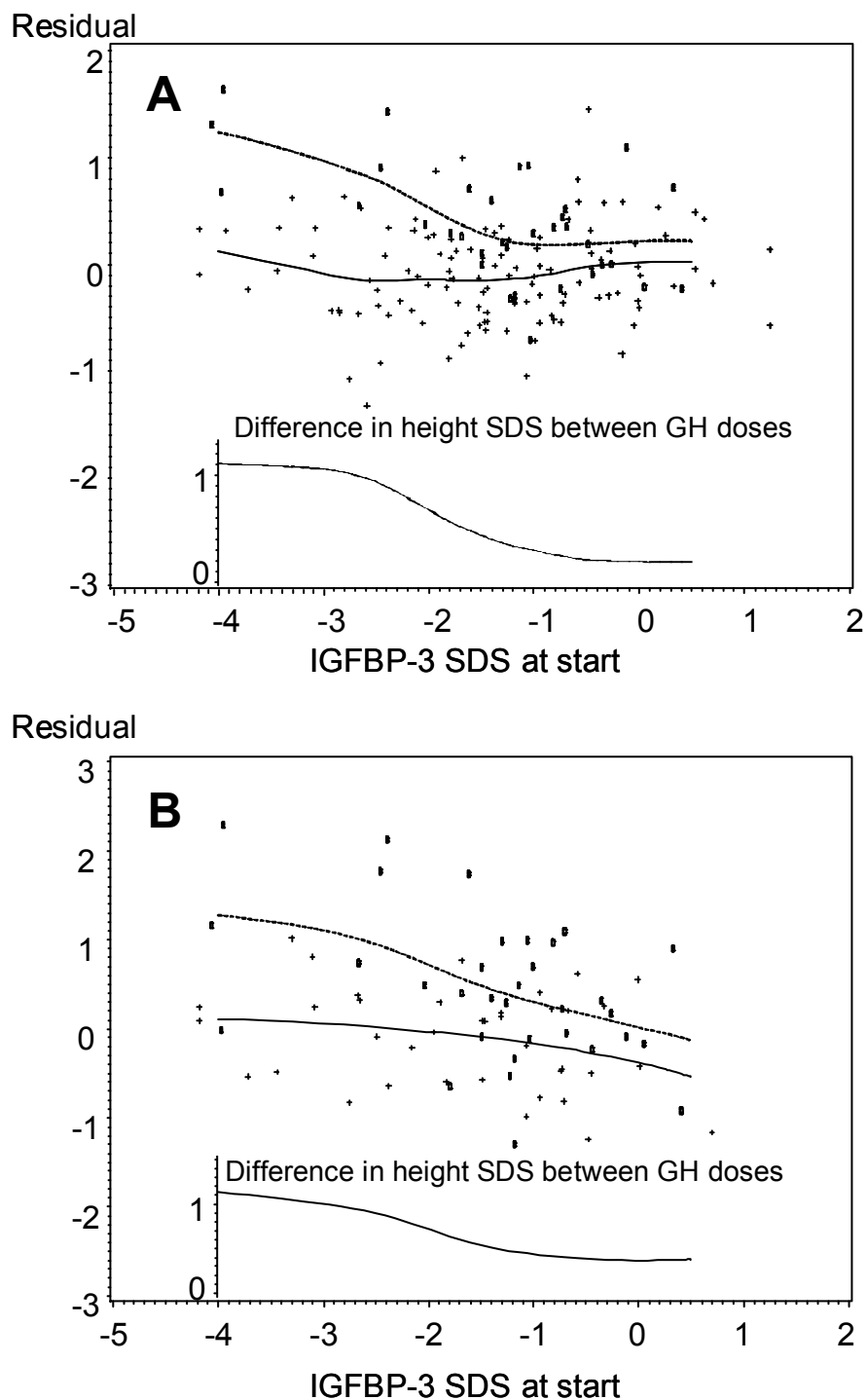
Figure 1. Illustration of the effects of GH dose in relation to IGFBP-3 levels (SDS). Plotted are height SDS at onset of puberty (A) and AH (B), both adjusted for gender, TH SDS, height SDS, age and CA-BA at start against IGFBP-3 SDS at start, dose 1 (+) and dose 2 (0). The distance between the splines indicates the dose-effect and is plotted at the bottom.

The model explained 74% of the variance in height SDS at onset of puberty and 68% of the variance in AH SDS. After correction for overfitting these percentages were 57% and 41%. Table 2 shows the results of the repeated measurements model.

Table 2. Results of the final model for prediction of height at onset of puberty and adult height

Predictor variable	Height SDS at onset of puberty			Adult height SDS		
	Estimated coefficient	SE	P-value	Estimated coefficient	SE	P-value
Intercept	3.17	0.29	<0.0001	0.014	0.43	0.97
Height SDS at start	0.71	0.07	<0.0001	0.71	0.07	<0.0001
Target height SDS	0.13	0.05	0.009	0.13	0.05	0.009
IGFBP-3 SDS at start	0.003	0.040	0.95	-0.14	0.07	0.06
GH dose[a]	0.16	0.14	0.25	0.16	0.14	0.25
GH dose*IGFBP-3 SDS at start[b]	-0.29	0.08	0.0003	-0.29	0.08	0.0003
Gender[c]	-0.34	0.08	<0.0001	0.09	0.15	0.57
Age at start	-0.27	0.02	<0.0001	0.02	0.04	0.57
CA-BA at start (yr)	0.07	0.04	0.07	0.23	0.07	0.002
Residual SD corrected (original)	0.49 (0.45)			0.72 (0.62)		
R^2 corrected (original)	0.57 (0.74)			0.41 (0.68)		
Shrinkage factor	0.97			0.92		

After correction for overfitting and removal of non-significant terms, the prediction formulas are:
H SDS at onset of puberty = 3.10 + 0.70 * H SDS_0 + 0.13 * TH SDS - 0.004 * IGFBP-3 SDS + 0.16 * GH dose - 0.28 * GH dose * IGFBP-3 SDS + 0.070 * (CA-BA) - 0.34 * Gender - 0.27 * Age,
AH SDS = 0.11 + 0.66 * H SDS_0 + 0.12 * TH SDS - 0.11 * IGFBP-3 SDS + 0.15 * GH dose - 0.27 * GH dose * IGFBP-3 SDS +0.21 * (CA-BA).

[a] Dose 0.033 mg/kg·d = 0, 0.066 mg/kg·d = 1
[b] This interaction term indicates that the GH dose effect is related to the value of IGFBP-3 SDS at start. Because of the negative coefficient of this term, the relation is: for lower values of IGFBP-3 the GH dose effect is higher.
[c] Male = 0, Female = 1

To obtain prediction formulas, we refitted the model using for each outcome only the relevant terms (p < 0.20) and applied the shrinkage factor determined by bootstrapping. The prediction formulas are also given in Table 2.

Table 3. Examples of predictions

Child			Data at start of GH therapy			
			Age (yr)	H SDS	IGFBP-3 SDS	CA-BA
1	Female	-0.5	4	-3.1	-2.0	1.0
2	Male	-2.0	6	-4.0	0.0	0.5
3	Male	0.0	5	-3.7	-2.5	0.0
4	Female	-0.5	8	-3.0	0.5	-2.0

	Predictions								
	H SDS at onset of puberty		H SDS at AH		THcAH SDS[a]		H SDS gain		
Child	Dose 1[b]	Dose 2[b]	Dose 1	Dose 2	Dose 1	Dose 2	Dose 1	Dose 2	D2-D1[c]
1	-0.48	0.24	-1.57	-0.88	-1.07	-0.38	1.53	2.22	0.69
2	-1.55	-1.39	-2.67	-2.52	-0.67	-0.52	1.34	1.49	0.15
3	-0.83	0.03	-2.06	-1.23	-2.06	-1.23	1.64	2.47	0.83
4	-1.71	-1.69	-2.41	-2.39	-1.91	-1.89	0.60	0.61	0.01

[a] Target Height corrected AH SDS = AH SDS – THSDS
[b] Dose 1 = 0.033 mg/kg·day, Dose 2 = 0.066 mg/kg·day
[c] D2-D1 is the difference in H SDS gain between Dose 2 and Dose 1

In Table 3 and Figure 2, examples of predictions for four children are shown, presuming that either 0.033 mg GH / kg·day (dose 1) or 0.067 mg GH / kg·day (dose 2) is given. The predicted height SDS at onset of puberty and the predicted AH SDS were plotted, both with 95% prediction interval. A reference line was drawn at -2 SDS and the TH range (defined as the TH SDS +/- 1.3) of the child was marked. For each child, AH SDS is lower than height SDS at onset of puberty. This decrease in height SDS during puberty is mainly related to age at start (for late starters the decrease is lower), and to less extent to gender (more decrease for males) and IGFBP-3 SDS (more decrease for higher levels). At the right of each plot, the total gain in height SDS during treatment is plotted.
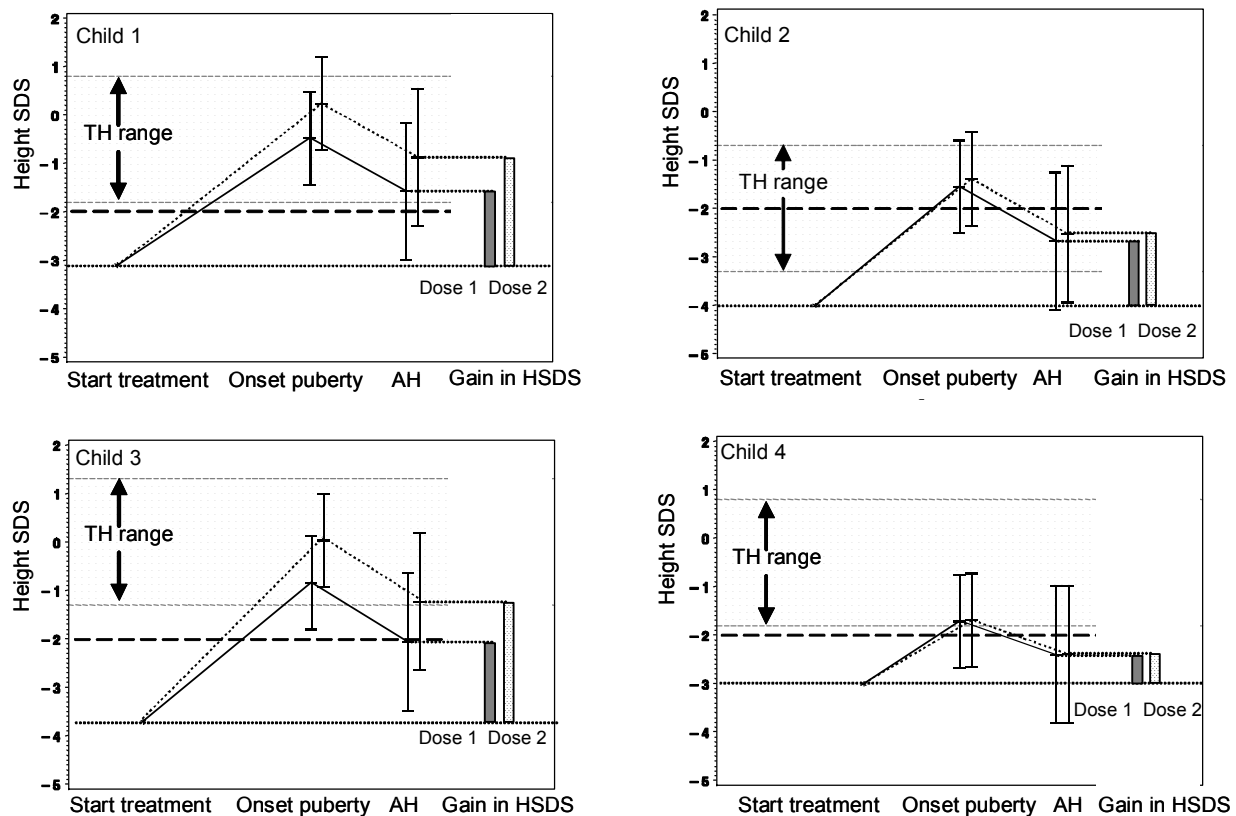
Figure 2. Examples of predictions for four children (see also Table 3). Predicted values at onset of puberty and at AH and 95% prediction interval are plotted for dose 1 (connected with solid lines) and for dose 2 (connected with dashed lines). The TH range (TH SDS +/- 1.3) is marked. At the right, the gain in height SDS is plotted.

Applying the prediction formula, it is expected that child 1 will achieve an AH above -2 SDS, with any of the GH dosages. For both dosages, predicted AH lies within the TH range, and the total gain in height SDS during treatment will be more than 1.5. Treatment with a GH dose of 0.033 mg/kg·day is likely to be sufficient for this child. Child 2 has a predicted AH below –2 SDS, even when dose 2 is given. However, with any dose, the predicted AH lies within the TH range. Dose 1 will result into a gain in height SDS of 1.34, whereas for dose 2 the predicted gain is 1.49 SDS. The small difference (1 cm) in the predicted growth response between the two doses is related to the relatively high serum level of IGFBP-3 (0 SDS). Therefore, for child 2, it seems not reasonable to prescribe the higher dose. On the contrary, for child 3, the dose effect is expected to be substantial. With dose 1, an AH below -2 SDS is predicted, whereas the prediction is -1.23 SDS with dose 2, so within the normal range and within the TH range. This may justify prescribing a dose of 0.067 mg GH /

kg·day. For child 4, aged 8 years, the prospects are rather low, and the effect of a higher dose is limited, again because of the relatively high level of IGFBP-3 (0.5 SDS). The predicted gain in height SDS is below 1, with either dose. One may doubt if GH treatment will be beneficial for this child.

Discussion

We have developed a model to predict height SDS at onset of puberty and adult height (AH SDS) for children born SGA with persistent short stature, in case GH treatment is considered. For height SDS at onset of puberty, the model explains 57% of the variance and for AH SDS 41%. These percentages are corrected for overfitting, in order to give a realistic presentation of the performance of the model in general practice.

Our model includes generally available factors as TH SDS and height SDS and BA at start of treatment. Serum IGFBP-3 levels are not always available for short children born SGA. When we removed IGFBP-3 SDS as potential determinant, the model included IGF-I SDS as a significant determinant (estimated coefficient = -0.08, $p = 0.03$). This model had an 11% lower explained variance compared to our final model including IGFBP-3 SDS. There was no significant interaction between IGF-I SDS and GH dose.

We found a significant interaction between GH dose and IGFBP-3 SDS, with a negative coefficient. This means that the effect of a higher dose is depending on the level of IGFBP-3 of the child: a lower IGFBP-3 SDS at start is related to a larger dose-effect. For a child with an IGFBP-3 SDS of -2, the expected difference in AH is 0.65 SDS (approximately 4.6 cm) when treated with 0.067 mg GH / kg·day compared to 0.033 mg GH / kg·day. For a child with an IGFBP-3 SDS of 0, this difference is only 0.11 SDS (0.8 cm). So estimating the effect of a higher dose of GH treatment for an individual child requires a measurement of serum IGFBP-3 at start of treatment. We therefore recommend measurement of IGFBP-3 and calculation of its SDS in SGA children for whom GH treatment is considered.

When a model was constructed only for the outcome height at onset of puberty (the first step in our procedure), IGF-I SDS was selected as significant predictor. However, in the model only for the outcome adult height, IGFBP-3 SDS was superior over IGF-I SDS as predictor. There was a significant correlation between IGFBP-3 SDS and IGF-I SDS. Also in the repeated measurement model, where we applied backward selection and started with both IGF-I SDS and IGFBP-3 SDS as candidate

predictors, IGF-I SDS was not significant and therefore not included in the final model.

Internal validation of the prediction model was performed by bootstrapping. With this procedure, the model is corrected for overfitting and will give valid predictions with correct confidence limits for new patients. It is not necessary to perform an external validation. Applying the model to data of an independent cohort is still interesting.

Ranke et al. (7) developed prediction models for short-term growth response to GH treatment in short children born SGA. For both first and second year response (in cm/yr), the predicting factors were: age at start (negatively related), weight SDS at start, mid-parental height SDS and GH dose (all positively related). In a Swedish study (9), a regression model for AH SDS was presented with paternal height SDS, height SDS at start (both positively related), age at start and maximal GH response during provocation tests (both negatively related) as predictors. The Swedish study group differed from our group with respect to the percentage of children with GH deficiency (37% with a maximal GH response below 5.3 μg/L, versus 21% in our study group) and the age range (2.5 to 15.1 years versus 3.0 to 11.2 years in our study group). In addition, dosage of GH treatment was only randomised during puberty and no significant influence of GH dose on the AH was found.

The interpretation of a predicted value for AH SDS depends on the objective of GH treatment. Firstly, the goal might be to achieve an AH within the normal range (above -2 SDS). Secondly, the aim could be an AH in the TH range, usually defined as the TH +/- 1.3 SDS. Thirdly, a substantial gain in height SDS during treatment might be regarded as a reasonable goal. When, for an individual child, the AH predicted for a GH dose of 0.033 mg/kg·day is unsatisfactory, a higher dose should be considered. Our study included only two fixed dosages (0.033 or 0.066 mg/kg·day). Therefore, for estimating results of intermediate doses, we have to specify the type of relation between GH dose and the effect on height SDS. Two assumptions are reasonable: a linear relation, meaning that each extra mg of GH has the same additive effect, or a loglinear relation, used in some studies (30, 31), meaning that the effect of an extra mg GH decreases with increasing dose. For example, the effect of 0.040 mg/kg·day compared to 0.035 mg/kg·day is than larger than the effect of 0.065 mg/kg·day compared to 0.070 mg/kg·day. According to our model, for a child with IGFBP-3 SDS at start of -2, the effect of 0.066 mg/kg·day compared to 0.033 mg/kg·day is 0.65 SDS. If the effect of GH dose is linear, the effect of a dose in the middle between these two doses, so 0.050 mg/kg·day, is half of 0.65, so 0.325 SDS. However, if the effect of GH dose is loglinear, the effect of a

dose of 0.05 mg/kg·day is 0.38 SDS. For doses in between 0.033 and 0.066 mg/kg·day, the difference between the two assumptions (linear or loglinear) is, however, small and for simplicity a linear relation can be assumed. Only for doses outside this range (extrapolation of dose-effect), the difference might become large. We therefore recommend the following: For a child born SGA, the predicted AH SDS for a GH dose of 0.033 mg/kg·day should be computed. If this prediction is satisfactory, according to the considerations mentioned above, the child is treated with this dose. If not, the predicted AH SDS for dose 0.066 mg/kg·day is computed. If this prediction is still relatively low, the benefit of any GH treatment should be doubted. If this prediction is in the desired range, the child is treated with this double dose. If the prediction is too far above the aimed AH SDS, the dose can be computed using the following formula:

$$GH\ dose\ (mg/kg\cdot day) = 0.033 * ((aimed\ AH\ SDS - 0.11 - 0.66 * Ht\ SDS$$
$$- 0.12 * TH\ SDS + 0.11 * IGFBP\text{-}3\ SDS - 0.21 * (CA - BA))$$
$$/ (0.15 - 0.27 * IGFBP\text{-}3\ SDS) + 1).$$

For example, a child with a height SDS of -3.86, age 7.35 years, BA 6 years, IGFBP-3 SDS -2.46 and TH SDS -1.57, has a predicted AH SDS for dose 0.033 mg/kg·day of -2.07, so still outside the normal range. If a dose of 0.066 mg/kg·day will be used, the predicted AH SDS is -1.26. If the aimed AH SDS is -1.5, the required GH dose is computed as 0.056 mg/kg·day.

The prediction formulas given in Table 2, as well as the formula for calculation of the suitable GH dose, described above, will be implemented in the next update of the Growth Analyser (www.growthanalyser.org).

In conclusion, the presented model predicts the expected AH of a short child born SGA who will be treated with GH. This may help in determining the optimal GH dose for each child.

References

1. Clayton PE, Cianfarani S, Czernichow P, Johannsson G, Rapaport R, Rogol A 2007 Management of the Child Born Small for Gestational Age Child (SGA) through to Adulthood: A Consensus Statement of the International Societies of Paediatric Endocrinology and the Growth Hormone Research Society. J Clin Endocrinol Metab

2. Hokken-Koelega A, van Pareren Y, Arends N, Boonstra V 2004 Efficacy and safety of long-term continuous growth hormone treatment of children born small for gestational age. Horm Res 62 Suppl 3:149-54

3. de Zegher F, Ong KK, Ibanez L, Dunger DB 2006 Growth hormone therapy in short children born small for gestational age. Horm Res 65 Suppl 3:145-52

4. Van Pareren Y, Mulder P, Houdijk M, Jansen M, Reeser M, Hokken-Koelega A 2003 Adult height after long-term, continuous growth hormone (GH) treatment in short children born small for gestational age: results of a randomized, double-blind, dose-response GH trial. J Clin Endocrinol Metab 88:3584-90

5. de Zegher F, Hokken-Koelega A 2005 Growth hormone therapy for children born small for gestational age: height gain is less dose dependent over the long term than over the short term. Pediatrics 115:e458-62

6. Cutfield WS, Lindberg A, Rapaport R, Wajnrajch MP, Saenger P 2006 Safety of growth hormone treatment in children born small for gestational age: the US trial and KIGS analysis. Horm Res 65 Suppl 3:153-9

7. Ranke MB, Lindberg A, Cowell CT, Wikland KA, Reiter EO, Wilton P, Price DA 2003 Prediction of response to growth hormone treatment in short children born small for gestational age: analysis of data from KIGS (Pharmacia International Growth Database). J Clin Endocrinol Metab 88:125-31

8. Sas T, de Waal W, Mulder P, Houdijk M, Jansen M, Reeser M, Hokken-Koelega A 1999 Growth hormone treatment in children with short stature born small for gestational age: 5-year results of a randomized, double-blind, dose-response trial. J Clin Endocrinol Metab 84:3064-70

9. Dahlgren J, Wikland KA 2005 Final height in short children born small for gestational age treated with growth hormone. Pediatr Res 57:216-22

10. Boguszewski M, Albertsson-Wikland K, Aronsson S, Gustafsson J, Hagenas L, Westgren U, Westphal O, Lipsanen-Nyman M, Sipila I, Gellert P, Muller J, Madsen B 1998 Growth hormone treatment of short children born small-for-gestational-age: the Nordic Multicentre Trial. Acta Paediatr 87:257-63

11. Chatelain P, Job JC, Blanchard J, Ducret JP, Oliver M, Sagnard L, Vanderschueren-Lodeweyckx M 1994 Dose-dependent catch-up growth after 2 years of growth hormone treatment in intrauterine growth-retarded children. Belgian and French Pediatric Clinics and Sanofi-Choay (France). J Clin Endocrinol Metab 78:1454-60

12. de Zegher F, Albertsson-Wikland K, Wilton P, Chatelain P, Jonsson B, Lofstrom A, Butenandt O, Chaussain JL 1996 Growth hormone treatment of short children born small for gestational age: metanalysis of four independent, randomized, controlled, multicentre studies. Acta Paediatr Suppl 417:27-31

13. de Zegher F, Maes M, Gargosky SE, Heinrichs C, Du Caju MV, Thiry G, De Schepper J, Craen M, Breysem L, Lofstrom A, Jonsson P, Bourguignon JP, Malvaux P, Rosenfeld RG 1996 High-dose growth hormone treatment of short children born small for gestational age. J Clin Endocrinol Metab 81:1887-92

14. de Zegher F, Butenandt O, Chatelain P, Albertsson-Wikland K, Jonsson B, Lofstrom A, Chaussain JL 1997 Growth hormone treatment of short children born small for gestational age: reappraisal of the rate of bone maturation over 2 years and metanalysis of height gain over 4 years. Acta Paediatr Suppl 423:207-12

15. de Zegher F, Albertsson-Wikland K, Wollmann HA, Chatelain P, Chaussain JL, Lofstrom A, Jonsson B, Rosenfeld RG 2000 Growth hormone treatment of short children born small for gestational age: growth responses with continuous and discontinuous regimens over 6 years. J Clin Endocrinol Metab 85:2816-21

16. Johnston LB, Savage MO 2004 Should recombinant human growth hormone therapy be used in short small for gestational age children? Arch Dis Child 89:740-4

17. Arends NJ, Boonstra VH, Mulder PG, Odink RJ, Stokvis-Brantsma WH, Rongen-Westerlaken C, Mulder JC, Delemarre-Van de Waal H, Reeser HM, Jansen M, Waelkens JJ, Hokken-Koelega AC 2003 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) 59:779-87

18. Usher R, McLean F 1969 Intrauterine growth of live-born Caucasian infants at sea level: standards obtained from measurements in 7 dimensions of infants born between 25 and 44 weeks of gestation. J Pediatr 74:901-10

19. Roede MJ VWJ 1980 Growth diagrams 1980; Netherlands third nation-wide biometric survey. Tijdschr Soc Gezondheidszorg 63(Suppl):1-34

20. Rikken B, Wit JM 1992 Prepubertal height velocity references over a wide age range. Arch Dis Child 67:1277-80

21. Tanner JM, Whitehouse RH 1976 Clinical longitudinal standards for height, weight, height velocity, weight velocity, and stages of puberty. Arch Dis Child 51:170-9

22. Hokken-Koelega AC, Hackeng WH, Stijnen T, Wit JM, de Muinck Keizer-Schrama SM, Drop SL 1990 Twenty-four-hour plasma growth hormone (GH) profiles, urinary GH excretion, and plasma insulin-like growth factor-I and -II levels in prepubertal children with chronic renal insufficiency and severe growth retardation. J Clin Endocrinol Metab 71:688-95

23. Rikken B, van Doorn J, Ringeling A, Van den Brande JL, Massa G, Wit JM 1998 Plasma levels of insulin-like growth factor (IGF)-I, IGF-II and IGF-binding protein-3 in the evaluation of childhood growth hormone deficiency. Horm Res 50:166-76

24. Tanner J WR, Cameron N, Marshall W, Healy M, Goldstein H 1983 Assessment of skeletal maturity and prediction of adult height (TW2-method), 2nd ed. Academic Press, London

25. Rubin DB 1987 Multiple Imputation for Nonresponse in Surveys, New York

26. Schäfer JL 1997 Analysis of Incomplete Multivariate Data, New York

27. SAS Online Doc 9.1.3 2006. SAS Institute Inc., Cary, NC

28. Harrell FE, Jr., Lee KL, Mark DB 1996 Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 15:361-87.

29. Efron B. TR 1993 An introduction to the bootstrap, New York

30. Blethen SL, Compton P, Lippe BM, Rosenfeld RG, August GP, Johanson A 1993 Factors predicting the response to growth hormone (GH) therapy in prepubertal children with GH deficiency. J Clin Endocrinol Metab 76:574-9

31. Ranke MB, Lindberg A, Chatelain P, Wilton P, Cutfield W, Albertsson-Wikland K, Price DA 1999 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. KIGS International Board. Kabi Pharmacia International Growth Study. J Clin Endocrinol Metab 84:1174-83.

