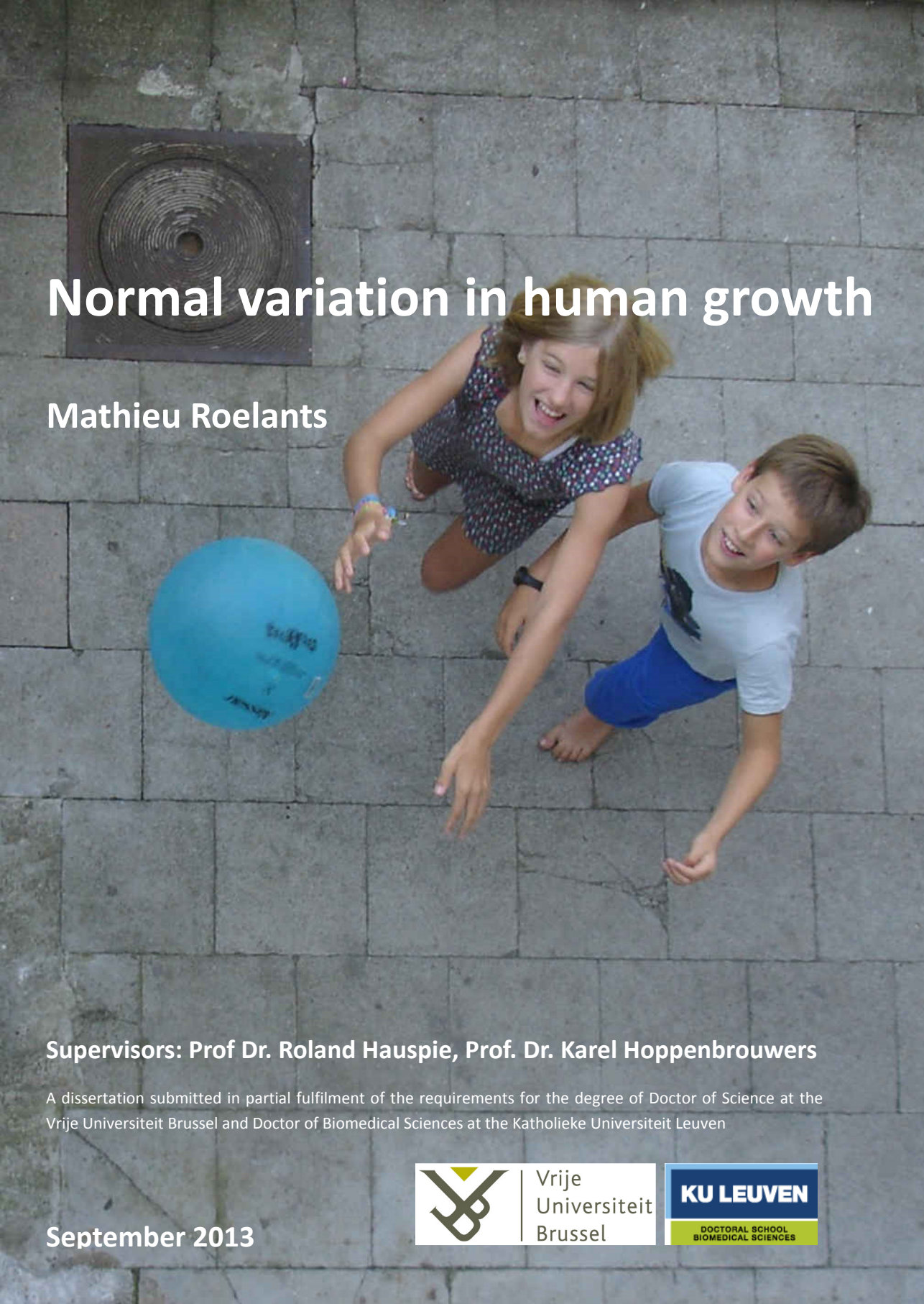


Normal variation in human growth

Mathieu Roelants

Supervisors: Prof Dr. Roland Hauspie, Prof. Dr. Karel Hoppenbrouwers

A dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Science at the Vrije Universiteit Brussel and Doctor of Biomedical Sciences at the Katholieke Universiteit Leuven

September 2013

Vrije Universiteit Brussel

KU LEUVEN
DOCTORAL SCHOOL BIOMEDICAL SCIENCES

VRIJE UNIVERSITEIT BRUSSEL
FACULTY OF SCIENCE AND BIO-ENGINEERING SCIENCES
DEPARTMENT OF BIOLOGY, LABORATORY OF ANTHROPOGENETICS
PLEINLAAN 2, B-1050 BRUSSELS, BELGIUM



ISBN 978-90-72325-07-5



KU LEUVEN FACULTY OF MEDICINE
DEPARTMENT OF PUBLIC HEALTH AND PRIMARY CARE
CENTRE OF YOUTH HEALTH CARE
KAPUCIJNENVOER 35, B-3000 LEUVEN, BELGIUM
mathieu.roelants@med.kuleuven.be



Normal Variation in Human Growth

Normal Variation in Human Growth

Mathieu Roelants, Roland Hauspie (sup), Karel Hoppenbrouwers (sup)

A dissertation submitted in partial fulfilment of the requirements for the joint degree of Doctor of Science at the Vrije Universiteit Brussel and Doctor of Biomedical Sciences at the Katholieke Universiteit