6

SGA children without early catch-up growth: spontaneous growth and prediction of height at 8 years

Abstract

Background/Aim

About 10-15% of children born SGA has at the age of 2 years a height SDS (HSDS$_{2y}$) still below -2. There is no model to predict which children will catch-up in height after two years of age. The aim of this study was to determine the percentage of children with catch-up growth to a normal height after the age of 2 years and to develop a prediction model for growth after that age.

Methods

In a cohort of 724 SGA children, the percentage of children with HSDS > -2 at 8 years of age was determined. In data of 97 children with HSDS$_{2y}$ < -2, a prediction model was developed for growth between 2 and 8 years.

Results

Thirty-nine percent of children with HSDS$_{2y}$ < -2 reached a HSDS above -2 between 2 and 8 years (6% of the total group). Determinants of growth after age 2 years, all with a positive influence, were the difference between TH SDS and HSDS$_{2y}$, change in height SDS during first 2 years of life, female gender and multiple birth.

Conclusions

Catch-up growth to a normal height occurred in 91% of SGA children, in 6% between 2 and 8 years of age. The difference between TH SDS and HSDS$_{2y}$ was the most important determinant. The presented prediction model can identify children with low or high probability of catch-up growth after the age of 2 years. This may assist to determine which children require medical follow-up.

Introduction

Studies have shown that growth hormone (GH) treatment improves adult height of short children born SGA (8-12). A product licence is available in EU and USA for the GH treatment of SGA children with persistent short stature. To manage GH treatment properly and efficiently, it is important to determine at a young age which children have a reasonable chance to reach an adult height above -2 SDS and for which children this chance is very low (13).

Previously, we have reported about the postnatal growth of 724 premature and full-term SGA infants of whom 85% showed catch-up growth to a height above -1.88 SDS during the first 2 years of life (3). Of the 111 with a height still below -1.88 SDS at the age of 2 years, 109 had a height below -2 SDS. The current study was performed to investigate which percentage of these children had catch-up growth after the age of 2 years and which factors determined this catch-up growth. We developed a prediction model for this growth, which is helpful for decisions if and when treatment with GH might be started.

Subjects and methods

The original cohort consisted of 724 infants born SGA, defined by a birth length (BL) below the third percentile for gestational age (14) (-1.88 SDS). In this group, 613 infants (85%) had catch-up growth to a height at or above the -1.88 SDS during the first 2 years of life (3).

Of the remaining 111 infants, 109 had a height < -2 SDS and 97 of them (52 boys, 45 girls) were enrolled in the present study. Ten children were lost to follow-up and the parents of two children refused to participate in our follow-up study. The remaining group (n = 97) consisted of premature as well as full-term infants. Twenty-one children were enrolled in a GH trial, at a mean age (range) of 7.2 years (4.2 - 12 years).

Data of height measurements between 1980 and 1991 were obtained from patient records at the Departments of Pediatrics of the participating hospitals and from Child Health Care Units/Area Health Authorities. From 1991, growth data were collected prospectively. Data were obtained at different ages and at different intervals for each child. Height measurements were performed with a Harpenden stadiometer or a standard measuring board and expressed as height SDS (15). The growth data used for analysis was restricted to the prepubertal stage, defined by breast development Tanner stage 1 in girls and by genital development Tanner stage 1 or testicular

volume less than 4 ml in boys (16). Furthermore, height measurements after the age of 10 years for boys and 9 years for girls were not used. The growth data of children who started GH therapy were used until the last height SDS before start of GH treatment.

The study was approved by the Medical Ethics Committees of the participating hospitals. Written informed consent was obtained from the parents.

Statistics

Student-t-test was used for comparing continuous outcomes between different groups and the chi-square test for categorical outcomes.

Catch-up growth was defined as a height > -2 SDS at the age of 8 years. For children with prepubertal height measurements beyond that age, height SDS at 8 years ($HSDS_{8y}$) was interpolated using the two adjacent observations. For children without a measurement at of after the age of 8 years, a repeated measurements analysis was performed to obtain an estimate of $HSDS_{8y}$. In this analysis, the available prepubertal growth data of all children was used. For each height measurement, the change in height SDS relative to the height SDS at the age of 2 years (dHSDS) was computed. A model was developed on these repeated outcomes. Firstly, we investigated if the pattern of these data could be described by a linear relation with age or whether use of a transformation of age would fit better. The growth curves were best described by a linear relation between dHSDS and the natural logarithm of age-1. Next, we developed a model including significant determinants of growth as fixed effects and a random age-effect for each child, reflecting the child-specific growth. Furthermore, the correlation between serial measurements of a child was taking into account. As potential determinants of growth, we considered age and its interactions with sex, premature/full-term, multiple birth, birth weight SDS, BL SDS (14), target height (TH) SDS, change in weight SDS during the first six months of life, height SDS at 2 years of age ($HSDS_{2y}$), change in height SDS during the first 2 years of life ($dHSDS_{0-2y}$), the difference between TH SDS and BL SDS (TH-BLSDS) and the difference between TH SDS and height SDS at 2 years of age (TH-$HSDS_{2y}$) (15). Missing values for the determinants were imputed using multiple imputation .

Forward selection was used to build a model with significant effects (p < 0.05). This model provided estimated $HSDS_{8y}$ for the children in the data set, based on the determinants in the model and the available growth data of a child. Combining the observed $HSDS_{8y}$ of part of the group and the estimated $HSDS_{8y}$ of the other part, we reported the proportion of children with catch-up growth at 8 years of age.

To use the model as prediction model for $HSDS_{8y}$ of other SGA children, additionally the standard deviation (SD) for the predicted $HSDS_{8y}$ was determined. Comparing this SD with the SD of all estimated $HSDS_{8y}$, the percentage explained variance was computed. Prediction and SD define the distribution of $HSDS_{8y}$, based on patient characteristics at 2 years of age. This enables to determine, for a child of at least 2 years of age, the probability to reach a $HSDS_{8y} > -2$.

Results

Clinical data are summarized in Table 1. Sixty-nine infants were born premature (before 37 weeks) and 28 were born full-term (between 37 and 43 weeks). Mean BL SDS (corrected for gestational age) was significantly lower for premature infants than for full-term infants (p = 0.001). The growth data between 2 and 10 (9) years of age are plotted in Figure 1.

Table 1. Characteristics of the study group

	Total	Premature	Full-term	P-value[a]
Total number of patients	97	69	28	
Male (%)	54%	57%	46%	0.37
Gestational age (wk)	34.4 ± 3.5	32.6 ± 2.2	38.8 ± 1.4	< 0.0001
Multiple births	17	14	3	0.38[b]
Target height SDS	-0.58 ± 0.94	-0.44 ± 0.92	-0.92 ± 0.91	0.02
Birth weight (g)	1347 ± 563	1069 ± 355	2031 ± 361	< 0.0001
Birth weight SDS	-3.09 ± 0.92	-3.19 ± 0.96	-2.85 ± 0.78	0.10
Birth length (cm)	38.6 ± 4.9	36.0 ± 3.6	44.0 ± 1.7	< 0.0001
Birth length SDS	-4.05 ± 1.34	-4.37 ± 1.41	-3.29 ±0.75	< 0.0001
Target height SDS - Birth length SDS	3.49 ± 1.70	4.03 ± 1.62	2.27 ± 1.19	< 0.0001
Height SDS at age 2 years	-2.62 ± 0.58	-2.68 ± 1.66	-2.47 ±0.31	0.04
Target height SDS - Height SDS at 2 yr	2.04 ± 1.03	2.25 ± 1.01	1.53 ± 0.93	0.002
Height SDS at 2 yr - Birth length SDS	1.44 ± 1.19	1.70 ± 1.23	0.81 ± 0.80	0.0001

Data expressed as mean ± SD
[a]Test of null-hypothesis: Means (or percentages) of premature and full-term children are equal
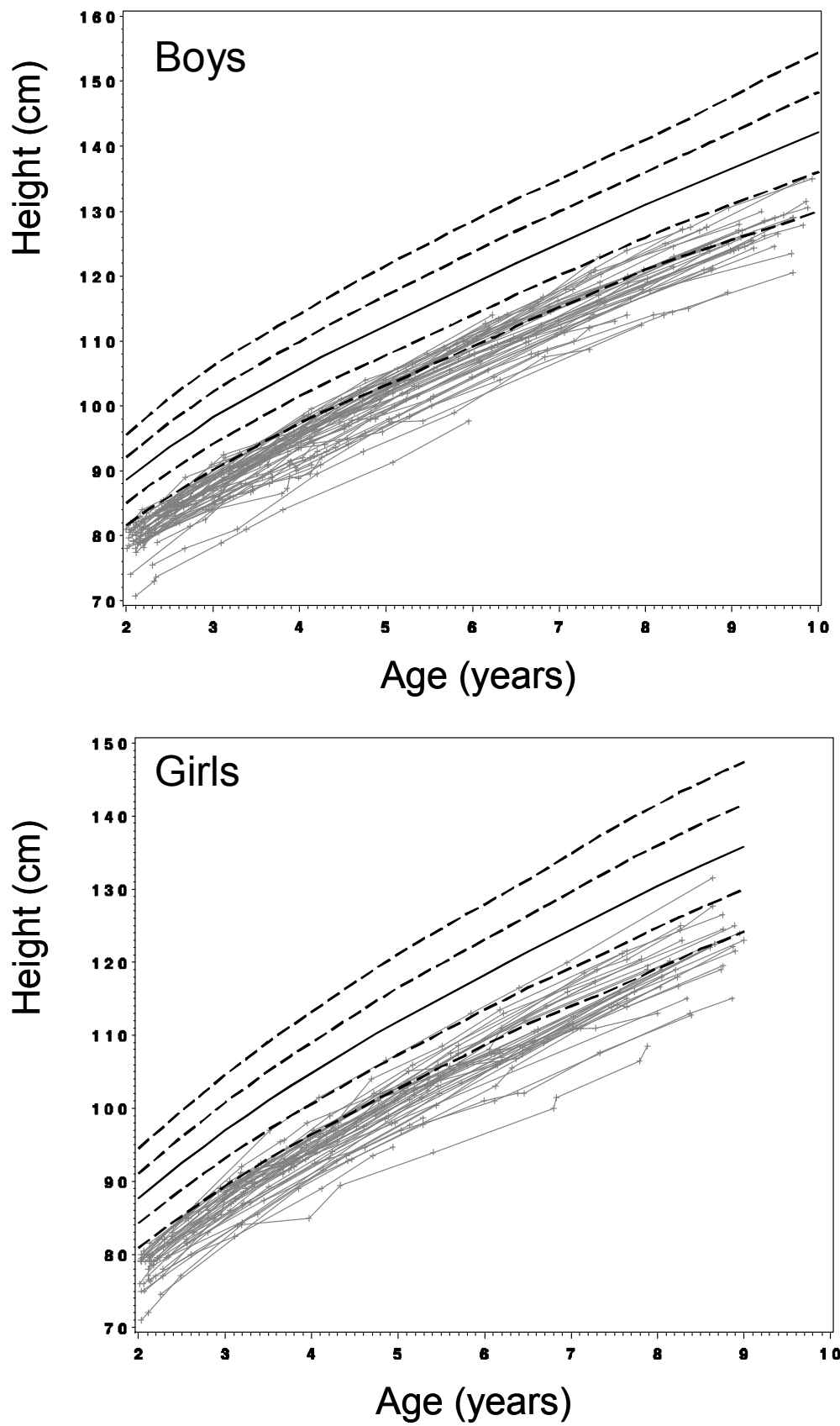[b]Fisher's exact test

Figure 1. Growth charts of the children in the study group. Bold lines indicate the reference lines at -2, -1, 0 +1 and +2 SDS of healthy Dutch children (15).

Catch-up growth in the study group

Thirty-eight (39%) of the children who were short at 2 years of age, had catch-up growth to a $HSDS_{8y}$ > -2. Twenty-seven of them were born premature and 11 were born full-term. Table 2 shows the percentages of premature and full-term SGA children with catch-up growth in three age periods: from birth until 2 years of age, from 2 until 8 years of age and from birth until 8 years of age. The percentages are corrected for the number of children lost to follow-up. In the first 2 years of life, the percentage with catch-up within the full-term infants (88.4%) was higher compared to the premature infants (82.5%) (p = 0.03). In the period from 2 until 8 years of age, the percentage in the full-term children (4.5%) was lower compared to those born preterm (6.9%) (p = 0.22). Over the entire period from birth until 8 years of age, the group of children born premature showed catch-up growth in 89.4%, versus 92.9% in the group of full-term children (p = 0.09). The overall percentage of SGA children with catch-up growth to a height > -2 SDS before the age of 8 years was 90.8%.

Table 2. Percentages of SGA children with catch-up growth to a height SDS > -2 in different age periods

| | | Percentage with catch-up growth > -2 SDS | | |
| | | Age-period (years) | | |
	N	0 – 2	2 – 8	0 – 8
Premature infants	423	82.5	6.9	89.4
Full-term infants	301	88.4	4.5	92.9
Total group	724	84.9	5.9	90.8

Prediction model for growth after 2 years of age

The prediction model for growth after 2 years of age included age (transformed) and its interactions with $TH\text{-}HSDS_{2y}$, sex, $dHSDS_{0\text{-}2y}$ and multiple birth. The estimated coefficients are given in Table 3.

In the right column in Table 3, the coefficients are given when the age of 8 years is filled in. These coefficients determine the prediction rule for $HSDS_{8y}$:

$$HSDS_{8y} = HSDS_{2y} - 0.72 + 0.26*TH\text{-}HSDS_{2y} + 0.25*Sex + 0.33*MultiB$$
$$+ 0.10*dHSDS_{0\text{-}2y}$$

The formula shows, that when children have a $HSDS_{2y}$ far below their TH SDS, they will have more catch-up between 2 and 8 years of age compared to children with a $HSDS_{2y}$ close to their TH SDS. $HSDS_{8y}$ will be 0.26 SDS greater for each SDS difference between TH SDS and $HSDS_{2y}$. Female gender is another positive factor. The difference in $HSDS_{8y}$ between girls and boys, with equal values for the other characteristics, will be on average 0.25. $HSDS_{8y}$ of children from multiple births will be 0.33 greater than of children from single births. The catch-up growth in height between birth and 2 years of age has a small positive effect on $HSDS_{8y}$, being 0.10 SDS for each SDS catch-up growth before the age of 2 years.

Table 3. Estimated coefficients of the final model for delta height SDS relative to height SDS at 2 years of age

Determinant	Estimated coefficient	SE	P-value	Coefficient for $HSDS_{8y}$[d]
AgeTr[a]	-0.3695	0.1031	0.0004	-0.7189
AgeTr* (TH SDS - $HSDS_{2y}$)	0.1348	0.0288	<0.0001	0.2623
AgeTr*Sex[b]	0.1298	0.0561	0.02	0.2525
AgeTr*Multiple birth[c]	0.1683	0.0765	0.03	0.3274
AgeTr*($HSDS_{2y}$ - BL SDS)	0.0519	0.0244	0.03	0.1010

[a]AgeTr = ln(Age - 1), Age in years
[b]Male = 1, Female = 2
[c]Single birth = 0, Multiple birth = 1
[d]For Age = 8: AgeTr = ln(8-1) = 1.95

The SD of the predicted values of $HSDS_{8y}$ was 0.54. The percentage variance in $HSDS_{8y}$ explained by the model was 44%. With a predicted value and SD, probabilities can be computed, for example, the probability to reach a $HSDS_{8y} > -2.5$, or to reach a $HSDS_{8y} > -2.0$.

In Figure 2, three examples are shown. The three (hypothetical) children, all single birth and all with $HSDS_{2y}$ equal to -2.5, have different predicted growth after the age of 2 years, due to their different characteristics. Figure 2A shows predicted growth of a girl with a TH SDS of 1.5 and a BL SDS of -4.5. At 2 years of age, her height SDS had increased by 2 relative to birth, but her height was still 4 SDS below TH. The predicted $HSDS_{8y}$ for this girl is -1.46, and the probability that her $HSDS_{8y}$ will be above -2 is 84%. Figure 2B shows a girl with a TH SDS of -0.5 and BL SDS of -4. The predicted $HSDS_{8y}$ for this girl is -2.04. This prediction is close to -2 SDS, and the probability that $HSDS_{8y}$ will be above -2 is 47%. Figure 2C shows predicted growth of a boy with TH SDS equal to -2 and BL SDS equal to -3. So there was only a minor change in height SDS during the first two years of life and $HSDS_{2y}$ is only 0.5

below TH SDS. His predicted $HSDS_{8y}$ is -2.78 and the probability of reaching a height > -2 SDS is only 7%.
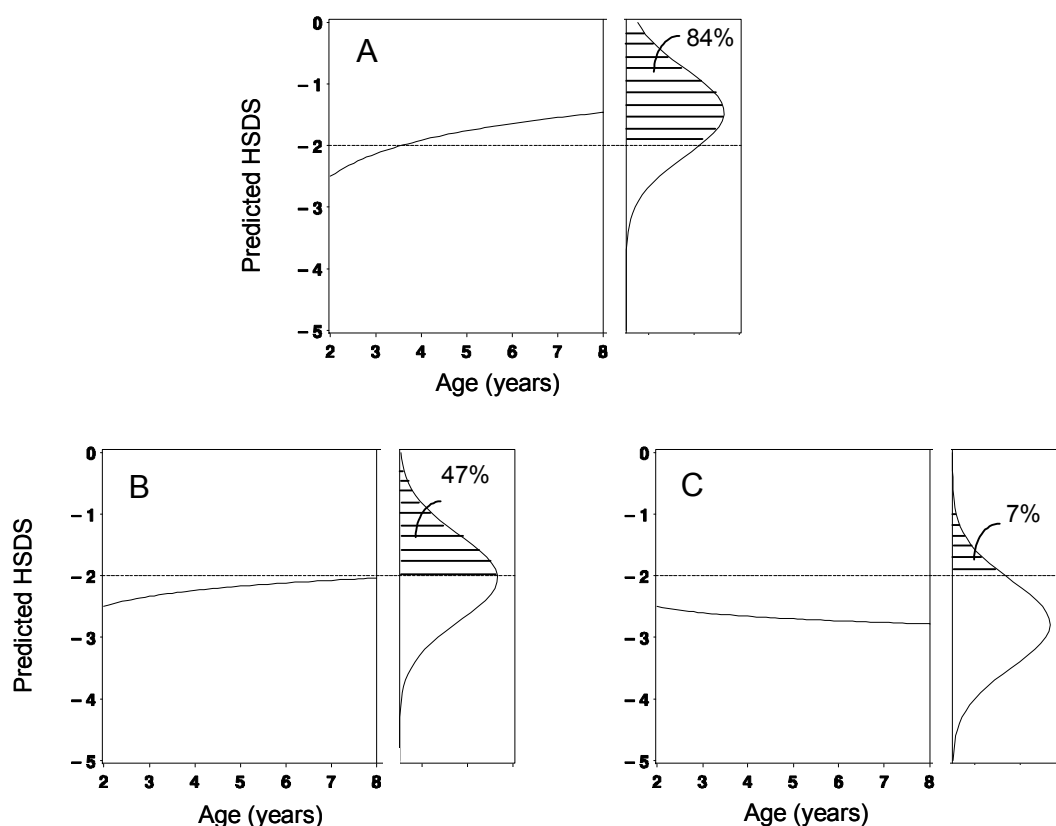

Figure 2. Examples of predicted growth. For the characteristics of the three examples see text. At the left of the vertical line, the predicted growth between 2 and 8 years of age is plotted. The Gauss-curve at the right of each plot gives the distribution of the predicted $HSDS_{8y}$. The marked area represents the probability to reach a $HSDS_{8y}$ > -2.

In Figure 3, the probability to reach a normal height > -2 SDS at 8 years of age, in relation to the predicted $HSDS_{8y}$, is depicted as line A. For the indication to start GH treatment, however, the height should be less than -2.5 SDS. The probability to reach the -2.5 SDS at 8 years of age is given as line B. For example, for a child with a predicted $HSDS_{8y}$ of -3.2, the probability to reach a normal height > -2 SDS is 1%, according to line A, whereas the probability to reach a height > -2.5 SDS is 10%, according to line B.
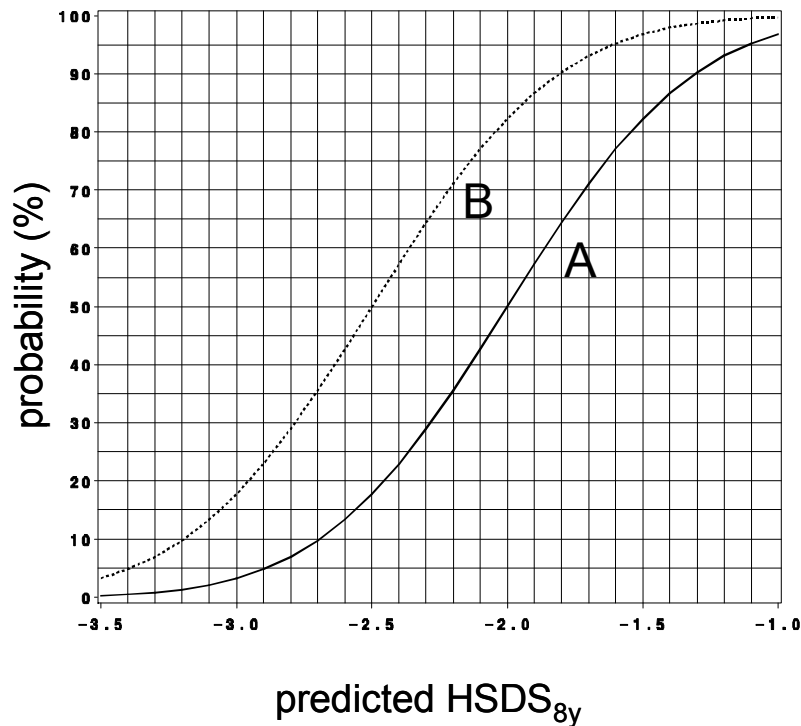
Figure 3. Relation between the predicted HSDS$_{8y}$ and the probability to reach a HSDS$_{8y}$ > -2 (line A) or to reach a HSDS$_{8y}$ > -2.5 (line B)

Discussion

Our study described longitudinal growth of a cohort of 724 healthy premature and full-term children born SGA. Catch-up growth occurred in 90.8%, in 6% between 2 and 8 years of age. So some SGA chidren with persistent short stature at 2 years of age will have spontaneous catch-up growth to a normal height after that age. For growth after the age of 2 years, the predicting factors, all with a positive influence, were: the distance between HSDS$_{2y}$ and TH SDS, female gender, multiple birth and the difference between HSDS$_{2y}$ and BL SDS. With these determinants, a prediction formula was derived for HSDS$_{8y}$. The large effect of the distance between HSDS$_{2y}$ and TH SDS reflects the importance of the genetic growth potential of the children. The characteristics used in the prediction formula are generally available. In the Netherlands, infants are regularly measured in well-baby clinics, including measurement of height and weight at the age of 2 years.

The total percentage of SGA children with catch-up growth to a height > -2 SDS (90.8%) is in line with the previously reported percentage of 89.8% catch-up growth before the age of 8 years (1, 17). Whereas during the first 2 years the percentage of catch-up growth was higher in the group of full-term children (88.4%) compared to the percentage in preterm children (82.5%), during the period after 2 years of age,

the percentage reaching a height > -2 SDS was lower for full-term children (4.5%) than for premature children (6.9%). The variable premature/full-term was not a significant predictor in the model. The difference in catch-up growth between premature and full term children is due to predictors that are different between the two groups: distance between $HSDS_{2y}$ and TH SDS and difference between $HSDS_{2y}$ and BL SDS. Qvigstad et al. also demonstrated a continuing catch-up growth in very preterm and very low-birth-weight infants after the age of 2 years (17).

In this study, we used $HSDS_{8y}$ as main outcome to prevent bias due to the beginning of puberty. SGA children with short stature at 8 years of age can be expected to remain short until adult height, when they remain untreated. Chaussain et al. reported a correlation of 0.81 between height at onset of puberty and adult height in SGA children.(18) For boys born SGA, a rather modest pubertal growth spurt is reported (19) and for girls a normal growth spurt (2, 19-21).

Based on longitudinal growth data, we developed a prediction model for $HSDS_{8y}$, including variables that are known at the age of 2 years. The repeated measurements analysis leads to valid results although the follow-up measurements were done at irregular time points and not all children had complete follow-up data until 8 years of age. With the model, the probability of catch-up to a normal height SDS > -2 or to a height SDS > -2.5 can be derived. This information can be used to determine further surveillance. For instance, when the probability to reach a height > -2 is high (arbitrarily > 80%), the frequency of follow-up visits might be low. It is very likely that children with this prospect will not become candidates for GH therapy. When the probability is, however, very low (arbitrarily < 10%), more frequent follow-up and referral to a pediatric endocrinologist for further diagnostics are desirable. When the probability to reach a height > -2.5 is < 10%, one may start to discuss with the parents the possibility of treatment with GH after the age of 4 years.

The explained variance of our model was 44%. Further analyses on larger data sets may lead to prediction models with higher accuracy. However, long-term follow-up of untreated SGA children becomes increasingly scarce as many SGA children are nowadays treated with GH. For that reason, more effort should be made to register pre-treatment growth data of SGA children.

In conclusion, 90.8% of a cohort of children born SGA (n = 724) showed catch-up growth to a normal height. This catch-up occurred mainly during the first 2 years of life but in 6% after the age of 2 years. For children with a height SDS at the age of 2 years still less than -2, the presented model enables prediction of future growth.

References

1. Karlberg J, Albertsson-Wikland K: Growth in full-term small-for-gestational-age infants: from birth to final height. Pediatr Res, 1995. 38(5): p. 733-9.

2. Paz I, Seidman DS, Danon YL, Laor A, Stevenson DK, Gale R: Are children born small for gestational age at increased risk of short stature? Am J Dis Child, 1993. 147(3): p. 337-9.

3. Hokken-Koelega AC, De Ridder MA, Lemmen RJ, Den Hartog H, De Muinck Keizer-Schrama SM, Drop SL: Children born small for gestational age: do they catch up? Pediatr Res, 1995. 38(2): p. 267-71.

4. Albertsson-Wikland K, Wennergren G, Wennergren M, Vilbergsson G, Rosberg S: Longitudinal follow-up of growth in children born small for gestational age. Acta Paediatr, 1993. 82(5): p. 438-43.

5. Fitzhardinge PM, Inwood S: Long-term growth in small-for-date children. Acta Paediatr Scand Suppl, 1989. 349: p. 27-33; discussion 34.

6. Tenovuo A, Kero P, Piekkala P, Korvenranta H, Sillanpaa M, Erkkola R: Growth of 519 small for gestational age infants during the first two years of life. Acta Paediatr Scand, 1987. 76(4): p. 636-46.

7. Karlberg JP, Albertsson-Wikland K, Kwan EY, Lam BC, Low LC: The timing of early postnatal catch-up growth in normal, full-term infants born short for gestational age. Horm Res, 1997. 48 Suppl 1: p. 17-24.

8. Carel JC, Chatelain P, Rochiccioli P, Chaussain JL: Improvement in adult height after growth hormone treatment in adolescents with short stature born small for gestational age: results of a randomized controlled study. J Clin Endocrinol Metab, 2003. 88(4): p. 1587-93.

9. Dahlgren J, Wikland KA: Final height in short children born small for gestational age treated with growth hormone. Pediatr Res, 2005. 57(2): p. 216-22.

10. de Zegher F, Hokken-Koelega A: Growth hormone therapy for children born small for gestational age: height gain is less dose dependent over the long term than over the short term. Pediatrics, 2005. 115(4): p. e458-62.

11. Hokken-Koelega A, van Pareren Y, Arends N, Boonstra V: Efficacy and safety of long-term continuous growth hormone treatment of children born small for gestational age. Horm Res, 2004. 62 Suppl 3: p. 149-54.

12. Van Pareren Y, Mulder P, Houdijk M, Jansen M, Reeser M, Hokken-Koelega A: Adult height after long-term, continuous growth hormone (GH) treatment in short children born small for gestational age: results of a randomized, double-blind, dose-response GH trial. J Clin Endocrinol Metab, 2003. 88(8): p. 3584-90.

13. Lee PA, Chernausek SD, Hokken-Koelega AC, Czernichow P: International Small for Gestational Age Advisory Board consensus development conference statement: management of short children born small for gestational age, April 24-October 1, 2001. Pediatrics, 2003. 111(6 Pt 1): p. 1253-61.

14. Usher R, McLean F: Intrauterine growth of live-born Caucasian infants at sea level: standards obtained from measurements in 7 dimensions of infants born between 25 and 44 weeks of gestation. J Pediatr, 1969. 74(6): p. 901-10.

15. Roede MJ VWJ: Growth diagrams 1980; Netherlands third nation-wide biometric survey. Tijdschr Soc Gezondheidszorg, 1980. 63(Suppl): p. 1-34.

16. Tanner J, Growth at adolescence. 2nd ed. 1962, Springfield: Charles C. Thomas.

17. Qvigstad E, Verloove-Vanhorick SP, Ens-Dokkum MH, Schreuder AM, Veen S, Brand R, Oostdijk W, Ruys JH: Prediction of height achievement at five years of age in children born very preterm or with very low birth weight: continuation of catch-up growth after two years of age. Acta Paediatr, 1993. 82(5): p. 444-8.

18. Chaussain JL, Colle M, Ducret JP: Adult height in children with prepubertal short stature secondary to intrauterine growth retardation. Acta Paediatr Suppl, 1994. 399: p. 72-3.

19. Bhargava SK, Ramji S, Srivastava U, Sachdev HP, Kapani V, Datta V, Satyanarayana L: Growth and sexual maturation of low birth weight children: a 14 year follow up. Indian Pediatr, 1995. 32(9): p. 963-70.

20. Albertsson-Wikland K, Karlberg J: Natural growth in children born small for gestational age with and without catch-up growth. Acta Paediatr Suppl, 1994. 399: p. 64-70; discussion 71.

21. Persson I, Ahlsson F, Ewald U, Tuvemo T, Qingyuan M, von Rosen D, Proos L: Influence of perinatal factors on the onset of puberty in boys and girls: implications for interpretation of link with risk of long term diseases. Am J Epidemiol, 1999. 150(7): p. 747-55.

7

**A new method to
determine mean adult
height from incomplete
follow-up data**

Abstract

Background/Aim

In long-term growth studies with adult height (AH) as outcome, reporting is often required while data are incomplete because some participants have not yet reached AH whereas others might be lost to follow-up. Current practice is to analyse only participants who did reach AH, which can easily give biased results. We introduce a new method into the area of growth research.

Methods

We used data of patients from a registration database and a growth study. The new method uses growth data in time intervals. The percentage of children still growing and the mean growth in each interval are used to determine mean AH.

Results

With the new method, estimated mean AHs had smaller bias and standard error than with commonly used methods. The method is not hampered by a correlation between AH and age at reaching AH, unlike methods merely using patients who have reached AH.

Conclusion

In contrast to commonly used methods, the new method provides valid results on mean AH when complete actual measurements of AH are not (yet) available, provided that drop-out, if any, is not related to (disappointing) growth. As it also uses observed data of children with incomplete follow-up, the method employs the data more effectively.

Introduction

In many growth studies the final outcome of interest is adult height (AH). However, when AH is the endpoint, incomplete data can easily cause a problem. Patients might be lost to follow-up or simply have not yet reached AH at the time of reporting.

Current practice in the medical literature is to base the analysis on the subgroup of patients with complete follow-up until AH (Uncensored Cases method) (1-3). The mean AH of these patients is only unbiased if these patients are a random sample of the complete group, so patients with incomplete follow-up do not have another growth pattern compared to the patients with complete follow-up. However, because AH and the age at which AH is reached are usually correlated and patients reaching AH at a relatively older age are more likely to have limited follow-up data, the group of patients with complete follow-up is usually not representative for the whole group. A correct method of analysis should take the information on patients with limited data (censored cases) into account.

Another method, less frequently used, is to include the actual height of the whole group, even though part of them have not yet reached AH (Full Sample method). In practice this is of course only done when all patients are at or very close to AH ("near-adult-height") (4-6). This method will always give an underestimation of the (unknown) mean AH of the complete sample, with the size of the bias highly depending on the percentage of patients censored and the ages at censoring.

Standard methods for the analysis of censored data, such as the well-known Kaplan-Meier method (7), do not work in this situation since the assumption of independent censoring is not fulfilled.

So the currently used statistical methods in growth research are inappropriate, and more sophisticated methods are needed. This problem of incomplete follow-up of growth data is completely analogous to the problem of estimating total medical costs (e.g. from the time of diagnosis until death) for a group of patients who may have censored survival times. In the area of cost-effectiveness research, a new statistically correct method for handling this type of data was recently developed by Lin et al. (8). This method uses the follow-up data of all patients, censored and uncensored, and properly accounts for the incompleteness of the data. In this paper we will show that the currently used methods for analysing AH are biased. We applied the new method (Lin's method) in order to get results that are more precise and valid, provided that loss to follow-up, if any, is not related to disappointing growth.

Patients and methods

The first group of patients (group I) was selected from the database of the Dutch Growth Foundation, containing data of all Dutch children treated with growth hormone (GH). Inclusion criteria were: boys born before January 1, 1980, girls born before January 1, 1982, available growth follow-up from at least the age of 14 until AH, and AH reached (defined as height velocity of < 2 cm/year) before January 1, 2000. The children usually had height measurements every 3 or 6 months during GH-treatment, which was routinely continued until AH was reached. If treatment was stopped before AH was attained, height measurements were usually done every year. This data set was used to show the effect of censoring in a large data set where the complete data are known. The mean AH for males and females was computed. Subsequently we used artificially censored data at time points where only a certain percentage of the patients (from 50% on) had reached AH. Using these censored data we calculated the mean AH of the patients who had reached AH before that date with the Uncensored Cases method and we calculated the mean AH with Lin's method using all follow-up data available at that date. We compared these estimated mean AHs with the mean of the actually measured AHs of the complete group.

The second group of patients (group II) consisted of 39 children participating in a clinical trial and treated with biosynthetic GH until AH (defined as growth of < 0.5 cm during the previous 6 months and bone age 15 years or more for girls and 16.5 years or more for boys). We evaluated this group because it represents a realistic situation of a growth study. The participants of this study were measured every 3 months. We computed the estimated mean AH for males and females using the data known at time points with 2-year intervals during the last years of follow-up of the study. Again we compared the results with the mean AH of the complete data.

Statistics

Lin's method requires growth data at fixed time intervals during follow-up. A data example is given in Table 1. For the first interval the attained height at the end of this interval is used and for subsequent intervals the growth during the interval. For each individual it must be known whether at the end of the last observed interval AH was reached or whether growth follow-up was censored.

Condition for valid application of the method is that censoring (stop of follow-up) of an individual is not related to his or her growth. However, it should be realised

that any other method of analysing the data fails as well if this condition is not satisfied. Secondly Lin's method requires consistency of growth patterns during the total observation period. If there is a time trend, such as changes in treatment or in measurement protocol during a study, results from data colleted early might not be fully representative for the final result.

Table 1. Example of growth follow-up data

Subject	Interval[*]	Age at end of interval (yr)	Growth in interval (cm)	AH reached at end of interval
1	1	14.0	150.2	no
1	2	14.5	4.9	no
1	3	15.0	3.3	no
1	4	15.5	1.3	no
1	5	16.0	0.2	no
1	6	16.5	0.2	no
1	7	17.0	0.2	no
1	8	17.5	0.1	yes
2	1	14.0	151.3	no
2	2	14.5	2.9	no
2	3	15.0	2.7	no
2	4	15.5	2.1	no
2	5	16.0	1.8	no
2	6	16.5	0.8	no
2	7	17.0	0.6	no
3	1	14.0	158.4	no
3	2	14.5	2.8	no
3	3	15.0	2.8	no
3	4	15.5	1.7	no
3	5	16.0	0.1	yes

[*]Interval 1 covers the time from conception until age 14 years, following intervals cover 0.5 years

Table 2 shows how Lin's method works. The first step in the analysis is to carry out a survival analysis for the outcome age at reaching AH (ageAH), where cases with incomplete follow-up are considered as censored at the age of last observed height (7). This gives, for all intervals 1 to K, estimations S_k (k = 1, …., K) for the proportion of patients that is still growing at the start of each interval. The second step is to calculate for each time interval the mean growth using the data of all

individuals contributing to that interval (G_k). In the next step, for each interval S_k is multiplied by G_k and the result for the mean AH is the sum of these products over all intervals:

$$\hat{E} = \sum S_k \cdot G_k$$

For mathematical details, including the calculation of the standard error and the statistical proof of validity, the reader is referred to Lin et al. (8).

Table 2. Illustration of Lin's method

Interval (k)	Age at end of interval (yr)	Proportion of patients still growing S_k	Mean growth in interval (cm) G_k	Contribution of interval to estimation of mean (cm) $S_k \cdot G_k$
1	12	1.00	148	148.00
2	13	1.00	6	6.00
3	14	0.98	7	6.86
4	15	0.84	7	5.88
5	16	0.63	6	3.78
6	17	0.49	4	1.96
7	18	0.34	3	1.02
8	19	0.15	2	0.30
9	20	0.10	1	0.10
12	21	0.00	0	0.00
			Estimation of AH	173.90

Simulation study

Because Lin's method has never been applied to growth data before, we illustrate its validity for growth data by applying it to artificially generated growth data in a simulation process. In that situation we know the true mean and standard deviation (SD) of AH in the population.

1. We generated AHs of 100 imaginary children, using a normal distribution with a mean of 170 cm and an SD of 10 cm. So an estimation of the mean of these AHs should ideally give 170 cm.
2. Next we used five different imaginary growth scenarios. In each scenario there was a different relation between AH and ageAH: no correlation (r (AH, ageAH) = 0), a (weak and strong) positive correlation, meaning that tall AH tends to be related to older ageAH ($r = 0.19$ and $r = 0.36$), and a negative

correlation (tall AH related to younger ageAH), again weak and strong (r = -0.19 and r = -0.36). According to this correlation we generated ageAHs for the 100 children.

3. For each scenario, growth data were generated from age 12 years until ageAH using a model based on real growth curves. Each child got a height at 12 year related to its AH. Growth data in between were generated using a quadratic curve, with random deviations from this curve.

4. Next we gave each child a random censoring age (drawn from a uniform distribution with range 15 - 25). If the censoring age was younger than the ageAH, growth data of this child were restricted to the intervals preceding the censoring age. If ageAH fell before the censoring age, the complete follow-up data of the child were included. This resulted in about 27% of children with incomplete follow-up data for each correlation scenario.

These four steps were repeated 5,000 times, resulting in 5,000 random data sets for each scenario. A large number of simulations is needed to compensate for possible accidental results if only one or a few simulations are done. The results of this simulation study are shown in Table 3.

For the complete data (without censoring) we computed for each simulated data set the mean AH, and calculated the average over the 5,000 simulations and the corresponding SD (Table 3, column 2). Since the estimation is unbiased for complete data sets, the average of the estimated means from the 5,000 simulations should be very close to 170 cm. In Table 3 we give the average bias (difference from the true mean of 170 cm), which, for the complete data, is of course almost zero in all scenarios. The SDs are very close to 1.0 cm (SD of 10 cm of the used distribution divided by the square-root of the sample size of 100) and the coverage probabilities of the 95% confidence intervals are indeed 95%.

For the censored data we first determined the mean AH using the mean of the full sample, i.e. using the AHs of the cases who have reached AH and the last observed height of the cases with incomplete follow-up (Full Sample method, Table 3, column 3). Using this method the calculated means (again for each correlation pattern the average over the 5,000 simulations) are always too small (negative bias), the SDs are larger than 1.0 cm and the coverage probabilities are much too low.

Secondly we computed the mean of the uncensored cases (Uncensored Cases method, Table 3, column 4). This is quite adequate when there is no correlation between AH and ageAH (r = 0: bias -0.013 cm, coverage probability 96%). However, in data sets with a correlation between AH and ageAH, a substantial bias is shown and the coverage probability is too low.

Thirdly we performed a standard survival analysis using all last observed heights, indicating AHs as events and last observed heights of the censored cases as censored (Kaplan-Meier method, Table 3, column 5). These results have substantial positive bias, large SDs and coverage probabilities that are too low.

Finally we used Lin's method (Table 3, column 6). In all situations we found only a very small bias (< 0.2 cm). The SDs are smaller than those of the other methods for censored data, showing that Lin's method is statistically more efficient. The SDs of Lin's method are only slightly larger than 1 (the best possible value for this data set, which can only be reached if complete data are observed), indicating that the data of the censored cases are used in a very efficient way. The coverage probabilities are very close to 95%. These results illustrate that Lin's method gives more valid and reliable results than currently used methods.

Table 3. Comparison of the various methods

	Method of analysis				
Scenario	Complete Data	Full Sample	Uncensored Cases	Kaplan-Meier	Lin's method
r = -0.36	-0.02 (1.01)	-2.50 (1.20)	1.03 (1.16)	2.16 (1.39)	-0.16 (1.03)
	95%	44%	86%	65%	95%
r = -0.19	-0.02 (0.99)	-2.50 (1.14)	0.51 (1.16)	1.82 (1.42)	-0.16 (1.01)
	95%	42%	93%	72%	96%
r = 0	0.01 (1.00)	-2.49 (1.12)	-0.01 (1.16)	1.57 (1.42)	-0.14 (1.02)
	95%	39%	96%	79%	96%
r = 0.19	0.00 (1.02)	-2.52 (1.11)	-0.53 (1.18)	1.25 (1.42)	-0.15 (1.04)
	95%	36%	92%	84%	95%
r = 0.36	0.02 (1.02)	-2.54 (1.06)	-1.02 (1.17)	0.99 (1.44)	-0.14 (1.04)
	95%	33%	86%	87%	95%

Each cell contains the bias of the estimated mean (cm) averaged over the 5,000 simulations, the standard deviation of the mean (in parentheses) and the coverage probability of the estimated 95% confidence interval (= percentage of the 5,000 confidence intervals containing the true mean). r is the correlation between AH and ageAH.

Results

Group I consisted of 194 patients (42% male) who started GH treatment at a median (25th; 75th percentile) age of 11.9 (9.9; 13.3) years and reached AH after treatment for 4.3 (3.1; 6.4) years. Thirty seven percent of the children had idiopathic GH deficiency (GHD), 32% had GHD with known cause, 7% had Turner syndrome, and

24% another diagnosis. The mean height SDS at start was -2.75 (9), mean height SDS at the age of 14 years (which was the starting point of the follow-up data we used) was -2.08 and mean height SDS of AH was -2.05. The median number of measurements per child (after age 14 years) was 11 (8; 15). The results for boys and girls are presented in Figure 1. On the x-axis is the percentage of patients who have reached AH. The calculated mean AHs are based on the data known at the corresponding date. The horizontal line indicates the value of the mean AH computed when the follow-up is complete.
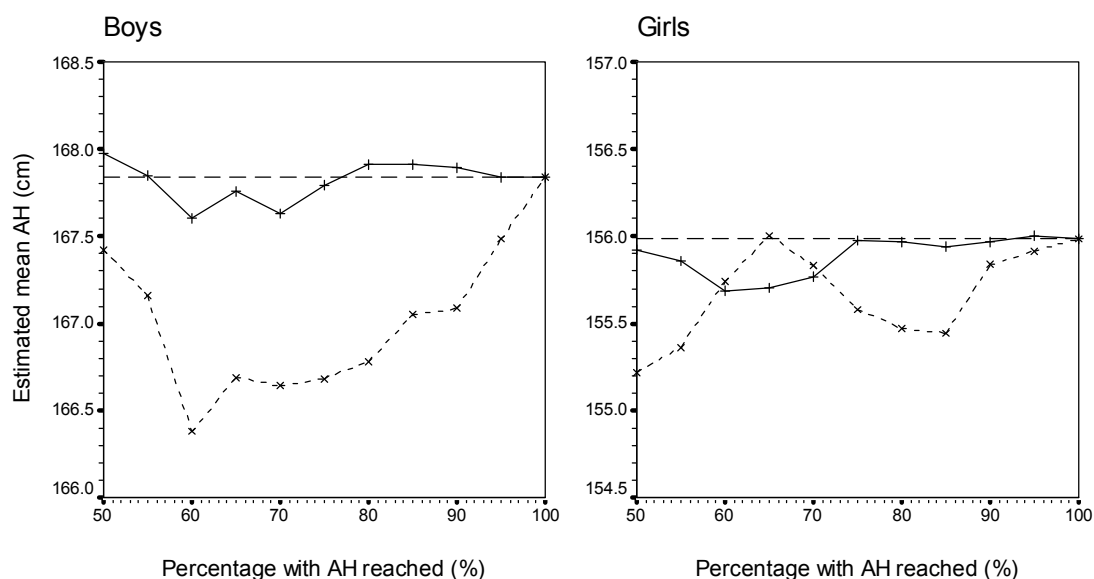

Figure 1. Estimated means for group I versus percentage of patients with complete follow-up. × − − × − − × Uncensored Cases method +—+—+ Lin's method

The Uncensored Cases method gives calculated means with larger deviation from the final mean AH than Lin's method. For the boys (Figure 1a), the final mean AH is 167.84 cm. The largest deviation when using the Uncensored Cases method is 1.46 cm, when 60% of the boys have reached AH. Lin's method has at the same time point a deviation of only 0.23 cm. For the girls (Figure 1b), the largest deviation using the Uncensored Cases method is 0.76 cm, and using Lin's method 0.30 cm. From the time point when 75% of the patients have reached AH, the result of Lin's method is within 0.10 cm of the mean of the complete data, for boys and girls.

Of the 39 children in group II (29 boys, 10 girls) calculated means are given in Table 4 which could be obtained at respectively 7, 9 and 11 years after start of the study (for girls the results at 11 years are omitted because after 9 years the follow-up was complete). In Table 4 the difference between the calculated mean and the

actual mean of the whole group (delta) is a good indicator of the validity of each method. For both boys and girls the delta of Lin's method is closer to zero than the delta of the Uncensored Cases method at each point in time presented. The standard error (SE) of Lin's method is smaller than the SE of the Uncensored Cases method, since the latter does not use all available data.

The obtained means might be converted to SD scores to make it possible to combine the results for boys and girls.

Discussion

In this paper we show that current methods used in growth research to analyse AH data are biased. The magnitude of the bias depends on the percentage of subjects who have dropped out or have not yet reached AH yet at the time of analysis and also on the correlation between AH and ageAH. We have shown that a more advanced method used in cost-effectiveness analyses can be applied to growth data to get valid results, provided that eventual loss to follow-up is not related to the outcome AH. If drop-out is due to disappointing growth, the method can not correct for this loss, nor can any other method.

Until now there has not been any effort to analyse the follow-up of growth data in a more sophisticated way than simply using the uncensored cases: to report the AH results of a growth study, the investigators wait until a sufficiently large percentage of the study group (ideally the number required for sufficient power) has reached AH and will then analyse the observed AHs of this subgroup. When reporting is done for patients from, e.g., a registry, the number of subjects who have reached AH is usually high, and one does not worry that this group in fact is a (non-representative) selection. However, this can give biased results because the patients who did reach AH early in follow-up might be different from the other patients. It also ignores all observed follow-up of the patients who had *not* yet reached AH, despite the fact that these data contain valuable information.

Since 1958 statistical techniques for censored data have been developed (7). These techniques are mostly used for the analysis of survival times, but can also be used for other censored data, e.g. bioassays dealing with a detection limit (left censored data). However, these techniques are not appropriate for the analysis of incomplete growth follow-up. In this situation there is censoring in two dimensions, namely growth and time. The *time* to reach AH is censored and that is the cause of observing incomplete data on *growth*. In this situation standard survival analysis

cannot be applied because the assumption that the attained outcome at censoring is independent of the outcome for complete data is violated.

In medical cost-effectiveness analyses the same phenomenon occurs, when the mean total costs of patients, e.g. from diagnosis until death, is estimated. Censoring occurs because not all patients have died at the time of analysis or some patients have been lost to follow-up, so for some of the patients total medical costs were not observed. Lin et al. (8) provided two methods to estimate mean total costs correctly, one for which no cost history is required and another (preferable) that requires observed costs over time for all patients within small intervals. The latter situation is analogous to growth studies where the height of the participants is usually measured at regular time intervals. Based on theoretical statistical considerations, Lin et al. (8) showed the method to be valid, as long as censoring on ageAH is uninformative. In practice this assumption will be fulfilled, since almost all censoring is due to the patients not having reached yet adult height at the time of the analysis. The validity of the method was illustrated in our simulation example. We showed that Lin's method gives estimated means with only very small bias, even in the case of a correlation between AH and ageAH. The efficiency of Lin's method is higher than currently used methods, as reflected by a smaller SD of the estimations. When there is no correlation between AH and ageAH, the Uncensored Cases method gives a valid estimation, but with 27% of the 100 cases censored the estimation has the accuracy of a study with 73 cases with complete data. The accuracy of Lin's method is comparable to a study with 87 cases. In simulations with 15% and with 40% censoring the accuracy of Lin's method was comparable to studies with n = 91 and n = 78, respectively.

In the first application of the method, to data from the registration database of the Dutch Growth Foundation, we analysed the follow-up data at various levels of censoring. The method we introduce here gave results closer to the mean AH of the complete data than the mean of the uncensored cases.

For the subjects in group II, we performed analyses using data available at different points in time, to see how the methods would perform if the data of a study were analysed when not all follow-up data are complete. Lin's method gave better results compared to the Uncensored Cases method.

In conclusion, when complete actual measurements of AH are not (yet) available and loss to follow-up, if any, is not related to growth, Lin's method provides valid results on the mean AH, in contrast to the commonly used methods and is also more accurate. Using also the observed data of study-participants with incomplete follow-

up, it makes more effective use of the data collected. To execute the computations for the method, standard statistical software is sufficient.

References

1. van Pareren YK, de Muinck Keizer-Schrama SM, Stijnen T, Sas TC, Jansen M, Otten BJ, Hoorweg-Nijman JJ, Vulsma T, Stokvis-Brantsma WH, Rouwe CW, Reeser HM, Gerver WJ, Gosen JJ, Rongen-Westerlaken C, Drop SL. Final height in girls with turner syndrome after long-term growth hormone treatment in three dosages and low dose estrogens. J Clin Endocrinol Metab 2003;88(3):1119-25.

2. Carel JC, Ecosse E, Nicolino M, Tauber M, Leger J, Cabrol S, Bastie-Sigeac I, Chaussain JL, Coste J. Adult height after long term treatment with recombinant growth hormone for idiopathic isolated growth hormone deficiency: observational follow up study of the French population based registry. BMJ 2002;325(7355):70.

3. Bain P, Toublanc JE. Adult height in congenital hypothyroidism: prognostic factors and the importance of compliance with treatment. Horm Res 2002;58(3):136-42.

4. Carel JC, Chatelain P, Rochiccioli P, Chaussain JL. Improvement in adult height after growth hormone treatment in adolescents with short stature born small for gestational age: results of a randomized controlled study. J Clin Endocrinol Metab 2003;88(4):1587-93.

5. Kelly BP, Paterson WF, Donaldson MD. Final height outcome and value of height prediction in boys with constitutional delay in growth and adolescence treated with intramuscular testosterone 125 mg per month for 3 months. Clin Endocrinol 2003;58(3):267-72.

6. Ranke MB, Partsch CJ, Lindberg A, Dorr HG, Bettendorf M, Hauffa BP, Schwarz HP, Mehls O, Sander S, Stahnke N, Steinkamp H, Said E, Sippell W. 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.

7. Kaplan EL, Meier P. Nonparametric estimator from incomplete observations. J Amer Statist Ass 1958;53:457-81.

8. Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. Biometrics 1997;53(2):419-34.

9. Roede MJ, van Wieringen JC. Growth diagrams 1980: Netherlands third nation-wide survey. Tijdschr Soc Gezondheidsz 1985;63(Suppl):1-34.

8

Individually customised fetal weight charts derived from ultrasound measurements. The Generation R Study
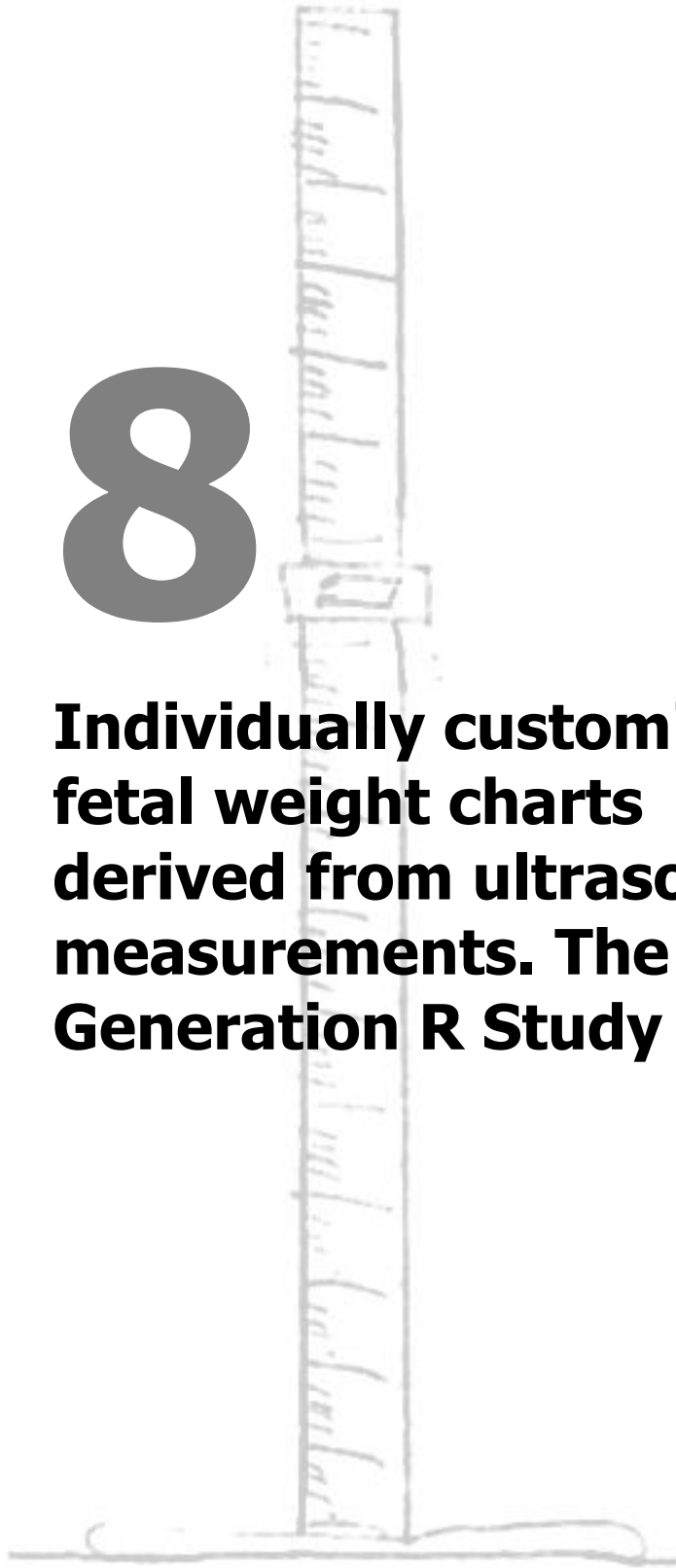
Abstract

Objective

To develop a model for individually customised growth charts for estimated fetal weight, which takes account for physiological fetal and maternal determinants that are fixed at the start of pregnancy.

Methods

Data was used from the Generation R Study, a prospective, population-based cohort study from early pregnancy onwards. In total, 8,162 singleton pregnancies were eligible for the present analysis, having excluded those with major congenital anomalies, termination of pregnancy or perinatal mortality. Of these, 5,473 had complete data on all determinants in our final model.

Results

The final model for estimated fetal weight included the following fetal and maternal determinants: gestational age, fetal gender, parity, ethnicity, maternal age, height, weight and smoking. At a gestational age of 20 weeks, the effects of the determinants were small and not significant, but at 30 weeks the effects were significant. By entering the characteristics into the model equation, we developed individually customised growth charts. In our study population, of the 495 fetuses who were classified as growth restricted (below the P10) when fetal weight was evaluated using the unadjusted reference chart, 80 (16%) were in the normal range when individually customised growth charts were used. On the other hand, 550 were classified as growth restricted using individually customised growth charts, and 135 (25%) were missed if the unadjusted reference was used.

Conclusion

We developed a model to construct individually customised growth charts, adjusted for physiological determinants. This is the first study using ultrasound measurements in a large population-based study to fit such a model. The use of these customised fetal growth charts may improve growth monitoring and prenatal care.

Introduction

Early and accurate detection of fetal growth restriction or macrosomia is important for prenatal and early postnatal care (1, 2). In clinical practice, size and weight of fetuses are evaluated using standard reference tables for fetal biometry measurements. These references do not take into account individual characteristics of the fetus. However, it is shown that gender, parity, ethnicity, maternal weight, maternal height, paternal height and maternal age are important determinants of non-pathological fetal growth variation (3-9). Using standard reference growth charts neglects normal variation in fetal growth due to these characteristics, which hampers the identification of fetuses with pathological growth abnormalities. Customisation of fetal growth charts attempts to adjust for physiological characteristics and so to estimate optimal fetal growth or growth potential for an individual. It is shown that the use of individually customised growth charts improves the distinction between physiological variation and pathological smallness and reduces the false-positive rate for the diagnosis of growth restriction (10-14).

Gardosi et al. developed a method to construct individually customised fetal growth charts (www.gestation.net) (15), based on a regression model for birth weight, using the determinants gender, parity, maternal height, maternal weight at booking and ethnic origin. In this method, the growth curve for estimated fetal weight for an individual is derived from the estimated optimal birth weight, using a formula for estimated fetal weight depending on gestational age (16). So the correctness of the growth curves is depending on the correctness of this formula. Furthermore, it is assumed that the effects of all determinants on fetal weight are proportional throughout pregnancy.

To overcome these limitations, we estimated the influences of physiological characteristics on fetal growth directly, using ultrasound measurements from a large population-based prospective cohort study. We modelled estimated fetal weight throughout pregnancy, obtained by an equation using abdominal circumference, head circumference and femur length (17), because this is the best overall measure of fetal size. In clinical practice estimated fetal weight is mostly used to describe growth anomalies (18). We identified which determinants are relevant to be included in a multiple regression model, considering statistical and clinical significance and availability. Subsequently, using this model, we constructed the individually customised growth charts for the participants in our study. We evaluated how their

fetal growth is assessed using these customised growth charts, compared to assessment using an unadjusted chart.

Material and methods

Design

The Generation R Study is a population-based prospective cohort study, designed to study growth, development and health from early fetal life until young adulthood (19, 20). Eligible mothers were resident in Rotterdam, the Netherlands, at their delivery date (between April 2002 and January 2006). In total, 9,778 mothers were enrolled, of which 8,880 during pregnancy. The response rate in the study, calculated at birth, was 61% (20). The mean maternal age and other characteristics at enrolment were similar to that of all pregnant women in the study area (21).

During the prenatal period, data were collected from physical examinations, questionnaires and fetal ultrasound assessments. The first visit usually took place before the 18th week of the pregnancy as part of routine care. Further ultrasound assessments took place in mid- (gestational age 18-25 weeks) and late-pregnancy (gestational age ≥ 25 weeks) in a research setting (20). Pregnancy outcome was obtained from the midwife or physician who attended the delivery.

The Medical Ethical Committee of the Erasmus Medical Centre, Rotterdam, approved the Generation R Study. Pregnant women and their partners received written and oral information about the study and gave written consent for use of the data.

Ultrasound measurements

Ultrasound examinations were carried out in a research setting at a regional health facility in early (before the 18th week of gestation), mid (18 - 25 weeks) and late (after 25 weeks) pregnancy. Fetal biometry including biparietal diameter (BPD), head circumference (HC), abdominal circumference (AC), and femur length (FL) was measured during each ultrasound examination using a transabdominal probe. Crown-rump length (CRL) was measured in early pregnancy if feasible.

Dating of the pregnancy was performed using the first ultrasound measurement of CRL (in case of a CRL measurement <65 mm, corresponding to 13 weeks of gestation) or BPD, using dating curves derived from this cohort. Establishing gestational age with fetal ultrasound examinations is the most accurate method for pregnancy dating (22-25).

Standardized ultrasound planes for HC, AC, and FL are described elsewhere (26-28). Estimated fetal weight was calculated using the formula of Hadlock with parameters AC, HC and FL (in cm): estimated fetal weight = $10^{**}(1.326-0.00326*AC*FL+0.0107*HC+0.0438*AC+0.158*FL)$ (17). The time period was restricted to gestational age of 18 weeks (earliest reliable estimated fetal weight) to 36 weeks. Visits after 36 weeks were excluded since they were probably performed because of suspected pathology.

Ultrasound exams were performed using an Aloka® Model SSD-1700 (Tokyo, Japan) or the ATL-Philips® Model HDI 5000 (Seattle, Washington, USA) equipped with a 5.0 MHz, high frequency curved array transducer.

Determinants

As determinants of fetal growth, we considered physiological factors that are fixed at the start of pregnancy: fetal gender (coded as male = 1, female = -1), gravidity (primigravida = -1, other = 1), parity (nullipara = -1, other = 1), ethnicity, maternal age (age groups: $\leq$ 27 yr, 28 to 32 yr, $\geq$ 33 yr), height and pre-pregnancy weight as well as paternal height. The pathological determinant maternal smoking was used because it has a substantial effect on fetal growth (29, 30). For the multivariate model, we excluded paternal height, because of its relatively low availability (in our study 73%). Because of the high correlation between gravidity and parity, only the one with the largest effect was used as potential determinant in the multivariate model.

Information about previous pregnancies, pre-pregnancy weight of the mother, smoking habits before pregnancy and ethnicity was collected by a questionnaire at enrolment. The response rate for this questionnaire was 91%. In questionnaires in mid- and late pregnancy mothers were asked whether they smoked in the past two months. Maternal smoking was coded as: smoking in mid or late pregnancy = 1, other = 0. Ethnicity of mothers was defined according to the classification of Statistics Netherlands (31), using country of birth of her parents. Mothers with Moroccan or Surinamese background were asked about their ethnic origin and further classified as Surinamese-Hindustani, Surinamese-Creole, Moroccan-Arabic or Moroccan-Berber. Height of mother and partner was examined at the first prenatal visit.

Population eligible for analysis

Of the total of 8,880 mothers who were enrolled during pregnancy, pregnancies were excluded in case of multiple pregnancy (n = 93), major fetal anomalies (n = 41),

termination of the pregnancy (n = 26) or perinatal mortality (n = 68). Because of difficulty in pregnancy dating, women who joined the study after the 24th week of pregnancy were excluded as well (n = 339). In 151 pregnancies ultrasound observations were only limited, making calculation of estimated fetal weight impossible. This resulted in 8,162 pregnancies eligible for analysis with 16,018 estimated fetal weight observations.

Statistics

An unadjusted reference curve for estimated fetal weight was constructed by modelling the relation between gestational age and estimated fetal weight, using repeated measurement analysis and fractional polynomials (32). Next, the effect of each determinant was estimated separately by adding the main term and its interaction with gestational age to the model. If the interaction term was significant (p < 0.05) it was tested whether adding the interaction with square of gestational age was significant. If the interaction term was not significant it was removed from the model.

For the multivariate model, we started with including the potential determinants together with maternal smoking, to get estimated effects adjusted for maternal smoking. For the construction of a customised growth chart, the term for smoking should be set to zero, whether the pregnant woman smokes or not. This provides that non-smoking is used as reference. As in the univariate models, interactions of the determinants with gestational age were included. Non-significant terms (p > 0.05) were removed using backward selection. Subsequently, we tested whether an additional interaction of a determinant and gestational age squared was significant.

Because of the increasing effects of the determinants during pregnancy, we computed, from the univariate models as well as from the multivariate model, estimated differences in estimated fetal weight at gestational ages of 20 and 30 weeks. For comparison, we also computed estimated differences in birth weight, obtained from models using the data of the neonates born after a gestation of 36 weeks or more. These models included a linear term for gestational age at birth.

Using the multivariate model, we virtually constructed customised growth charts for the participants in our study. Of each participant, the first observation of estimated fetal weight after 27 weeks of gestation was assessed, using the unadjusted growth reference, derived from our data, as well as using the individually customised growth chart.

For all analyses SAS 8.2 (SAS Institute Inc., Cary, NC, USA) was used.

Results

Table 1 depicts descriptives of the study population. Table 2 provides estimated mean, SD, 5th and 95th percentile and width of the 90% reference interval for estimated fetal weight at gestational ages of 20 and 30 weeks. For comparison, the same descriptives are given for birth weight at 40 weeks.

In the univariate models, all considered determinants had significant influence and all effects were significantly increasing with advancing gestational age. In Table 3, the estimated differences in estimated fetal weight between categories of the determinants are given for three time points during gestation. At 20 weeks gestation, the differences were small and only significant for maternal weight. At 30 weeks, almost all differences were significant.

There were 5,473 subjects with complete data on all candidate determinants for the multivariate model (fetal gender, parity, ethnicity, maternal age, maternal height, pre-pregnancy weight and smoking habits during pregnancy). The difference between the unadjusted curve for estimated fetal weight in this group of complete cases and the unadjusted curve for the total 8,162 subjects was less than 4 grams. All determinants had a significant contribution to the model and all interactions with gestational age were significant. For parity and maternal weight, the interaction with gestational age squared was significant. The coefficients of the multivariate model are given in the Appendix. In Table 4, the estimated differences in estimated fetal weight are presented, derived from this model. Because of the correlations between many of the determinants, most of the differences are smaller than the univariate differences. It appears that for a Moroccan Berber or a Turkish mother the expected estimated fetal weight is larger than for a Dutch or other European mother with all other characteristics equal. At 30 weeks gestation, maternal height and weight both had significant influence on estimated fetal weight, independent of each other.

Filling in the individual maternal and fetal characteristics in the regression equation given in the Appendix provides an individually customised growth chart. The expected mean value and standard deviation of estimated fetal weight at a certain gestational age can be computed and these can be used to convert an observed estimated fetal weight into a standard deviation (SD) score and a percentile. These provide a measure of fetal size, relative to fetuses with the specified characteristics. The calculations can easily be done in an Excel-worksheet available on our website.

Table 1. Characteristics of the study population

	Median (P5; P95) or percentage	
Male fetus	50.4%	
Primigravida	43.4%	
Nullipara	56.6%	
Ethnicity		
Dutch	59.4%	
Other European	5.9%	
Dutch Antilles	2.4%	
Cape Verdian	3.5%	
Moroccan-Arabic	1.9%	
Moroccan-Berber	3.7%	
Surinamese-Creole	3.2%	
Surinamese-Hindustani	3.4%	
Turkish	8.0%	
Others	8.6%	
Maternal age (year)	30.3	(20.4; 37.8)
Maternal height (cm)	167	(155; 180)
Pre-pregnancy weight (kg)	64	(50; 91)
Paternal height (cm)	182	(169; 195)
Maternal smoking during pregnancy	17.0%	

Table 2. Descriptives of the distribution of estimated fetal weight (EFW) and birth weight

	GA (weeks)	Mean (gr)	SD (gr)	90% Reference Interval	
				(P5; P95)	Width
EFW	20	326	29	(277; 374)	97
	28	1201	124	(998; 1405)	406
	36	2568	291	(2091; 3046)	955
Birth weight	40	3443	447	(2710; 4176)	1466

The unadjusted reference for EFW is described by:
mean EFW = $13735 - 5.434 \cdot 10^7 * GA^{-2} + 4.297 \cdot 10^7 * GA^{-2} \cdot \log(GA) - 0.889 \cdot 10^7 * GA^{-2} \cdot (\log(GA))^2$ and
SD EFW = $-24.659 + 0.00677 * GA^3$.
The unadjusted equation for birth weight, 36 < GA < 44 weeks, is described by:
mean birth weight = $3443 + 178*(GA - 40)$ and SD birth weight = 447.
GA = Gestational age, P5 = 5th percentile, P95 = 95th percentile.

Table 3. Estimated differences for mean EFW and birth weight, using univariate models

Determinant	n of cases	Differences (gr) in estimated fetal weight						Differences (gr) in birth weight		
		Gestational age (weeks)						Gestational age (weeks)		
		20			30			40		
		Diff	95% CI	P-value	Diff	95% CI	P-value	Diff	95% CI	P-value
Gender fetus (Male − Female)	8111	1	(-5; 6)	0.84	12	(7; 17)	<0.0001	108	(88; 127)	<0.0001
Parity (Para 1 or more − Para 0)	7817	3	(-3; 9)	0.33	25	(19; 31)	<0.0001	177	(157; 197)	<0.0001
Gravidity (Multi − Primi)	7817	1	(-5; 7)	0.68	14	(9; 20)	<0.0001	126	(106; 146)	<0.0001
Ethnicity (ref: Dutch & other European)	6768									
Dutch Antilles		3	(-17; 22)	0.79	-34	(-53; -15)	0.0004	-132	(-202; -61)	0.0003
CapeVerdian		-7	(-23; 10)	0.44	-74	(-90; -59)	<0.0001	-233	(-292; -175)	<0.0001
Moroccan-Arabic		2	(-20; 25)	0.84	-21	(-42; 0)	0.05	-82	(-159; -4)	0.04
Moroccan-Berber		3	(-14; 19)	0.76	15	(0; 30)	0.06	9	(-47; 65)	0.75
Surinamese-Creole		-1	(-19; 16)	0.91	-78	(-94; -61)	<0.0001	-200	(-262; -138)	<0.0001
Surinamese-Hindustani		-10	(-27; 7)	0.26	-75	(-91; -59)	<0.0001	-298	(-358; -238)	<0.0001
Turkish		2	(-9; 13)	0.74	-10	(-21; 0)	0.06	-57	(-97; -18)	0.005
Other		-2	(-14; 9)	0.66	-28	(-40; -16)	<0.0001	-105	(-144; -66)	<0.0001
Maternal age (reference: <= 27 yr)	8162									
28 to 32 yr		-1	(-7; 6)		25	(19; 31)	<0.0001	92	(69; 115)	<0.0001
>= 33 yr		1	(-7; 8)	0.89	44	(37; 50)	<0.0001	131	(106; 155)	<0.0001
Maternal height (10 cm)	8135	4	(0; 8)	0.04	34	(30; 37)	<0.0001	140	(127; 153)	<0.0001
Paternal height (10 cm)	5960	2	(-2; 6)	0.36	31	(27; 35)	<0.0001	106	(92; 120)	<0.0001
Maternal weight (10 kg)	6644	3	(0; 5)	0.03	26	(23; 28)	<0.0001	81	(73; 89)	<0.0001
Maternal smoking (Yes − No)	7197	-2	(-10; 6)	0.58	-47	(-55; -40)	<0.0001	-165	(-193; -137)	<0.0001

[a] All models are adjusted for gestational age (GA) including terms GA^{-2}, $GA^{-2}*\ln(GA)$, $GA^{-2}*\ln(GA)^2$
CI = Confidence interval

Table 4. Estimated differences for mean EFW and birth weight, using a multivariate model

Determinant	Differences (gr) in estimated fetal weight						Differences (gr) in birth weight		
	Gestational age (weeks)						Gestational age (weeks)		
	20			30			40		
	Diff	95% CI	P-value	Diff	95% CI	P-value	Diff	95% CI	P-value
Gender fetus (Male – Female)	0	(-7; 7)	0.98	15	(10; 22)	<0.0001	112	(90; 134)	<0.0001
Parity (Para 1 or more – Para 0)	3	(-5; 10)	0.46	16	(9; 229)	<0.0001	176	(153; 200)	<0.0001
Ethnicity (ref: Dutch & other European)									
Dutch Antilles	5	(-18; 28)	0.69	-9	(-30; 13)	0.43	-92	(-170; -14)	0.02
CapeVerdian	-2	(-20; 17)	0.87	-36	(-54; -19)	<0.0001	-117	(-180; -55)	0.0002
Moroccan-Arabic	5	(-23; 32)	0.75	-8	(-33; 17)	0.52	-33	(-124; 58)	0.48
Moroccan-Berber	3	(-17; 23)	0.79	23	(5; 41)	0.01	15	(-49; 80)	0.65
Surinamese-Creole	-2	(-22; 18)	0.84	-68	(-87; -49)	<0.0001	-161	(-228; -94)	<0.0001
Surinamese-Hindustani	-6	(-26; 14)	0.54	-33	(-51; -14)	0.0006	-163	(-230; -97)	<0.0001
Turkish	2	(-12; 16)	0.75	18	(5; 31)	0.007	30	(-17; 76)	0.21
Other	0	(-14; 13)	0.95	-4	(-16; 8)	0.51	38	(-6; 83)	0.09
Maternal age (reference: <= 27 yr)									
28 to 32 yr	-2	(-10; 7)	0.67	10	(2; 17)	0.02	1	(-27; 30)	0.93
>= 33 yr	-1	(-10; 9)	0.89	28	(19; 36)	<0.0001	12	(-19; 43)	0.45
Maternal height (10 cm)	2	(-3; 8)	0.45	16	(11; 21)	<0.0001	101	(82; 119)	<0.0001
Maternal weight (10 kg)	2	(-1; 5)	0.34	21	(19; 24)	<0.0001	56	(46; 65)	<0.0001
Maternal smoking (Yes – No)	-1	(-10; 8)	0.80	-46	(-54; -37)	<0.0001	-164	(-194; -133)	<0.0001

As example, we examined a hypothetical observation for estimated fetal weight of 1500 g, obtained at a gestational age of 32 weeks for a female fetus, first child of a Surinamese-Creole mother, with maternal height 155 cm, pre-pregnancy weight 45 kg and age 20 yr. Using the unadjusted formula derived in our data (Table 2), the expected estimated fetal weight at 32 weeks is 1865 g and the SD 197 g. So an estimated fetal weight of 1500 g at 32 weeks is converted to an SD score of (1500 - 1865) / 197 = - 1.85. The corresponding percentile is P3. When a customised growth chart is constructed, using the characteristics of this fetus, the expected estimated fetal weight at 32 weeks is 1669 g, with SD 191 g. So taking the physiological determinants into account, the SD score for this observed estimated fetal weight is (1500 - 1669) / 191 = - 0.89, corresponding percentile P19.

The relation between the SD score from the unadjusted reference chart and the SD score from the individual customised growth chart for the same estimated fetal weight observation is depicted in figure 1. From the subgroup with complete data, for each fetus the SD scores of the ultrasound measurement in the third trimester (gestational age $\geq$ 27 weeks) was plotted (n = 5,300). Using the unadjusted reference, 495 fetuses were classified as too small (compartment A and B). However, using individually customised references, 80 of these (16%, A) were classified as normal. If the classification was based on individually customised references, 550 were too small (compartment B and C) and 135 of them (25%, C) were missed when the unadjusted reference was used.
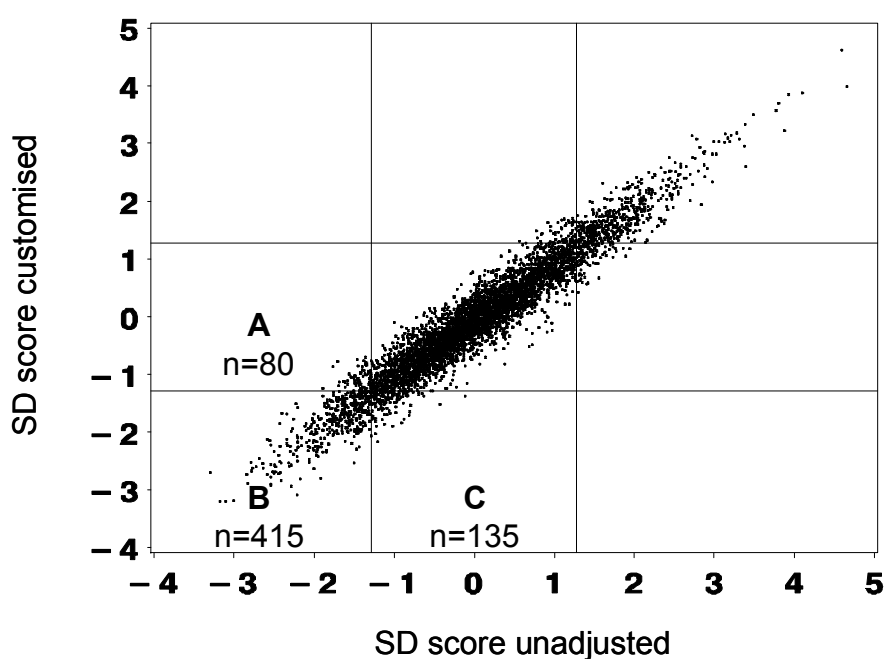

Figure 1. Relation between SD score obtained by the unadjusted reference (x-axis) and by individually customised references (y-axis). Reference lines are drawn at P10 and P90.

Discussion

In this study, we developed a model to construct individually customised fetal growth charts. It is the first time that such a model is fitted using observations of estimated fetal weight obtained by ultrasound measurements in a large population-based prospective study. With the customised growth charts, fetal growth can be evaluated, taking into account the following physiological characteristics: fetal gender, parity, ethnicity, maternal height, pre-pregnancy weight and age. All effects increased with advancing gestational age. The effects were largest for parity and ethnicity.

Our study cohort comprises contemporary urban children including about 40% from ethnic minorities in the Netherlands. The largest ethnic minority groups in this population are from the Turkish, Moroccan and Surinamese groups. Of all eligible children at birth, 61% participated in the study. National and regional registries do not have subject characteristics in all eligible children and their parents that enable detailed non-response analyses in our study. The percentages of mothers from ethnic minorities and lower socio-economic status among the participants were slightly lower than expected from the population figures in Rotterdam (21). This resulted in a more healthy study population, possibly affecting the generalizibility of the results.

Pregnancy dating in this study was done by ultrasound measurements of CRL or BPD at first visit, which is found to be superior over dating by last menstrual period (22-25). However, this procedure neglects possible differences in CRL or BPD, which might be correlated with fetal size, at the time of dating. It is possible that this caused underestimation of the effects of the determinants, especially in early pregnancy. This bias is expected to be smaller when pregnancy dating is performed in early pregnancy. In the present study, we excluded pregnancies that were dated later than at a gestational age of 24 weeks. In total, in our study population 73% of the pregnancies was dated before 18 weeks of gestation. Therefore, we think underestimation of the effects in mid pregnancy is possible, but will be very small in relation to the effects in late pregnancy.

We assumed that the formula for estimated fetal weight (17) is appropriate for the whole study group. This assumption could be questioned. Body-proportions may differ between subgroups. Schild et al. (33) derived formulas for each gender. In an evaluation group, these gender-specific formulas fitted better than established methods. Unfortunately, they did not derive and evaluate a not-gender-specific

formula from their data. As they stated, the gender-specific formulas need to be tested in different settings and population. Differences between ethnic groups might also be reasonable. To our knowledge, no different formulas for estimated fetal weight for ethnic groups or other subgroups are available and it might be doubted to which extend this would be realistic.

The customised antenatal growth charts developed by Gardosi et al. (15) are based on a regression model for birth weight, fitted in a very large group of over 40,000 neonates. The determinants in this model are maternal height and weight, ethnic origin, parity and fetal gender. After calculation of the "term optimal weight" for a child, a fetus-specific intra-uterine growth chart for estimated fetal weight can be constructed using a proportionality equation linking estimated fetal weight during gestation to birth weight. So an important assumption for this approach is that this proportionality equation is correct for each fetus. It also assumes that the effect of each determinant is proportional during pregnancy. Applying the model of Gardosi, we derived the effects of the determinants at gestational ages of 24 and 30 weeks and compared these with the effects estimated by our model. For example, using the model of Gardosi, the estimated difference between male and female fetuses at 24 weeks is 18 g, using our model it is 5 gr. At 30 weeks these differences are estimated as 42 g and 15 g respectively. For the other determinants, our estimated effects were also smaller. The model of Gardosi does not include maternal age and parity is coded in a different way (separate effects for parity 1, 2, 3 and ≥ 4). This may influence the estimations of some effects (e.g. maternal weight, which might be related to maternal age), but not all (e.g. fetal gender). Because our model is based on estimated fetal weight derived from ultrasound measurements, like in clinical practice, we think our charts are better applicable than the charts of Gardosi.

Other studies on factors influencing fetal growth, using ultrasound measurements, have been published before (4-8). Jacquemyn et al. (4) only studied the differences between some ethnic groups, while Schwärzler et al. (7) only studied sex-differences. Mongelli et al. (5) compared different subgroups based on maternal weight, maternal height, fetal gender, parity or ethnicity, but did not develop a model with all determinants included. Pang et al. (6) developed models for BPD, HC, AC and FL but not for estimated fetal weight to avoid potential problems in erroneous estimation of fetal weight and to be able to assess growth restriction in biometric parameters separately. Most comparable to our study is the study of Johnson et al. (8). This study used data of 635 women visiting a low-risk antenatal clinic. It did not comprise different ethnic groups, the effects of maternal weight and parity were not significant and continuous determinants were categorised.

Using customised growth charts enables to identify pathological smallness instead of constitutional small size, with normal intrauterine growth. This can prevent unnecessary classification as growth restricted. On the other hand, studies have shown that failure to detect fetal growth restriction is an important reason for suboptimal perinatal care (34). It is not obvious, which factors should be taken into account when evaluating fetal growth. Some factors may represent both physiological and pathological effects. For example, ethnic differences exist in feeding habits and other life-style factors. However, ethnicity also reflects constitutional growth potential. Studies have shown, that in ethnic groups with lower mean birth weight, the optimal birth weight, defined as birth weight with the lowest perinatal mortality, is also lower (35, 36). Parity is another determinant of which inclusion in the model is debatable. We adjusted for parity, because it is a determinant of the growth potential of a fetus. However, nulliparous women are at higher risk of obstetric complications (37) and stillbirths (38). We are not aware of any study relating perinatal mortality of first born children and second or later born siblings to their birth weight. Therefore, it is not clear whether nulliparity must be seen as a physiological determinant, or should be considered as a possible cause of intrauterine growth restriction. The effects on estimated fetal weight of parity and ethnicity are the largest in our model. Our model can also be used when a customised growth chart adjusted for not all but only a selection of determinants is preferred.

In conclusion, we developed a model for individual customised growth charts that can be used in clinical obstetric settings. This may improve fetal growth monitoring and prenatal care. Further studies are needed to examine whether and to what extend the use of customised growth charts can improve predicting which children are at risk for perinatal or later morbidity.

Appendix

The formula for the mean estimated fetal weight of an individually customised growth chart is:

$$EFW = 17877 - 62328362 \cdot GA^{-2} + 49529740 \cdot GA^{-2} \cdot \ln(GA) - 10323705 \cdot GA^{-2} \cdot (\ln(GA))^2$$

$$15 \cdot Sexe + 150 \cdot Parity + 24 \cdot Age_2 - 33 \cdot Age_3 - 2.58 \cdot Height + 11.27 \cdot Weight$$

$$+ 32 \cdot Ethn_2 + 68 \cdot Ethn_3 + 30 \cdot Ethn_4 - 38 \cdot Ethn_5 + 130 \cdot Ethn_6 + 46 \cdot Ethn_7 - 29 \cdot Ethn_8 + 7 \cdot Ethn_9$$

$$+ GA \cdot (0.78 \cdot Sexe - 12.83 \cdot Parity - 52.45 \cdot Age_1 - 53.58 \cdot Age_2 - 50.76 \cdot Age_3$$

$$+ 0.1395 \cdot Height - 1.0473 \cdot Weight$$

$$- 1.35 \cdot Ethn_2 - 3.48 \cdot Ethn_3 - 1.28 \cdot Ethn_4 + 2.02 \cdot Ethn_5 - 6.60 \cdot Ethn_6$$

$$- 2.63 \cdot Ethn_7 + 1.57 \cdot Ethn_8 - 0.37 \cdot Ethn_9)$$

$$+ GA^2 \cdot (0.2694 \cdot Parity + 0.02476 \cdot Weight)$$

EFW	Estimated Fetal Weight (gr)
GA	Gestational age (weeks)
Sexe	Female = -1, Male = 1
Parity	Nulliparity = -1, Other = 1
Agegroup	Age_1: <= 27 yr = 1, other = 0
	Age_2: 28 to 32 yr = 1, other = 0
	Age_3: >= 33 yr = 1, other = 0
Height	Maternal height - 167 (cm)
Weight	Pre-pregnancy weight - 64 (kg)
Ethnicity	$Ethn_2$: Dutch Antilles = 1, other = 0
	$Ethn_3$: Cape Verdian = 1, other = 0
	$Ethn_4$: Morrocan-Arabic = 1, other = 0
	$Ethn_5$: Morrocan-Berber = 1, other = 0
	$Ethn_6$: Surinamese-Creole = 1, other = 0
	$Ethn_7$: Surinamese-Hindustani = 1, other = 0
	$Ethn_8$: Turkish = 1, other = 0
	$Ethn_9$: Other non-European = 1, other = 0

For construction of a chart without adjustment for one or more determinants, terms for sexe, parity, maternal height and weight can just be neglected, because these are centered around zero. For agegroup and ethnicity a reference category has to be chosen.

The formula for the standard deviation (SD) for the individually customised growth charts is:

$$SD\ EFW = -23.0315 + 0.006523 \cdot GA$$

References

1. Fang S. Management of preterm infants with intrauterine growth restriction. Early Hum Dev 2005 Nov;81(11):889-900.

2. Boulet SL, Salihu HM, Alexander GR. Mode of delivery and birth outcomes of macrosomic infants. J Obstet Gynaecol 2004 Sep;24(6):622-9.

3. Gardosi J, Mongelli M, Wilcox M, Chang A. An adjustable fetal weight standard. Ultrasound Obstet Gynecol 1995 Sep;6(3):168-74.

4. Jacquemyn Y, Sys SU, Verdonk P. Fetal biometry in different ethnic groups. Early Hum Dev 2000 Jan;57(1):1-13.

5. Mongelli M, Gardosi J. Longitudinal study of fetal growth in subgroups of a low-risk population. Ultrasound Obstet Gynecol 1995 Nov;6(5):340-4.

6. Pang MW, Leung TN, Sahota DS, Lau TK, Chang AM. Customizing fetal biometric charts. Ultrasound Obstet Gynecol 2003 Sep;22(3):271-6.

7. Schwarzler P, Bland JM, Holden D, Campbell S, Ville Y. Sex-specific antenatal reference growth charts for uncomplicated singleton pregnancies at 15-40 weeks of gestation. Ultrasound Obstet Gynecol 2004 Jan;23(1):23-9.

8. Johnsen SL, Rasmussen S, Wilsgaard T, Sollien R, Kiserud T. Longitudinal reference ranges for estimated fetal weight. Acta Obstet Gynecol Scand 2006;85(3):286-97.

9. Wilcox MA, Newton CS, Johnson IR. Paternal influences on birthweight. Acta Obstet Gynecol Scand 1995 Jan;74(1):15-8.

10. de Jong CL, Francis A, van Geijn HP, Gardosi J. Fetal growth rate and adverse perinatal events. Ultrasound Obstet Gynecol 1999 Feb;13(2):86-9.

11. de Jong CL, Francis A, van Geijn HP, Gardosi J. Customized fetal weight limits for antenatal detection of fetal growth restriction. Ultrasound Obstet Gynecol 2000 Jan;15(1):36-40.

12. Gardosi J, Francis A. Controlled trial of fundal height measurement plotted on customised antenatal growth charts. Br J Obstet Gynaecol 1999 Apr;106(4):309-17.

13. Mongelli M, Gardosi J. Reduction of false-positive diagnosis of fetal growth restriction by application of customized fetal growth standards. Obstet Gynecol 1996 Nov;88(5):844-8.

14. Bukowski R. Fetal growth potential and pregnancy outcome. Semin Perinatol 2004 Feb;28(1):51-8.

15. Gardosi J, Francis A. Customised Antenatal Growth Charts - GROW-Chart. 7.1 ed: Gestation Network, www.gestation.net; 2006.

16. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. Radiology 1991 Oct;181(1):129-33.

17. Hadlock FP, Harrist RB, Sharman RS, Deter RL, Park SK. Estimation of fetal weight with the use of head, body, and femur measurements--a prospective study. Am J Obstet Gynecol 1985 Feb 1;151(3):333-7.

18. Nyberg DA, Abuhamad A, Ville Y. Ultrasound assessment of abnormal fetal growth. Semin Perinatol 2004 Feb;28(1):3-22.

19. Hofman A, Jaddoe VW, Mackenbach JP, Moll HA, Snijders RF, Steegers EA, et al. Growth, development and health from early fetal life until young adulthood: the Generation R Study. Paediatr Perinat Epidemiol 2004 Jan;18(1):61-72.

20. Jaddoe VW, Mackenbach JP, Moll HA, Steegers EA, Tiemeier H, Verhulst FC, et al. The Generation R Study: Design and cohort profile. Eur J Epidemiol 2006;21(6):475-84.

21. Centre for Research and Statistics RC. http://www.cos.rotterdam.nl; 2005.

22. Campbell S, Warsof SL, Little D, Cooper DJ. Routine ultrasound screening for the prediction of gestational age. Obstet Gynecol 1985 May;65(5):613-20.

23. Taipale P, Hiilesmaa V. Predicting delivery date by ultrasound and last menstrual period in early gestation. Obstet Gynecol 2001 Feb;97(2):189-94.

24. Tunon K, Eik-Nes SH, Grottum P. A comparison between ultrasound and a reliable last menstrual period as predictors of the day of delivery in 15,000 examinations. Ultrasound Obstet Gynecol 1996 Sep;8(3):178-85.

25. Verburg BO, Steegers EAP, de Ridder MAJ, Snijders RJM, Hofman A, Moll HA, et al. New charts for ultrasound dating of pregnancy and assessment of fetal growth, longitudinal data from a population-based cohort study. Ultrasound Obstet Gynecol 2007;Accepted for publication.

26. Hadlock FP, Deter RL, Harrist RB, Park SK. Fetal abdominal circumference as a predictor of menstrual age. AJR Am J Roentgenol 1982 Aug;139(2):367-70.

27. Hadlock FP, Harrist RB, Deter RL, Park SK. Fetal femur length as a predictor of menstrual age: sonographically measured. AJR Am J Roentgenol 1982 May;138(5):875-8.

28. Shepard M, Filly RA. A standardized plane for biparietal diameter measurement. J Ultrasound Med 1982 May;1(4):145-50.

29. Jaddoe VW, Verburg BO, de Ridder M, Hofman A, Mackenbach JP, Moll HA, et al. Maternal Smoking and Fetal Growth Characteristics in Different Periods of Pregnancy: The Generation R Study. Am J Epidemiol 2007 Feb 28.

30. Bernstein IM, Plociennik K, Stahle S, Badger GJ, Secker-Walker R. Impact of maternal cigarette smoking on fetal growth and body composition. Am J Obstet Gynecol 2000 Oct;183(4):883-6.

31. Statistics Netherlands. Allochtonen in Nederland. Voorburg/Heerlen; 2004.

32. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. Int J Epidemiol 1999 Oct;28(5):964-74.

33. Schild RL, Sachs C, Fimmers R, Gembruch U, Hansmann M. Sex-specific fetal weight prediction by ultrasound. Ultrasound Obstet Gynecol 2004 Jan;23(1):30-5.

34. Richardus JH, Graafmans WC, Verloove-Vanhorick SP, Mackenbach JP. Differences in perinatal mortality and suboptimal care between 10 European regions: results of an international audit. Bjog 2003 Feb;110(2):97-105.

35. Platt RW, Ananth CV, Kramer MS. Analysis of neonatal mortality:is standardizing for relative birth weight biased? BMC Pregnancy Childbirth 2004 Jun 4;4(1):9.

36. Graafmans WC, Richardus JH, Borsboom GJ, Bakketeig L, Langhoff-Roos J, Bergsjo P, et al. Birth weight and perinatal mortality: a comparison of "optimal" birth weight in seven Western European countries. Epidemiology 2002 Sep;13(5):569-74.

37. Bai J, Wong FW, Bauman A, Mohsin M. Parity and pregnancy outcomes. Am J Obstet Gynecol 2002 Feb;186(2):274-8.

38. Raymond EG, Cnattingius S, Kiely JL. Effects of maternal age, parity, and smoking on the risk of stillbirth. Br J Obstet Gynaecol 1994 Apr;101(4):301-6.

9

General discussion

During the last decade, several prediction models are presented in the medical literature to predict growth response to growth hormone (GH) treatment. In the statistical literature, progress is made in methodologies used to develop and validate prediction models. The general aim of the studies presented in this thesis was to join knowledge from these two fields. In this way, at one hand, we investigated how growth response prediction models can be improved and at the other hand, we explored how newly developed methods for modelling could be applied to predict growth response.

This chapter provides a more general discussion of the main findings, considers general methodological issues and gives suggestions for further research.

Validation and calibration of prediction models for growth response

Several prediction models have been developed in order to predict the growth response to GH treatment in children with short stature due to GH deficiency (GHD) or other causes (1-10). It is well-known that prediction models often perform less well than expected when applied to new patients.

We validated a widely used prediction model for first-year response to GH therapy in children with idiopathic GHD (4). As validation data, we used data of 136 children registered in the National Registry of Growth Hormone Treatment in Children by the Dutch Growth Foundation. Children were selected according to the same in- and exclusion criteria as were used for the cohort on which the prediction model was developed.

The observed height velocities (HV) in the first year of GH treatment of the cases in the validation group were plotted versus the predicted values. The regression line through the points in this plot (called a calibration plot) was fitted (11). The line had a slope of 0.808 and was significantly different from the line of identity, where observed values are equal to predicted values. This indicated underestimation of the lowest predictions and overestimation of the highest predictions. This phenomenon, called overfitting or overoptimism of a prediction model, is common when the selection of predictors and the estimation of their coefficients in the model are both guided by the same data set (12, 13). The predictions obtained by a model that suffers from overfitting can be corrected using a simple formula that can be derived from the equation of the regression line in the calibration plot (11). This formula defines a modified, calibrated prediction model. Compared to the original predictions, the predictions obtained by the calibrated model are shrunk towards the mean. In

our validation study, the corrections ranged from +0.92 cm/yr, if the predicted HV was 6.4 cm/yr (the lowest prediction in our validation group), to 1.2 cm/yr, if the predicted HV was 17.3 cm/yr (highest prediction). Corrections for predictions of in-between values were smaller, and for a predicted HV of 11.2 cm/yr the correction term was zero.

Subsequently, we obtained the standard deviation (SD) of the residuals (observed minus predicted value) of the validation data. As expected, this SD was higher than the SD that was reported for the cohort used for the development of the model (4). It is well-known that a prediction model is less precise for new data than for the data on which it was developed. Another important finding was the dependency of the SD of the residuals on the magnitude of the prediction. We found that the residual SD was ranging from 1.1 cm/yr for a predicted HV of 7.3 cm/yr (lowest prediction after correction) to 3.6 cm/yr when the predicted HV was 16.1 cm/yr (highest prediction after correction). This indicated that the higher a predicted HV, the less its accuracy. We gave the formula to calculate the residual SD depending on the value of the prediction. This SD should be used to calculate the prediction interval when a prediction of first-year HV is desired.

The validity of the calibrated model that was obtained by the analysis presented in chapter 2 was checked in a third data set. Patient data (n = 226) were selected from another registration database. Predictions of first-year HV were calculated using the original prediction model as well as the calibrated model and the performance of both models was investigated.

The analysis showed that, using the original prediction model, the regression line in the calibration plot was significantly different from the line of identity, indicating overfitting. In the calibration plot of the calibrated model, however, this difference was not significant.

Thus, this study confirmed that, in an independent cohort of children, the calibrated prediction model, presented in Chapter 2, did not show overfitting, in contrast to the original prediction model.

Previously reported prediction models for growth response were validated but none was adjusted based on the results of the validation. None of the models was validated according to the method we described in chapters 2 and 3. The issue of overfitting was not addressed. Only for the model presented by Albertsson-Wikland et al. (1), it was mentioned that prevention of overfitting was part of their method.

Conclusion: Overfitting was present in a prediction model for first-year growth response to GH treatment. Using the results from the validation, a calibrated model was obtained. Predictions calculated with this calibrated model are corrected for

overfitting and are less extreme compared to the original predictions. Validation of both the original and the calibrated model in a third data set confirmed the improvement of the model after calibration. These two studies showed that this method of validation and calibration is well applicable and improves the quality of prediction models.

Since almost all growth prediction models currently used in practice are not corrected for overfitting, one should take care in using them, in particular if the resulting prediction is relatively large of small.

Prediction of adult height in GH treated children with GHD

Models to predict long-term growth response to GH treatment in children with GHD did not exist. We therefore developed models predicting adult height (AH). The models are described in chapter 4. It is desirable to have a prediction of AH before starting GH treatment, in order to get information about the long-term results of the treatment. Application of the prediction model (Start model) will give a patient his or her individual predicted value along with a prediction interval. It can identify patients with a high or low chance to benefit from GH treatment. Prior to our study, several studies found that the growth response in the first year of treatment is highly predictive for the response in the second year (4, 14) and in the third and fourth year (4, 14). Therefore, we also developed prediction models using the patient information available after one year of treatment (First-year model).

For the development of the models, we used advanced statistical methods. This included dealing with missing data (15, 16), flexible modelling of the relationships between continuous determinants and the outcome (17), checking the significance of pre-specified interactions and correction for overfitting (18). Because of the large influence of puberty on growth, separate models were needed for prepubertal and for pubertal children. The data used in this study came from the National Registry of Growth Hormone Treatment in Children by the Dutch Growth Foundation. In total, 342 patients were selected with a diagnosis of GHD or with a maximum GH response during provocation tests of less than 11 ng/ml. They should have been treated with biosynthetic GH for at least one year and should be at an age that guaranteed AH had been reached. Stepwise forward selection was used to determine the predictors in the models. For continuous predictors, we tested whether adding a transformation of the variable improved the model significantly (19). Next, the significance of some pre-specified interactions was tested.

As expected, the four prediction models included height SDS (at start or after the first year of GH treatment) and target height (TH) SDS as most important predictive factors. In the models for prepubertal children, we found a positive effect of female gender and of multiple pituitary hormone deficiencies. The negative effect of bone age in the Start model for prepubertal children shows that start of GH treatment at a younger age gives a higher growth response. The positive coefficients for bone age delay (chronological age minus bone age) in the other three models reflect that children with delayed bone age have more growth potential. Maximum response to GH provocation tests is included in the Start model for prepubertal children (negative effect), but is not significant anymore in the First-year model. A similar finding was reported by Cole et al. (14). Both First-year models include change in height SDS during the first year, which is in line with several other studies (8, 10, 20, 21).

In none of the models, GH dose was found to be a significant predictor variable. First of all, it should be noted that the data on which the models were developed were not suitable to estimate a dose effect of GH, as this can only be done in data from randomised clinical trials. Our data came from a registration database. The GH dose at start and during treatment was determined by the treating physician. The criteria used for dose assessments were unknown. It is possible that higher initial doses were prescribed to patients with supposed worse prospects, like children close to or entering puberty, children with only mild GHD or children with very short parents. During treatment, a change of GH dose might have been guided by the obtained growth response. Even if any GH dose effect would have been found, this could not be interpreted as a causal effect of GH dose on AH. Furthermore, although a significant positive effect of GH dose on short-term growth response is often found (4, 22, 23), a dose-effect on long-term response is less established (8, 20, 24). It might be that the short-term dose effect wanes over the years because other influences become more important. Or it might be that a high dose of GH treatment causes more acceleration of bone maturation on the long term, leading to earlier cessation of growth.

We performed internal validation of the models by bootstrapping (25). This is a recently developed advanced statistical method to correct for overfitting. It is already a well-established method in biostatistics, while practical applications in medical research are still quite rare, especially in the field of growth research. To the best of our knowledge, we were the first to apply this method in the field of growth hormone research.

This is a well-established statistical method to come to a prediction model corrected for overfitting. It results into a prediction formula in which the coefficients

are shrunk towards zero compared to the uncorrected model. It provides corrected values for the explained variance (R^2) and for the residual SD. With these corrections, the true performance of the model in new data is given.

For our models, the corrected percentages of explained variance in AH SDS ranged from 37% (prepubertal group, Start model) to 60% (pubertal group, First-year model). Given the fact that the outcome AH is usually several years ahead from the moment of prediction, it was reasonable that a substantial part of the variance remained unexplained. During GH treatment, several factors, which are unknown at the time of prediction, may influence the growth of a child.

One reason to develop the First-year models was to investigate if such a model could provide a criterion for the first-year growth response required to attain an AH in the normal range. Indeed, our study showed that change in height SDS during the first year of treatment is highly related to the attained AH. This might indicate that, as a positive correlation between GH dose and first-year response is well established (4, 14, 26), one might tend to give a high GH dose in order to increase the first-year response, with the idea that this will subsequently increase AH. However, we found an inverse relation between the GH dose in the first year and AH, although this did not reach significance. This indicates that a first-year growth response obtained by a high GH dose will give a lower predicted AH than the same response obtained by a low GH dose. The First-year models can be used as a response criterion for first-year growth response only when the GH dose given during the first year is determined in the same way as done for the patients in our data set and when the GH dose is in the same range.

Conclusion: We developed models for AH prediction in children with GHD treated with GH. Separate models for prepubertal and pubertal children were developed, and models to obtain a prediction at start of treatment as well as models that provide a prediction using the available data after one year of treatment. These models can be a useful tool for decisions about GH treatment of children with GHD.

Prediction of adult height in GH treated SGA children

For children born SGA with persistent short stature, for which GH treatment is considered, it is very useful to get some information about the expected AH for an individual child and how this expected AH is related to the dosage of GH treatment. In chapter 5, we presented a model developed to predict height SDS at onset of puberty and AH.

We used data of 150 SGA children who were involved in two clinical trials concerning GH treatment. Children were selected who were prepubertal for at least one year after start of GH treatment and had started or completed puberty. Furthermore, birth year had to be prior to a specified year, to avoid overrepresentation of children with early puberty. We developed a model for the outcomes height SDS at onset of puberty and AH SDS, using mixed effects repeated measurements analysis.

The determinants for AH SDS were height SDS at start, TH SDS, GH dose, bone age delay at start (positive effects), and IGFBP-3 SDS at start (negative effect). A significant interaction was found between GH dose and IGFBP-3 SDS, with a negative coefficient. This means that the effect of a higher dose depends on the level of IGFBP-3 of the child: a lower IGFBP-3 SDS at start is related to a larger GH dose-effect. For a child with an IGFBP-3 of -2 SDS, the expected difference in AH is 0.65 SDS (approximately 4.6 cm) when treated with 0.067 mg GH / kg·day compared to 0.033 mg GH / kg·day. For a child with an IGFBP-3 of 0 SDS, this difference is only 0.11 SDS (0.8 cm).

For height SDS at onset of puberty, the model explained 57% of the variance and for AH SDS 41%. These percentages gave a realistic presentation of the performance of the model because we did correct for overfitting, as explained before.

When this prediction model is used in deciding on the GH dose to be prescribed, one has to determine the objective of GH treatment. This objective can be defined in terms of the absolute AH, for which the goal will generally be an AH above -2 SDS. The TH range, usually defined as the TH +/- 1.3 SDS, could also be taken as goal. Thirdly, a substantial gain in height SDS during treatment might be regarded as a reasonable goal. When, for an individual child, the predicted AH is unsatisfactory in case of a GH dose of 0.033 mg/kg·day, a higher dose might be considered.

Conclusion: We developed a prediction model for long-term response to GH treatment in short children born SGA. The model included GH dose as determinant and the effect of GH dose was depending on the level of serum IGFBP-3. The model can be helpful in the decision about the optimal GH dose for a child. Because the data set used in this study was of limited size, it could be very useful to perform a meta-analysis combing data of several randomised clinical trials. This could possibly provide more information about the GH dose effect and its relation to certain patient characteristics.

Spontaneous growth of SGA children without early catch-up growth

Eighty-five percent of children born SGA will show catch-up growth to a height in the normal range before 2 years of age. In the other 15%, there are some children with late catch-up growth. It would be very informative if, in an early stage, it could be predicted which children will have this catch-up growth and which children will not.

In chapter 6, we presented a study of the growth of SGA children with a height still below -2 SDS at 2 years of age. These 97 children came from a cohort of 724 SGA children, of whom the growth during the first two years of life was previously described (27). We found that 39% of these 97 children (6% of the original cohort) had catch-up growth to a normal height between 2 and 8 years of age. While during the first two years of life the percentage children with catch-up growth was highest in infants born full-term, in the period after 2 years of age, the percentage was highest in infants born preterm.

Furthermore, we developed a model to predict the spontaneous growth of SGA children after 2 years of age using the prepubertal height measurements of these children between 2 and 8 years of age. Predicting factors in this model were the difference between height SDS at the age of 2 years and TH SDS, gender of the child, multiple birth and the difference between height SDS at the age of 2 years and birth length SDS. We showed how this model can be used to predict the height SDS at 8 years of age when a child is 2 years of age. In addition, the model provides the probability that, at 8 years of age, a height in the normal range will be reached.

A larger data set may provide a prediction model with higher accuracy. However, long-term follow-up of untreated SGA children becomes increasingly scarce as many SGA children are nowadays treated with GH. Good registration of pre-treatment growth data of SGA children might be useful.

Conclusion: Catch-up growth to a normal height occurred in 91% of SGA children, in 6% between 2 and 8 years of age. The difference between height SDS at 2 years of age and TH SDS was the most important determinant. The presented prediction model can identify children with low or high probability of catch-up growth to a normal height. This may assist to determine which children require medical follow-up.

Determining mean adult height from incomplete follow-up data

In long-term growth studies with AH as outcome, often reporting is required while data are still incomplete because some participants have not yet reached AH whereas others might be lost to follow-up. Until now, there has not been any effort to analyse incomplete follow-up of growth data in a more sophisticated way than simply using the cases that have reached AH and disregarding the other cases. Also when reporting is done for patients from e.g. a registry, usually the same procedure is followed: cases that have reached AH are selected. However, in these situations, the subgroup used for analysis is not a representative selection of the complete group. This can give biased results because the AHs of the patients who did reach AH earlier in follow-up might be different from the AHs of the other patients. Using only the observed AHs also ignores all observed follow-up of the patients that did not yet reach AH, despite the fact that these data contain valuable information.

In chapter 8, we introduced a more advanced method to estimate the mean AH of a group with partly incomplete follow-up data. The method was originally developed for cost-effectiveness analyses (28). We illustrated the validity of the method for growth data by applying it to artificially generated growth data in a simulation process. In that situation, the true mean and standard deviation (SD) of AH in the population is known. Furthermore, we showed results of the new method applied to two data sets. One data set was selected from the National Registry of Growth Hormone Treatment in Children by the Dutch Growth Foundation. Inclusion criteria were birth year and the availability of growth follow-up until AH. We censored the follow-up data artificially. We compared the mean AH obtained from the complete follow-up with the estimated means using the censored follow-up, based on the commonly used method as well as based on the presented new method. The other data set consisted of 39 children participating in a clinical trial and treated with biosynthetic GH until AH. These represented a realistic situation of a growth study. We computed the estimated mean AH for males and females using the data known at time points with two-year intervals during the last years of follow-up of the study. Again we compared the results with the estimated mean AH using the complete data.

We showed that the new method gives estimated means with only very small bias, even in case of correlation between AH and the age of reaching AH. The efficiency of

143

the method was higher than currently used methods, reflected in a smaller standard deviation of the estimations.

The method will give valid results, provided that eventual loss to follow-up is not related to the outcome AH. In practice, this assumption will be fulfilled, since almost all censoring is due to the patients not having reached AH yet at the time of the analysis. If drop-out is due to disappointing growth, the method cannot correct for this type of loss, nor can any other method. The method can be carried out in standard statistical software.

Conclusion: Incompleteness of follow-up data in growth studies is until now more or less neglected. The method we presented provides valid mean AH estimations if complete actual measurements of AH are not (yet) available, provided that and loss to follow-up is not related to growth. While in this method also the observed data of study-participants with incomplete follow-up data are used, it makes more effective use of the data collected. More research on the analysis of incomplete growth follow-up data would be worthwhile.

Individually customised fetal weight charts derived from ultrasound measurements

While normal fetal growth is influenced by several characteristics of the fetus and its parents (29-35), fetal size should be evaluated taking into account important physiological characteristics, like gender, parental anthropometrics and ethnicity. Individually customised fetal growth curves are in fact reference growth charts that are adjusted for important characteristics. Use of these charts makes it possible to identify pathological smallness instead of constitutional small size with normal intrauterine growth. This can prevent unnecessary classification as growth restriction (36-39). On the other hand, studies have shown that failure to detect fetal growth restriction is an important reason for suboptimal perinatal care (40).

In chapter 7, we presented a model that can be used to construct individually customised fetal growth charts. For the development of the model, we used data from the Generation R Study, a prospective, population-based cohort study. We included ultrasound measurements, taken repeatedly during pregnancy, of 5,473 pregnancies.

Our final model for estimated fetal weight between 18 and 36 weeks of gestational age included as determinants fetal gender, parity, ethnicity, maternal height, pre-pregnancy weight, age of the mother and smoking. The effects of the determinants all increased with advancing gestational age. We evaluated the fetal

sizes measured in the third trimester of pregnancy (gestational age ≥ 27 weeks). In the group of fetuses classified as too small (≤ -1.28 SDS) using the unadjusted reference chart, 16% had a size in the normal range when individually customised reference charts were used.

Our study was the first one fitting a model for customised fetal growth charts using observations of estimated fetal weight obtained by ultrasound measurements in a large population-based prospective study.

It is not obvious which factors to take into account when evaluating fetal growth. Adjustment has to be done only for physiological factors, but some factors may represent both physiological as well as pathological effects. For example, ethnicity reflects constitutional growth potential, but may also represent differences in feeding habits and other life-style factors. Studies have shown that in ethnic groups with mean birth weight lower compared to other groups, the optimal birth weight, defined as birth weight with the lowest perinatal mortality, is also lower (41, 42). This points to the physiological aspect of ethnicity as determinant of fetal growth. Parity is another determinant of which inclusion in the model can be discussed. We adjusted for parity, because it is a determinant of the growth potential of a fetus. However, nulliparous women are at higher risk of obstetric complications (43) and stillbirths (44). We are not aware of any study relating perinatal mortality of first born children and second or later born siblings to their birth weight. Therefore, it is not clear whether nulliparity must be seen as a physiological determinant, or should be considered as a possible cause of intrauterine growth restriction. The parameterisation of our model makes it possible to disregard a determinant if one does not want to adjust for it.

The customised antenatal growth charts developed by Gardosi et al. (45) were based on a regression model for birth weight, fitted in a very large group of over 40,000 neonates. Some assumptions were needed to derive intrauterine growth charts from a calculated "optimal birth weight" for an infant. Our model is based on estimated fetal weights derived from ultrasound measurements. In clinical practice, fetal size is determined in the same way. Therefore, we think our curves suit better.

Conclusion: Our study showed that individually customised fetal growth charts can very well be derived from ultrasound measurements collected during pregnancy. It has been demonstrated in other studies that the use of customised growth charts may improve fetal growth monitoring and prenatal care. Our model, derived using data from a large population based study comprising several ethnic groups, will contribute to this.

Prediction models for growth response: methodological aspects

Imputation of missing values

Imputation of missing values, which enables to perform analyses on a total data set, is hardly done for growth analyses, although the advantages are well established. Usually, the analysis is done on the subset of patients with complete data. This can lead to considerable bias and loss of power. Imputation methods can avoid the disadvantages of such a complete cases analysis.

Imputation can be performed in an extensive way, using for example the procedure MICE (46), implemented in S-Plus and in Stata. With MICE, imputation can be performed in a multiple fashion using all relations between all variables in the data set. It is possible to specify which variables are categorical and to indicate which variables should be used as predictors for the imputation of others. Such an extensive procedure is worthwhile when a large data set can be imputed as a whole. The resulting set of imputed data sets (usually five) can be used for several analyses. In many cases, however, a data set is "dynamic". For instance, in a registration database, patients are added continuously and follow-up of the registered patients is extending throughout the years. In longitudinal clinical studies, follow-up is increasing and analyses with the available data will often be performed before the study has finished. This means that the data set is continuously growing and it does not suffice to impute the data set only once, as one extensive project.

The data sets we used in the studies presented in chapter 4, 5 and 6, were selections from larger databases, which will extend further in number of patients and/or in follow-up. We used the procedure Proc MI in SAS (47). This procedure assumes that all variables are normally distributed, and all can be correlated to each other. The correlations are estimated from the data and subsequently used for the imputation of the missings. We included all available variables that could hold information about variables with missings. The variables with missings in our data sets were all continuous variables. We checked the distribution of all variables and if needed a transformation was performed to get a normally distributed variable. The assumption of a normal distribution for a dichotomous variable like gender or for an ordinal variable, like severity of GHD is of course artificial, but is adequate for describing the relation with the continuous variables to be imputed.

Conclusion: Applying multiple imputation will avoid bias in results that can occur if only cases with complete data are selected. It will improve the quality of growth prediction models. For situations that are not too complex, a simple imputation method will satisfy.

Modelling continuous predictors

Using only straight-line relations between a continuous determinant and the outcome in a regression model might be insufficient. Though this is the first relation to be considered, another type of relation might fit significantly better. For example, for treatment with small doses of GH, each unit increase in dose might have a substantial effect on the growth response, while in the range of higher doses the effect of one unit increase might be much smaller. One way to model such a relation more flexibly is the use of polynomial models, in practice adding a quadratic term of the determinant. This is, however, often unsatisfactory. Another simple possibility is to investigate the relation using a transformation of the determinant, like the logarithm. Royston and Altman (17) proposed to test a range of transformations of the covariable for possible superiority over the straight line relation. Then, next to a first term, the significance of adding a second term could be tested. So their proposed method involves comparing several models, e.g. when considering two terms, 36 models should be evaluated.

Our analyses presented in chapters 4 and 5 included stepwise forward selection procedures, involving several continuous covariables. The selection procedures had to run automatically because they were repeated in the bootstrap procedure hundreds of times. Therefore we simplified the procedure of finding the best relation. If a continuous covariable was selected as determinant in a model, we tested the addition of a transformed term. In this way only a limited number of fractional polynomials with two terms were considered. In three of the models presented in chapter 4, quadratic terms of height SDS were included. The quadratic term was the term we tested first. In the model presented in chapter 5, no transformation terms turn out to be included. It might have been that the data set was too small to have sufficient power for a second term coming in.

Conclusion: Growth prediction models will improve if more flexible relations between continuous predictors and the outcome will be considered. We used an approach related to the theory on fractional polynomials. For these types of analyses, this was a good compromise between the theoretical and the practical optimum.

Interactions terms

In the development of growth prediction models, consideration of interaction terms is hardly done. One reason might be that this complicates the procedure. Another obstructing factor might be that interaction terms are often not well understood and sometimes criticized as artificial and incomprehensible. On the other hand, it is well recognised that there is heterogeneity in patient groups. Therefore, sometimes the size of the effect of a determinant depends on another determinant and it is not sufficient to estimate just one constant effect. For example, the effect of GH dose might be depending on the age of the child. In some situations, models are developed for different subgroups (48). However, the use of interaction terms is often much more efficient.

For the models presented in chapter 4, we tested the interactions between age (chronological age and bone age) and TH SDS, and between age and the change in height SDS during the first year. For the model in chapter 5, all interactions with GH dose were tested. These possible interactions were pre-specified based on clinical knowledge and not on the data.

Conclusion: Considering interaction terms and including them when relevant, will improve the quality of growth prediction models. With good explanation and illustration of the meaning of an interaction term, one will become more familiar with this extension of modelling.

Correction for overfitting

A growth prediction model developed on one specific data set should be corrected for overfitting. Only after such a correction, the model will give good predictions for other children and the predictive performance, indicated by the percentage of explained variability and the residual standard deviation, can be given correctly. Correction for overfitting of a linear regression model can be done either by internal validation, using bootstrapping (applied in chapters 4 and 5) or cross-validation, or by external validation (applied in chapter 2). A corrected model will give better predictions and will provide the actual accuracy of the prediction. For external validation of models, which is still very useful after internal validation and correction for overfitting, collaboration between research groups and organisations holding registries with data of GH-treated children, would be very helpful in order to obtain larger validation groups.

Conclusion: Checking the presence of overfitting and correction for it is necessary for any growth prediction model. Adopting this as common practice, the

knowledge about response to GH treatment will improve and so will the clinical care for children with growth impairment.

Concluding remarks

Application of prediction models for growth response

In the last decade, progression has been made in the area of prediction of growth response to GH treatment. With caution, the following suggestions can be given for the use of the models presented in this thesis:

In case a child with GHD is going to start GH treatment, it would be very informative to calculate the predicted HV during the first year of treatment, using the calibrated model (a correction of the KIGS model) presented in chapter 2, as well as the predicted AH, using the Start-model presented in chapter 4. Both predictions, with prediction interval, provide information for the patient, its parents and the treating physician, about the prospects on the short-term as well as on the long-term. The prediction of first-year HV might be used to adjust the dose of GH, in case the standard dose gives a low prediction. However, one should be reluctant to give high doses. After one year of treatment, the first-year HV should be compared with its prediction. If the observed HV is substantially lower than the prediction, one should trace possible causes, like compliance and somatic or psychological factors influencing growth. Next, a new prediction of AH can be calculated using the First-year model presented in chapter 4. Taking into account this new prediction for AH, one might decide about the continuation of GH treatment or adjustment of the dose. A higher dose might result in a higher growth response, but unfortunately we are not yet able to specify a quantitative dose effect. Also, it is likely that the effect of a higher dose depends on patient characteristics, for example age or the maximum response to GH provocation tests. When the patient has reached AH, it is interesting to compare the observed AH and the prediction and to look for explanations. This might provide information for the clinical care of future patients.

For children born SGA with a height still below -2 SDS at 2 years of age, a prediction, along with prediction interval, of the height at 8 years of age should be calculated using the model presented in chapter 6. It might be that this prediction does not provide very specific information, because the predicted value is rather close to -2 SDS, so the prediction interval is more or less half in the normal range and half below. However, for some children, the prediction interval will be for the larger part, or as a whole, below the -2 SDS, indicating a very low chance that a normal height will be reached. This information can be given to the parents of the

child, and might have consequences for the clinical care. For some others, the prediction interval will mainly be in the normal range. The frequency of follow-up visits might be decreased for these children. When a short child born SGA starts GH treatment, predictions of height at start of puberty and of AH can now be calculated (chapter 5). Especially the prediction of AH should be used to determine whether a dose of 0.033 mg GH / kg·day (= 1 mg/m^2·day) is sufficient. Otherwise, the required dose can be calculated using the provided formula, with a maximum of a 0.066 mg GH / kg·day.

It is still the case that prediction models are an underused tool in clinical practice. One of the causes is that it takes some effort to compute a prediction, e.g. because patient characteristics have to be converted using reference data. It may be worthwhile to provide easy to use tools for the calculation of predictions. For patients registered in a central database, the treating physician can be provided with predictions for short-term as well as long-term response. Well-developed prediction models will certainly provide useful information. Regular use of prediction models and comparing predictions and observations of patients will in the long run show the value of a prediction model.

Recommendations for future research

The lack of evidence about the size of the GH dose effect on AH in children with GHD might be one of the challenges for the coming years. Only a few randomised clinical trials on GH dose in children with GHD(22, 23) have been published, and new trials, covering the whole period until AH will not easily be set up. It may be feasible, to set up a study in which randomisation into two or more doses is done based on the prediction of AH. For example, only patients with a prediction below a certain level could be randomised into a group receiving the standard dose and a group receiving a higher dose. The other patients could receive the standard dose. Trying to estimate the causal effect of the GH dose from non-randomised data sets is not impossible but very complicated and needs very large data sets with sufficient follow-up measurements.

For the treatment of SGA children, we were able to estimate a dose effect, because we could use data of patients that were randomised into two doses of GH treatment (chapter 5). We found that the dose effect depends on the level of serum IGFBP-3. We realise, that the size of our sample was limited and recommend further analyses on larger data sets.

Future research in larger patient groups may identify more significant determinants of AH in GH treated children. Continuous determinants may be

modelled more flexibly and more interactions may be detected. In addition, it can be expected that information on genetic characteristics of children will contribute to better prediction (49).

For the development of growth prediction models, data of sufficient quantity and quality are needed. Data of children using GH treatment are included in registries of some pharmaceutical industries and some national registries, like in the Dutch National Registry of Growth Hormone Treatment in Children. Although these registration databases comprise heterogeneous groups, the data sets are often large and they do reflect clinical practice. Detailed information on patient characteristics is preferable to provide high quality analyses. Collection and registration of data require financing, but the results of the analyses will improve our knowledge and will lead to recommendations that might save money.

For indications for which GH treatment is approved, it will be more difficult and in some situations unethical to set up clinical trials with randomisation of patients into different treatment doses. However, studies could be set up for subgroups of patients, for which there is no consensus about the optimal dosing. Another possibility is to randomise the patients into different dosage strategies, for example constant dosing throughout the treatment period versus adjusting the dosage depending on the growth response.

In conclusion, the recommendations of this thesis may serve to contribute more to the progress which has already been made during the last decade, in the prediction of growth response to GH treatment. However, many challenges still exist. The improvement of prediction models for growth response will be attributable to the greater availability of genetic and biochemical data, coupled with advanced methods for the development of such models.

References

1. Albertsson-Wikland K, Kristrom B, Rosberg S, Svensson B, Nierop AFM. Validated Multivariate Models Predicting the Growth Response to GH Treatment in Individual Short Children with a Broad Range in GH Secretion Capacities. Pediatric Research 2000;48(4):475-84.

2. Dahlgren J, Wikland KA. Final height in short children born small for gestational age treated with growth hormone. Pediatr Res 2005 Feb;57(2):216-22.

3. Carel JC, Ecosse E, Nicolino M, Tauber M, Leger J, Cabrol S, et al. Adult height after long term treatment with recombinant growth hormone for idiopathic isolated growth hormone deficiency: observational follow up study of the French population based registry. Bmj 2002 Jul 13;325(7355):70.

4. Ranke MB, Lindberg A, Chatelain P, Wilton P, Cutfield W, Albertsson-Wikland K, et al. 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. KIGS International Board. Kabi Pharmacia International Growth Study. J Clin Endocrinol Metab 1999;84(4):1174-83.

5. Ranke MB, Lindberg A, Chatelain P, Wilton P, Cutfield W, Albertsson-Wikland K, et al. Prediction of long-term response to recombinant human growth hormone in Turner syndrome: development and validation of mathematical models. KIGS International Board. Kabi International Growth Study. J Clin Endocrinol Metab 2000 Nov;85(11):4212-8.

6. Ranke MB, Lindberg A, Cowell CT, Wikland KA, Reiter EO, 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). J Clin Endocrinol Metab 2003 Jan;88(1):125-31.

7. Ranke MB, Lindberg A, Martin DD, Bakker B, Wilton P, Albertsson-Wikland K, et al. The mathematical model for total pubertal growth in idiopathic growth hormone (GH) deficiency suggests a moderate role of GH dose. J Clin Endocrinol Metab 2003 Oct;88(10):4748-53.

8. Reiter EO, Price DA, Wilton P, Albertsson-Wikland K, Ranke MB. Effect of Growth Hormone (GH) Treatment on the Final Height of 1258 Patients with Idiopathic GH Deficiency: Analysis of a Large International Database. J Clin Endocrinol Metab 2006 Mar 14.

9. Schonau E, Westermann F, Rauch F, Stabrey A, Wassmer G, Keller E, et al. A new and accurate prediction model for growth response to growth hormone treatment in children with growth hormone deficiency. Eur J Endocrinol 2001;144(1):13-20.

10. Thomas M, Massa G, Bourguignon JP, Craen M, De Schepper J, de Zegher F, et al. Final height in children with idiopathic growth hormone deficiency treated with recombinant human growth hormone: the Belgian experience. Horm Res 2001;55(2):88-94.

11. van Houwelingen JC. Validation, calibration, revision and combination of prognostic survival models. Stat Med 2000;19(24):3401-15.

12. Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med 2000;19(4):453-73.

13. Chatfield C. Model uncertainty, data mining and statistical Inference. Journal of the Royal Statistical Society, Series A 1995;158(3):419-66.

14. Cole TJ, Hindmarsh PC, Dunger DB. Growth hormone (GH) provocation tests and the response to GH treatment in GH deficiency. Arch Dis Child 2004 Nov;89(11):1024-7.

15. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York; 1987.

16. Schäfer JL. Analysis of Incomplete Multivariate Data. New York; 1997.

17. Royston P, Altman DG. Regression using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling. Applied Statistics 1994;43(3):429-67.

18. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15(4):361-87.

19. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. Int J Epidemiol 1999 Oct;28(5):964-74.

20. Cutfield W, Lindberg A, Albertsson Wikland K, Chatelain P, Ranke MB, Wilton P. Final height in idiopathic growth hormone deficiency: the KIGS experience. KIGS International Board. Acta Paediatr Suppl 1999 Feb;88(428):72-5.

21. Blethen SL, Baptista J, Kuntze J, Foley T, LaFranchi S, Johanson A. Adult height in growth hormone (GH)-deficient children treated with biosynthetic GH. The Genentech Growth Study Group. J Clin Endocrinol Metab 1997 Feb;82(2):418-20.

22. Cohen P, Bright GM, Rogol AD, Kappelgaard AM, Rosenfeld RG. Effects of dose and gender on the growth and growth factor response to GH in GH-deficient children: implications for efficacy and safety. J Clin Endocrinol Metab 2002 Jan;87(1):90-8.

23. de Muinck Keizer-Schrama S, Rikken B, Hokken-Koelega A, Wit JM, Drop S. Comparative effect of two doses of growth hormone for growth hormone deficiency. The Dutch Growth Hormone Working Group. Arch Dis Child 1994 Jul;71(1):12-8.

24. Wit JM, Kamp GA, Rikken B. Spontaneous growth and response to growth hormone treatment in children with growth hormone deficiency and idiopathic short stature. Pediatr Res 1996 Feb;39(2):295-302.

25. Efron B. TR. An introduction to the bootstrap. New York; 1993.

26. de Muinck Keizer-Schrama SM, Rikken B, Wynne HJ, Hokken-Koelega AC, Wit JM, Bot A, et al. Dose-response study of biosynthetic human growth hormone (GH) in GH-deficient children: effects on auxological and biochemical parameters. Dutch Growth Hormone Working Group. J Clin Endocrinol Metab 1992 Apr;74(4):898-905.

27. Hokken-Koelega AC, de Ridder MA, Lemmen RJ, den Hartog H, de Muinck Keizer-Schrama SM, Drop SL. Children born small for gestational age: do they catch up? Pediatr Res 1995 Aug;38(2):267-71.

28. Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. Biometrics 1997 Jun;53(2):419-34.

29. Gardosi J, Mongelli M, Wilcox M, Chang A. An adjustable fetal weight standard. Ultrasound Obstet Gynecol 1995 Sep;6(3):168-74.

30. Johnsen SL, Rasmussen S, Wilsgaard T, Sollien R, Kiserud T. Longitudinal reference ranges for estimated fetal weight. Acta Obstet Gynecol Scand 2006;85(3):286-97.

31. Jacquemyn Y, Sys SU, Verdonk P. Fetal biometry in different ethnic groups. Early Hum Dev 2000 Jan;57(1):1-13.

32. Mongelli M, Gardosi J. Longitudinal study of fetal growth in subgroups of a low-risk population. Ultrasound Obstet Gynecol 1995 Nov;6(5):340-4.

33. Pang MW, Leung TN, Sahota DS, Lau TK, Chang AM. Customizing fetal biometric charts. Ultrasound Obstet Gynecol 2003 Sep;22(3):271-6.

34. Schwarzler P, Bland JM, Holden D, Campbell S, Ville Y. Sex-specific antenatal reference growth charts for uncomplicated singleton pregnancies at 15-40 weeks of gestation. Ultrasound Obstet Gynecol 2004 Jan;23(1):23-9.

35. Wilcox MA, Newton CS, Johnson IR. Paternal influences on birthweight. Acta Obstet Gynecol Scand 1995 Jan;74(1):15-8.

36. de Jong CL, Francis A, van Geijn HP, Gardosi J. Fetal growth rate and adverse perinatal events. Ultrasound Obstet Gynecol 1999 Feb;13(2):86-9.

37. de Jong CL, Francis A, van Geijn HP, Gardosi J. Customized fetal weight limits for antenatal detection of fetal growth restriction. Ultrasound Obstet Gynecol 2000 Jan;15(1):36-40.

38. Mongelli M, Gardosi J. Reduction of false-positive diagnosis of fetal growth restriction by application of customized fetal growth standards. Obstet Gynecol 1996 Nov;88(5):844-8.

39. Bukowski R. Fetal growth potential and pregnancy outcome. Semin Perinatol 2004 Feb;28(1):51-8.

40. Richardus JH, Graafmans WC, Verloove-Vanhorick SP, Mackenbach JP. Differences in perinatal mortality and suboptimal care between 10 European regions: results of an international audit. Bjog 2003 Feb;110(2):97-105.

41. Platt RW, Ananth CV, Kramer MS. Analysis of neonatal mortality:is standardizing for relative birth weight biased? BMC Pregnancy Childbirth 2004 Jun 4;4(1):9.

42. Graafmans WC, Richardus JH, Borsboom GJ, Bakketeig L, Langhoff-Roos J, Bergsjo P, et al. Birth weight and perinatal mortality: a comparison of "optimal" birth weight in seven Western European countries. Epidemiology 2002 Sep;13(5):569-74.

43. Bai J, Wong FW, Bauman A, Mohsin M. Parity and pregnancy outcomes. Am J Obstet Gynecol 2002 Feb;186(2):274-8.

44. Raymond EG, Cnattingius S, Kiely JL. Effects of maternal age, parity, and smoking on the risk of stillbirth. Br J Obstet Gynaecol 1994 Apr;101(4):301-6.

45. Gardosi J. New definition of small for gestational age based on fetal growth potential. Horm Res 2006;65 Suppl 3:15-8.

46. van Buuren S, Oudshoorn CGM. Multivariate imputation by chained equations. MICE V1.0 User's Manual. Leiden: TNO Preventie en Gezondheid; 2000. Report No.: PG/VGZ/00.038.

47. SAS. version 9.1. SAS Institute, 2002 , Cary, NC USA.

48. Ranke MB, Lindberg A, Albertsson-Wikland K, Wilton P, Price DA, Reiter EO. Increased response, but lower responsiveness, to growth hormone (GH) in very young children (aged 0-3 years) with idiopathic GH Deficiency: analysis of data from KIGS. J Clin Endocrinol Metab 2005 Apr;90(4):1966-71.

49. Clayton PE, Whatmore AJ. The Genomic Approach to Growth Prediction. Hormone Research 2007;67(Suppl. 1):10-5.
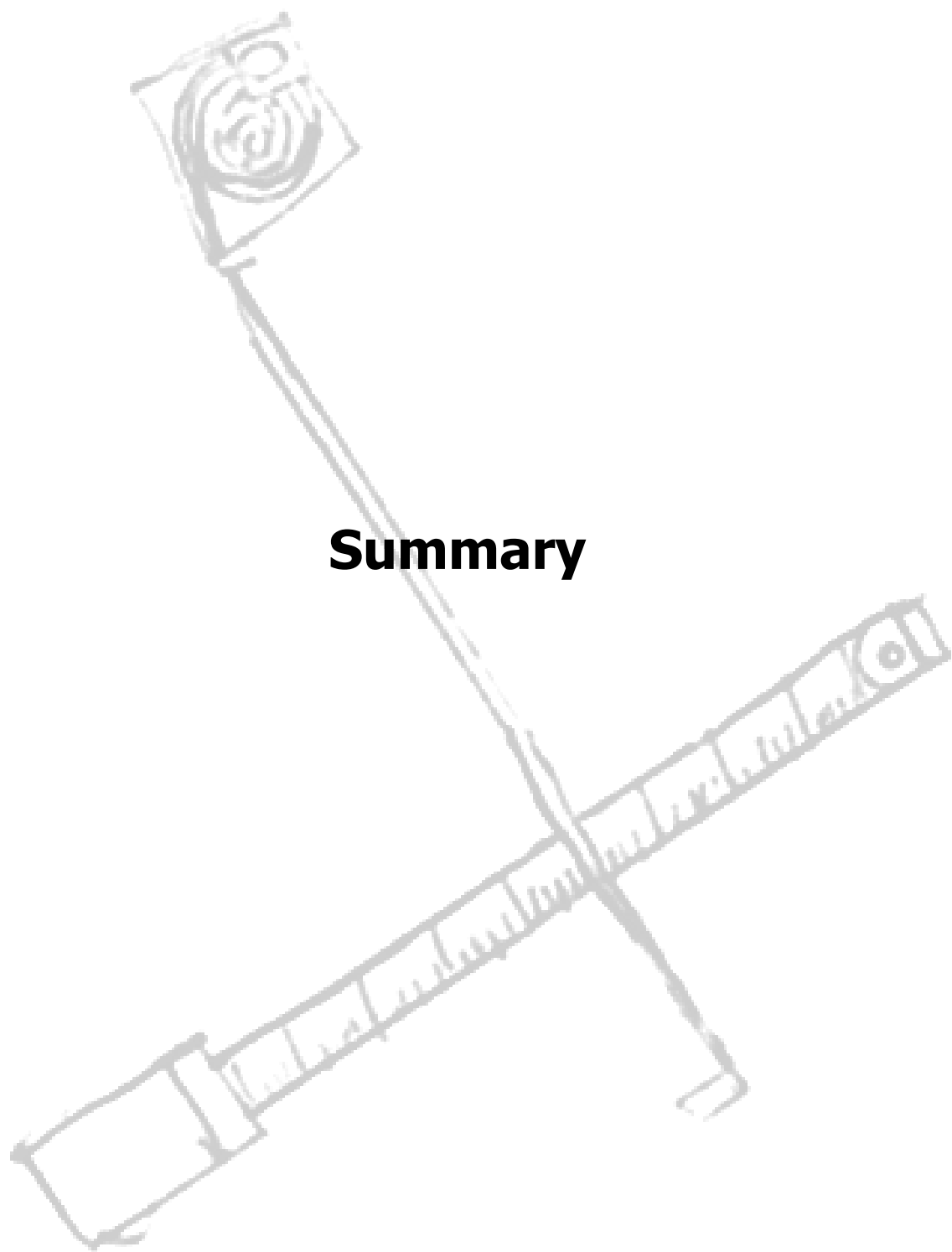
Summary

In **Chapter 1**, the background of the topics of this thesis is presented. Some aspects of growth, prenatal as well as postnatal, and growth regulation are described. The studies in this thesis concern two groups of children with impaired growth, namely children with growth hormone deficiency (GHD) and children born short for gestational age (SGA). Some information about these patient populations is given, including treatment with growth hormone (GH). The current state of art with regard to prediction models for growth response to GH treatment is described.

Chapter 2 describes validation and calibration of a model published in 1999, for prediction of growth in children with idiopathic growth hormone deficiency (IGHD) during GH therapy. This model, predicting growth in the first year of GH therapy, was validated using data of 136 Dutch children with IGHD. We plotted the observed height velocities (HV) of the validation group versus the predicted HV in a calibration plot. A regression analysis was done with observed HV as outcome. The regression line was significantly different from the line of identity (where observed HV is equal to predicted HV) and the estimated slope was 0.808. This indicated that the model suffered from overfitting, a well-known phenomenon in the context of prediction models. Therefore, the estimated coefficients from the regression analysis were used to determine a calibration correction for the model. This correction was described by the formula: $Y_{cal} = Y_{orig} + (2.153 - 0.192 * Y_{orig})$, where Y_{cal} is the calibrated prediction and Y_{orig} the original prediction. The calibrated prediction will be higher than the original prediction when the original prediction was below 11.2 cm/yr and lower otherwise. The variability of the prediction errors of the calibrated predictions was positively related to the value of the prediction ($p < 0.001$), described by the equation $SD_{pred\ err} = -1.017 + 0.286 * Y_{cal}$.

Our modification of the original prediction rule will give better predictions for children with IGHD fulfilling the same criteria.

In **Chapter 3**, we validated the modified prediction model from Chapter 2 in an independent cohort of 226 patients. For the modified as well as the original prediction method, observed first-year height velocity (HV) was plotted vs. predicted HV in a calibration plot. The regression line in the calibration plot of the modified model was not significantly different from the line of identity ($p = 0.43$), in contrast to the original model ($p < 0.001$). For the modified model, the mean (SD) prediction error was -0.11 (2.05) cm/yr and for the original model 0.28 (2.11) cm/yr.

In conclusion, the modified prediction method, obtained after calibration of the original model, performed well in an independent patient sample and gave more accurate predictions than the original model.

Since a prediction model for long-term response to GH treatment in children with GHD had not yet been presented, in **Chapter 4** we describe models we developed for prediction of adult height (AH), using information available at start of GH treatment (Start model) or after one year of treatment (First-year model). Data were collected from the National Registry of Growth Hormone Treatment in Children, which contained data of Dutch children treated with GH. Main inclusion criterion was a diagnosis of GHD or a maximum GH response during provocation tests < 11 ng/ml. The children had to be treated with biosynthetic GH for at least one year. To be able to use the complete selection of 342 children for the development of the models, multiple imputation was used for missing values. The predictions models, obtained by stepwise forward selection of determinants, were corrected for overfitting by bootstrapping. We developed separate models for prepubertal and for pubertal children, with outcome AH SDS.

In each prediction model TH SDS as well as current height SDS were significant determinants. In the First-year models, the change in height SDS during the first year proved to be an important predictor. The percentage explained variance, after correction for overfitting, ranged from 37% (prepubertal children, Start model) to 60% (pubertal children, First-year model).

In none of the models, GH dose was selected as significant predictor. For the patients in our data set, the GH dose at start and during treatment was assessed by the clinician based on unknown criteria. It is possible that higher initial doses were prescribed to patients with supposed worse prospects and that during treatment a change of GH dose might have been guided by the obtained growth response. Our data were therefore not suitable to estimate the effect of GH dose on AH SDS.

The presented prediction models give accurate predictions of AH for children with GHD at start and after one year of GH treatment. They are useful tools in the treatment of these children.

In **Chapter 5**, a model is presented for prediction of height at onset of puberty and of AH in short children born SGA and treated with GH. While GH treatment is approved for these children, the optimal dose is not yet established.

Data to develop this model came from two GH trials in short SGA children. We selected 150 children with height at start of treatment below the -2 SDS, age 3 years

or older, no signs of catch-up growth, available height at onset of puberty and at least one year of GH treatment prior to onset of puberty. In one of the trials, patients were randomly assigned to either 0.033 or 0.067 mg GH / kg·day; in the other trial all received 0.033 mg GH / kg·day. In 71 children, AH was reached.

Important finding was the significant contribution of serum IGFBP-3 SDS at start and the significant interaction between GH dose and IGFBP-3 SDS, indicating a smaller GH dose effect for higher levels of IGFBP-3. The final model explained 57% of the variance in height SDS at onset of puberty and 41% of AH SDS (percentages corrected for overfitting).

These findings implicated that measurement of IGFBP-3 in SGA children is needed to establish the effect of GH dose and to obtain a prediction for AH. The presented model provides useful information about the expected long-term growth. Because GH dose is one of the determinants, the model aids to determine the optimal GH dose for each child.

Chapter 6 describes a study on catch-up growth in SGA children. While the majority of children born SGA has catch-up growth during the first years of life, about 10-15% still has a height below the normal range (< -2 SDS) at the age of 2 years. The aim of this study was to determine the percentage of children with catch-up growth to a normal height after the age of 2 years and to develop a prediction model for growth after that age.

In a cohort of 724 SGA children, 109 had a height below -2 SDS at the age of 2 years. Ninety-seven could be included in this study. Thirty-nine percent of them reached a height above -2 SDS between 2 and 8 years (6% of the total group). Most important determinant of growth after the age of 2 years was the difference between TH SDS and height SDS at 2 years of age. The residual SD was 0.54 and the percentage variance explained by the model was 44%. With a predicted value and SD, probabilities can be computed, for example, the probability to reach a height SDS at 8 years of age above -2.5, or above -2.0.

The presented prediction model can identify children with low or high probability of catch-up growth. This may be useful to determine which children require medical follow-up.

In long-term growth studies with adult height (AH) as outcome, often reporting is required while data are incomplete. Some participants have not yet reached AH whereas others might be lost to follow-up. Current practice is to analyse only

participants who did reach AH, which can easily give biased results. In **Chapter 7**, we describe a new method, developed for analyses of cost-effectiveness.

The method requires growth data at fixed time intervals during follow-up. The mean AH is estimated using the estimated probability of still growing at the start of each interval and the mean growth in each interval.

We illustrated the validity of the method for growth data by applying it to artificially generated growth data in a simulation process. In that situation we know the true mean and standard deviation (SD) of AH in the population. We used five different imaginary growth scenarios, with different relations between AH and the age at which AH is reached. With this simulation study, we showed that the proposed new method had only a very small bias (less than 0.2 cm) in each scenario. The standard deviations were smaller than those of the other methods used for censored data, showing that the method is statistically more efficient.

We applied the method to data of patients from a registration database and data of a GH trial. The estimated means had smaller bias and standard error than estimations with commonly used methods. The method is not hampered by a correlation between AH and the age at reaching AH. So in contrast to commonly used methods, the new method provides valid results on mean AH when complete actual measurements of AH are not (yet) available. As it also uses observed data of children with incomplete follow-up, the method employs the data more effectively.

Chapter 8 describes a model concerning prenatal growth. In clinical practice, size and weight of fetuses are evaluated using standard reference tables for fetal biometry measurements. However, this neglects normal variation in fetal growth due to characteristics like gender, parity, ethnicity, maternal weight, height and age, and paternal height, which are important determinants of fetal growth. Customisation of fetal growth charts attempts to adjust for physiological characteristics and so to estimate the growth potential for an individual fetus. We developed a model for the construction of fetus specific growth charts.

Data for this study came from the Generation R Study, a prospective, population-based cohort study, following children born in Rotterdam, the Netherlands, from early fetal life until young adulthood. We could use data of 8,162 singleton pregnancies of which 5,473 had complete data on all determinants in our final model.

Determinants in the model, with outcome fetal weight, estimated using ultrasound measurements, were: fetal gender, parity, ethnicity, maternal age, height, weight.

The model was adjusted for the maternal smoking during pregnancy, because it has a substantial adverse effect on fetal growth.

This study was the first study fitting a model to obtain individually customised growth charts using ultrasound measurements in a large population-based study. The use of these charts may improve monitoring of fetal growth and prenatal care.

Finally, in **Chapter 9,** the main findings from this thesis are discussed and recommendations are given for the development of future prediction models for growth response and their use in clinical practice.

Samenvatting
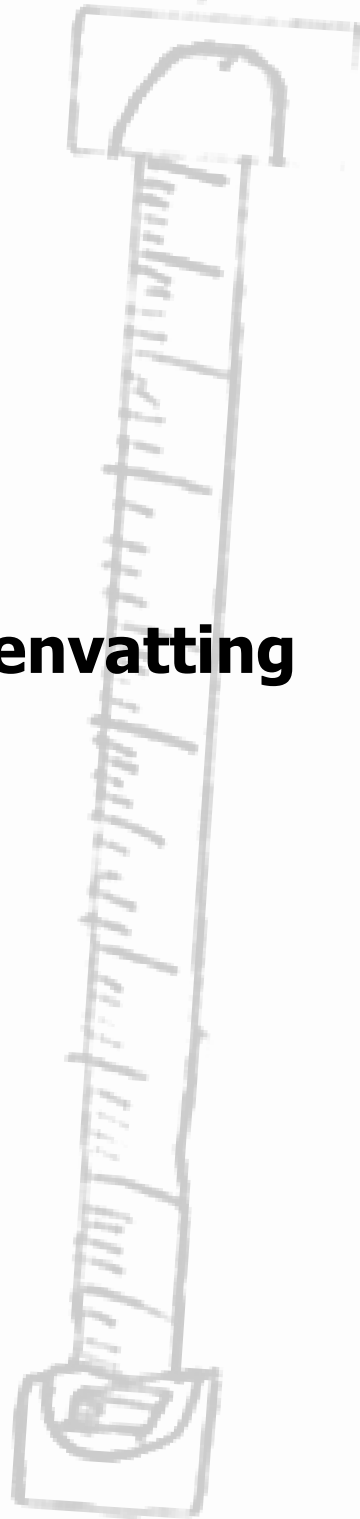
In **Hoofdstuk 1** worden de achtergronden gegeven van de onderwerpen die in dit proefschrift ter sprake komen. Enkele aspecten van groei, zowel prenataal als postnataal, en van groeiregulatie worden beschreven. De studies in dit proefschrift betreffen twee groepen van kinderen met groeistoornissen, nl. kinderen met groeihormoon deficiëntie (GHD) en kinderen die bij geboorte te klein zijn, gegeven de zwangerschapsduur (short for gestational age, SGA). Over deze groepen wordt enige informatie gegeven, o.a. over de behandeling met groeihormoon (GH). Ook wordt de huidige stand van zaken m.b.t. predictiemodellen voor groeirespons op GH behandeling beschreven.

    Hoofdstuk 2 beschrijft de validatie en calibratie van een model gepubliceerd in 1999, voor het voorspellen van groei van kinderen met idiopathische groeihormoon deficiëntie (IGHD), tijdens GH behandeling. Dit model, dat de groei tijdens het eerste jaar GH behandeling voorspelt, hebben wij gevalideerd met data van 136 Nederlandse kinderen met IGHD. Voor deze validatiegroep werd de geobserveerde groeisnelheid uitgezet tegen de voorspelde groeisnelheid in een zgn. calibratieplot. Er werd een regressie-analyse uitgevoerd met als uitkomst de geobserveerde groeisnelheid en als determinant de voorspelde groeisnelheid. De regressielijn was statistisch significant van de diagonaal y = x (waar de geobserveerde groeisnelheid gelijk is aan de voorspelde groeisnelheid) en de geschatte helling van de lijn was 0.808. Dit betekent dat in het model overfitting aanwezig was, wat een bekend verschijnsel is bij predictiemodellen. Met de geschatte coëfficiënten van de regressielijn werd een calibratiecorrectie voor het model opgesteld. De formule voor deze correctie was: $Y_{cal} = Y_{orig} + (2.153 - 0.192 * Y_{orig})$, met Y_{cal} de gecalibreerde predictie en Y_{orig} de originele predictie. De gecalibreerde predictie is hoger dan de originele predictie indien de originele predictie kleiner is dan 11.2 cm/jr en in de andere gevallen lager. Ook bleek dat de variantie van de predictiefouten positief gerelateerd was aan de waarde van de predictie (p < 0.001), volgens de vergelijking $SD_{pred\ fout} = -1.017 + 0.286 * Y_{cal}$.

    De gevonden modificatie van de originele predictieregel geeft betere predicties voor kinderen met IGHD die voldoen aan dezelfde criteria.

    Hoofdstuk 3 beschrijft de validatie van het aangepaste predictiemodel uit Hoofdstuk 2 in een nieuw onafhankelijk cohort van 226 patiënten. Zowel voor de aangepaste als voor de oorspronkelijke predictieregel werd de geobserveerde groeisnelheid in het eerste jaar uitgezet tegen de voorspelde groeisnelheid in een calibratieplot. De regressielijn in de calibratieplot van het aangepaste model was niet

significant verschillend van de lijn y = x (p = 0.43), in tegenstelling tot het oorspronkelijke model (p < 0.001). De gemiddelde (SD) predictiefout van het aangepaste model was -0.11 (2.05) cm/jr en voor het oorspronkelijke model 0.28 (2.11) cm/jr.

Geconcludeerd werd dat de aangepaste predictiemethode, verkregen door calibratie van het oorspronkelijke model, goed voldeed in een onafhankelijke groep patiënten en accuratere predicties gaf dan het oorspronkelijke model.

Voor langetermijnrespons op GH behandeling van kinderen met GHD was nog geen predictiemodel voor handen. Wij ontwikkelden modellen, die worden beschreven in **Hoofdstuk 4**, voor de predictie van volwassen lengte**.** Het betreft zowel modellen die voorspellen op basis van gegevens die bekend zijn bij de start van de GH behandeling (Start model) als modellen die de gegevens na één jaar behandeling gebruiken (First-year model). De data voor het ontwikkelen van de modellen werden geselecteerd uit de Landelijke Registratie Groeihormoonbehandeling bij Kinderen, waarin alle kinderen in Nederland die behandeld worden met GH geregistreerd zijn. Belangrijk inclusiecriterium was de diagnose GHD of een maximum GH respons tijdens provocatietesten van minder dan 11 ng/ml. De kinderen waren tenminste één jaar behandeld met biosynthetisch GH. Om de complete geselecteerde groep van 342 kinderen te kunnen gebruiken voor de ontwikkeling van de modellen, gebruikten we multipele imputatie voor het imputeren van missende gegevens. De predictiemodellen, die verkregen werden d.m.v. stepwise forward selectie van de determinanten, werden gecorrigeerd voor overfitting met bootstrapping.

Alle predictiemodellen bevatten zowel de doellengte (target height,TH) SDS als de huidige lengte SDS. In de First-year modellen was de verandering in lengte SDS tijdens het eerste jaar een belangrijke predictor. De percentages verklaarde variantie, gecorrigeerd voor overfitting, lagen tussen de 37% (prepubertaire kinderen, Start model) en 60% (pubertaire kinderen, First-year model).

In geen van de modellen werd de dosis GH geselecteerd als significante predictor. Voor de patiënten in onze dataset werd de GHdosis bij start en gedurende de behandeling vastgesteld door de behandelend arts, gebaseerd op niet-geregistreerde criteria. Mogelijk werd er vaak een hogere startdosis voorgeschreven aan patiënten voor wie de verwachtingen laag waren, en werd de GHdosis gedurende de behandeling gewijzigd afhankelijk van de geobserveerde groeirespons. Onze data waren daarom niet geschikt om een schatting te geven van het effect van GH dosering op volwassen lengte.

De gepresenteerde predictiemodellen geven accurate predicties van volwassen lengte voor kinderen met GHD, bij start of na een jaar GH behandeling en leveren daarmee nuttige informatie voor de behandeling van deze kinderen.

In **Hoofdstuk 5** wordt een model gepresenteerd voor predictie van de lengte bij de start van de puberteit en van de volwassen lengte van te klein geboren en te klein gebleven kinderen (SGA) die behandeld worden met GH. SGA is 2005 een indicatie voor GH behandeling maar de optimale GH dosis staat nog niet vast.

Voor het ontwikkelen van dit model werd data gebruikt van twee GH studies. We selecteerden 150 kinderen met een lengte bij start van de behandeling onder de -2 SDS en een leeftijd van 3 jaar of ouder, die geen tekenen van spontane inhaalgroei vertoonden. De lengte bij de start van de puberteit moest bekend zijn en de behandeling met GH diende tenminste één jaar vóór de start van de puberteit begonnen te zijn. In een van de studies waren de patiënten gerandomiseerd voor 0.033 of 0.067 mg GH / kg·dag, in de andere studie werden alle kinderen behandeld met 0.033 mg GH / kg·dag. 71 kinderen hadden hun volwassen lengte bereikt.

Een belangrijke bevinding was dat serum IGFBP-3 SDS bij start een significante determinant was en dat er een significante interactie was tussen GH dosis en IGFBP-3 SDS. Dit betekent dat bij hogere waarden van IGFBP-3 het effect van de GH dosis kleiner was. Met het model werd 57% verklaard van de variantie in lengte SDS bij start van de puberteit en 41% van de variantie in volwassen lengte SDS (percentages gecorrigeerd voor overfitting).

Deze resultaten betekenen dat de bepaling van IGFBP-3 in SGA kinderen noodzakelijk is om het effect van GH dosis vast te kunnen stellen en om een predictie voor volwassen lengte te kunnen geven. Het model verschaft belangrijke informatie over de te verwachten langetermijngroei. Omdat de GH dosis een van de determinanten is, kan het model gebruikt worden bij het bepalen van de optimale GH dosis voor een kind.

Hoofdstuk 6 beschrijft een studie naar de inhaalgroei van SGA kinderen. De meerderheid van te klein geboren kinderen vertoont inhaalgroei gedurende de eerste levensjaren, maar ongeveer 10-15% heeft op de leeftijd van 2 jaar nog een lengte onder het normale gebied (< -2 SDS). Het doel van deze studie was om het percentage kinderen die na de leeftijd van 2 jaar alsnog een voor de leeftijd normale lengte bereiken vast te stellen en om een predictiemodel te ontwikkelen voor groei vanaf 2 jaar.

In een cohort van 724 SGA kinderen hadden er 109 een lengte onder de -2 SDS op de leeftijd van 2 jaar. Van hen konden er 97 worden geïncludeerd in deze studie. Negenendertig procent van hen bereikte een lengte boven de -2 SDS op een leeftijd tussen 2 en 8 jaar (6% van de totale groep). De belangrijkste determinant van groei na de leeftijd van 2 jaar was het verschil tussen doellengte SDS en lengte SDS op de leeftijd van 2 jaar. De residuele SD was 0.54 en het percentage verklaarde variantie was 44%. Met een voorspelling van de lengte SDS op 8 jarige leeftijd en de SD kan de kans worden bepaald dat de lengte op 8 jarige leeftijd bijvoorbeeld boven de -2.5 SDS zal zijn, of boven de -2.0 SDS.

Het beschreven predictiemodel kan identificeren welke kinderen een kleine of grote kans hebben op spontane inhaalgroei. Dit kan helpen om te bepalen welke kinderen in de kliniek gevolgd dienen te worden.

In langetermijnstudies naar groei, met volwassen lengte als uitkomst, is het vaak nodig om al resultaten te rapporteren op een moment dat nog niet alle data compleet zijn. Enkele respondenten hebben hun volwassen lengte nog niet bereikt en/of anderen kunnen niet meer gevolgd worden. De huidige aanpak is om alleen de respondenten mét volwassen lengte te analyseren. Dit kan echter gemakkelijk verkeerde resultaten geven. In **Hoofdstuk 7** wordt een nieuwe methode beschreven, oorspronkelijk ontwikkeld voor kosteneffectiviteitsanalyses.

Voor deze methode zijn lengtemetingen nodig op vaste tijdstippen. De gemiddelde volwassen lengte wordt geschat m.b.v. de geschatte kans om nog aan het groeien te zijn aan het begin van ieder tijdsinterval en de gemiddelde groei in ieder interval.

We lieten zien dat deze methode valide is voor groeidata door de methode toe te passen op kunstmatig gegenereerde groeigegevens in een simulatiestudie. In een dergelijke situatie zijn het werkelijke gemiddelde en de standaard deviatie (SD) van de volwassen lengte in de populatie bekend. Er werden vijf verschillende groeiscenario's gebruikt, met verschillende relaties tussen volwassen lengte en de leeftijd waarop deze bereikt wordt. Met deze simulatie studie werd aangetoond dat de voorgestelde nieuwe methode in ieder scenario slechts een zeer klein verschil gaf met het bekende populatiegemiddelde (minder dan 0.2 cm). De standaard deviaties waren kleiner dan die van de andere methoden die gebruikt worden voor gecensureerde data. Dit toont aan dat de methode statistisch gezien efficiënter is.

De methode werd toegepast op data van patiënten uit een registratie database en op data van een GH studie. De geschatte gemiddelden van de volwassen lengte waren minder onzuiver en hadden een kleinere standaard fout dan verkregen werd met de gebruikelijke methoden. De validiteit van de schattingen werd niet beïnvloed

door een correlatie tussen de volwassen lengte en de leeftijd waarop deze bereikt wordt.

Samenvattend geeft de nieuwe methode valide resultaten voor de gemiddelde volwassen lengte in situaties waarin (nog) niet alle volwassen lengtes geobserveerd zijn. Omdat de methode ook de geobserveerde gegevens meeneemt van kinderen van wie de follow-up niet compleet is, worden de data effectiever gebruikt.

Hoofdstuk 8 beschrijft een model voor prenatale groei. In de klinische praktijk worden grootte en gewicht van een foetus beoordeeld aan de hand van standaard referentie tabellen voor foetale groeiparameters. Hierbij wordt geen rekening gehouden met de normale variatie in foetale groei die veroorzaakt wordt door eigenschappen als geslacht, pariteit, etniciteit, gewicht, lengte en leeftijd van de moeder en lengte van de vader, wat belangrijke determinanten zijn van foetale groei. Foetusspecifieke groeicurven daarentegen houden wel rekening met de eigenschappen van een foetus. Hiermee kan de groeipotentie voor een individuele foetus beter bepaald worden. Wij ontwikkelden een model voor de constructie van foetusspecifieke groeicurven.

Data voor deze studie kwamen van de Generation R Studie, een prospectieve studie onder kinderen geboren in Rotterdam, waarin gegevens verzameld worden vanaf de vroege zwangerschap tot jongvolwassenheid. Er konden gegevens gebruikt worden van 8,162 eenlingen, van wie er 5,473 complete gegevens hadden voor alle determinanten die in het model gebruikt werden.

De determinanten in het model, met als uitkomst het foetale gewicht, geschat o.b.v. echometingen, waren: foetaal geslacht, pariteit, etniciteit, leeftijd, lengte en gewicht van de moeder. Het model werd gecorrigeerd voor roken van de moeder gedurende de zwangerschap, omdat dit een substantieel negatief effect heeft op de foetale groei.

Deze studie was de eerste waarin een model voor de constructie van foetusspecifieke groeicurven ontwikkeld werd gebruikmakend van echometingen uit een groot bevolkingsonderzoek. Het gebruik van deze curven kan leiden tot verbetering van de prenatale zorg.

Tenslotte worden in **Hoofdstuk 9** de belangrijkste bevindingen uit dit proefschrift besproken en worden aanbevelingen gedaan voor het gebruik van predictiemodellen in de klinische praktijk en voor het ontwikkelen van toekomstige predictiemodellen voor groeirespons.

Dankwoord
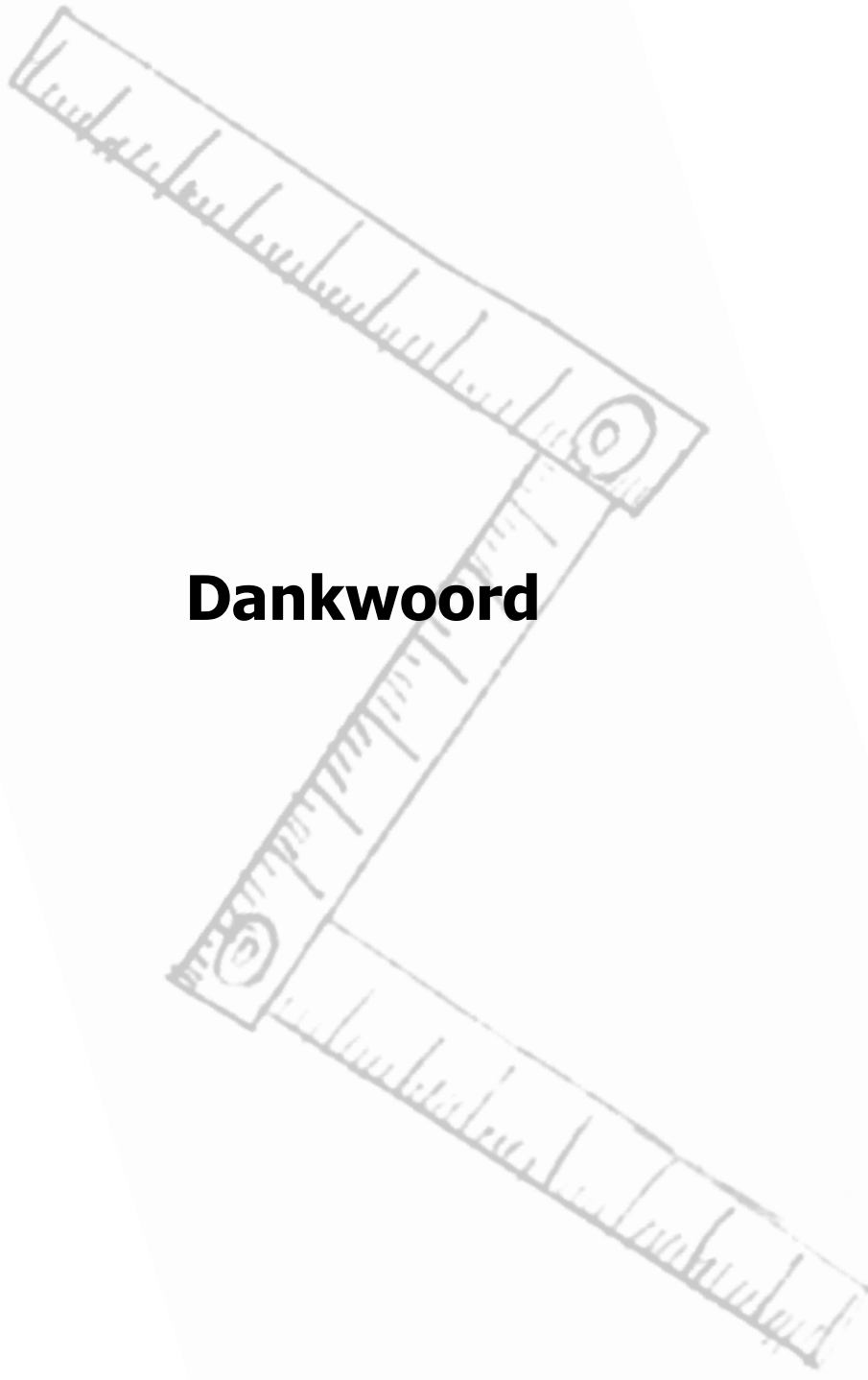
Bij het afronden van dit boekje wil ik graag een aantal mensen bedanken.

In de eerste plaats zijn dat mijn promotoren Prof.dr. A.C.S. Hokken-Koelega en Prof.dr. Th. Stijnen. Beste Anita, dank dat ik indertijd een plek bij de Stichting Kind en Groei heb gekregen. Analyses over groei vond en vind ik heel interessant en ik ben blij dat ik in dit gebied kan werken. Ook vind ik het bijzonder dat jij, als hoogleraar in de kinder-endocrinologie, een "buitenstaander" als ik, een biostatisticus, wilde begeleiden. Dank voor al je geduld en je vertrouwen. Beste Theo, bedankt voor je uitstekende begeleiding, je praktische oplossingen voor wat soms onoverkomelijke problemen leken, en natuurlijk voor het vele dat ik van je heb geleerd de afgelopen jaren. Dat betreft de inhoud van de statistiek, maar ook meer algemeen hoe je als biostatisticus grote en kleine statistische vragen en problemen aan kunt pakken. Je vertrek naar Leiden neemt mij die basis natuurlijk niet meer af.

Prof.dr. E. Steyerberg en Prof.dr. C.M. Moons wil ik bedanken voor het plaatsnemen in de kleine commissie en het beoordelen van het manuscript. Prof.dr. M. Ranke, thank you for your willingness to read the manuscript and to participate in the thesis defence ceremony.

Ook wil ik alle mensen bedanken die (soms al ver vóór mijn tijd) de data hebben verzameld waarmee ik de analyses heb kunnen doen. Dit zijn o.a. medewerkers van de Stichting Kind en Groei, van de Generation R studie en van de afdeling kinder-endocrinologie van het Sophia kinderziekenhuis. Zoals arts-onderzoekers beseffen dat hun onderzoek niet mogelijk was geweest zonder de medewerking van patiënten, zo was ik nergens geweest zonder dat alle gegevens van die patiënten door anderen gemeten en geregistreerd waren: lengte-metingen, lab-waarden, diagnoses, echo-metingen enz.

Ook de leden van de Adviesgroep Groeihormoon hebben een belangrijk aandeel in het verzamelen van de gegevens gehad en ben ik daarvoor zeer erkentelijk. Daar-naast heb ik het als heel nuttig ervaren dat zij respons gaven op de verschillende analyses. Op- en aanmerkingen van mensen uit de praktijk zijn onmisbaar voor iemand die alleen met de getallen aan het werk is.

Alle co-auteurs wil ik bedanken voor hun bijdragen aan de studies en artikelen. Bero, bedankt voor de prettige samenwerking.

Mijn kamergenoten bij Biostatistiek, Lidia en Bettina: wat is het boffen om collega's te hebben die maken dat je graag naar je werk gaat. Dat komt niet alleen door onze gesprekken over belangrijke en minder belangrijke dingen, maar ook doordat we in het werk met elkaar mee willen denken en zonodig bijspringen. Bettina, ik ben heel blij jou als paranimf te hebben, dat geeft vertrouwen. Lidia, dank voor alle hulp, steun en tips.

Ook mijn andere collega's van Biostatistiek (Wim, Paul, Sten, Taye, Marek en Emmanuel), ex-collega's en ex-kamergenoten wil ik bedanken voor hun collegialiteit, vriendelijkheid en hulpvaardigheid. Daarnaast dank aan mijn collega's van Epidemiologie, die zeker ook bijdragen aan een gezellige en interessante werk-omgeving. Aan Nano speciaal dank voor de onmisbare hulp bij computer-problemen.

Alle collega's van de Stichting Kind en Groei wil ik graag bedanken voor de prettige samenwerking. Ik vind het bijzonder dat ik, ondanks mijn soms erg onregelmatige komst naar het mooie pand, me er eigenlijk altijd heel snel weer thuis voel. Gladys, ik ben heel blij met jou als paranimf en ik hoop nog veel met je samen te werken.

Familie, vrienden en buren, bedankt voor al jullie belangstelling en meelevendheid, maar zeker ook voor de broodnodige activiteiten naast het werk. Vooral de laatste maanden was het goed dat jullie (misschien onbewust) mij af en toe achter de computer vandaan haalden. Clare, many thanks for your help with the English during the last phase.

Lieve pap en mam, dank voor jullie steun, liefde en belangstelling, dat heeft echt energie gegeven.

En dan het thuisfront: jullie hebben het maar allemaal moeten doorstaan de afgelopen jaren. Dank voor jullie geduld! Lieve Jori en Matti, bedankt voor al het plezier dat ik met jullie heb, voor jullie vrolijke en jullie verstandige opmerkingen, en natuurlijk voor de prachtige tekeningen voor in en op het boekje. Lieve Pieter, wat een geluk dat ik met jou samen ben.

Curriculum Vitae
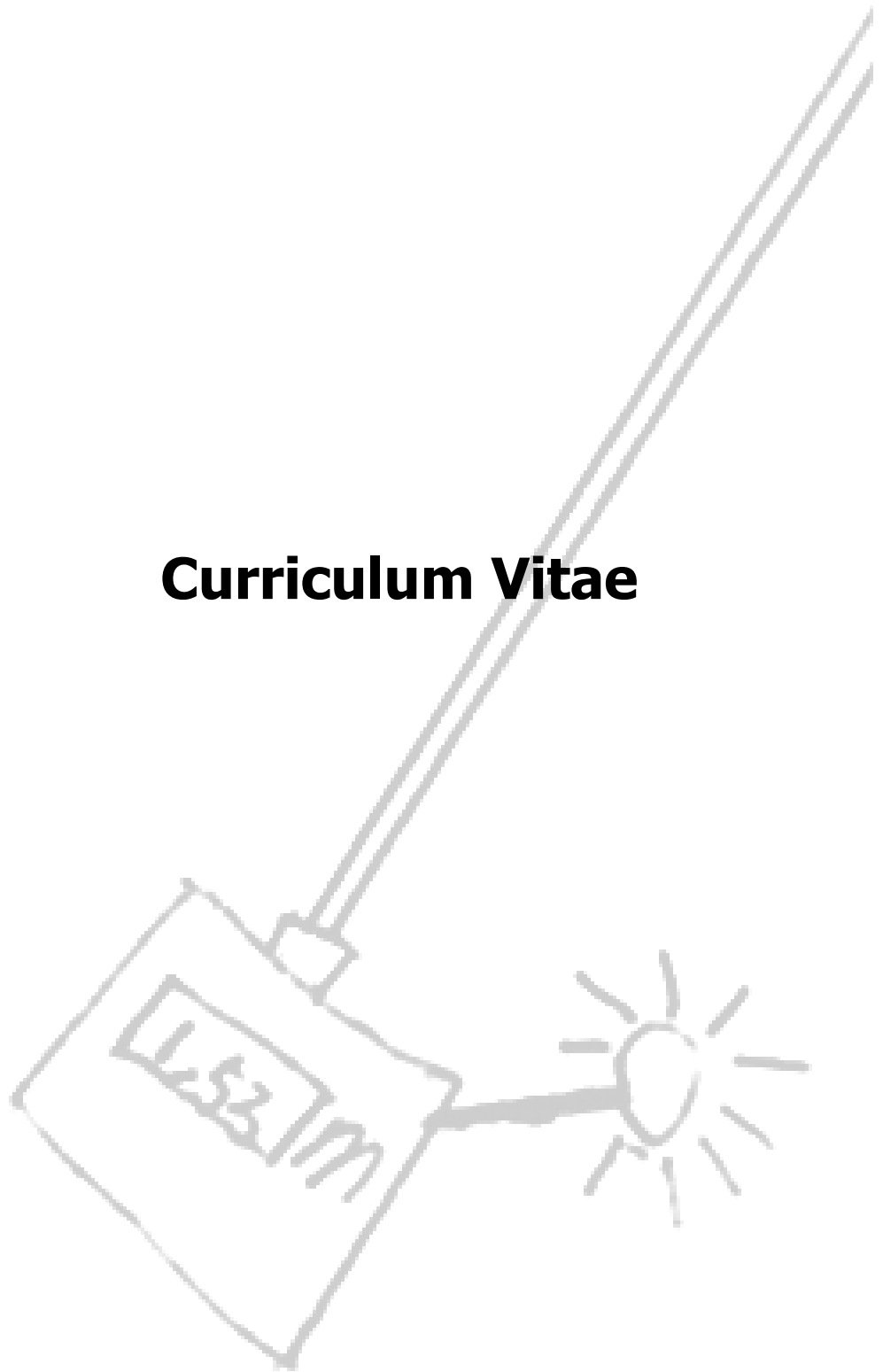
Maria de Ridder was born on February 8th, 1961 in Gouda. After finishing high school at the Coornhert Gymnasium in Gouda in 1979, she studied architecture for some months, but then decided to change her course of study. She subsequently went to a College of Education in Delft and in 1984 she graduated in mathematics and physics. Following her graduation she worked as a secondary school mathematics teacher for a number of years.

In September 1986 she started at the Vrije University in the faculty of Psychology and Pedagogical Science, as a computer programmer and assistant in the research group dealing with theoretical pedagogy. From September 1988 until March 1991, she worked at the Gemeenschappelijk Administratie Kantoor (GAK) as a statistical analyst. In March 1991, she started to work at the medical faculty of the Erasmus University Rotterdam (now called Erasmus MC), in the Department of Epidemiology and Biostatistics. In 1992 she graduated as a Statistician VVS and in 1995 she obtained a Master of Science in Clinical Epidemiology at the Netherlands Institute for Health Sciences in Rotterdam. Since 2000 she is registered as a Biostatistician VVS.

In December 2000 she started studying and researching the background information relating to the subject matter of this thesis namely the development of prediction models for growth response during growth hormone treatment. Her work was carried out under the supervision of Prof. Dr. A.C.S. Hokken-Koelega (Dutch Growth Research Foundation) and Prof. Dr. Th. Stijnen (Department of Epidemiology and Biostatistics).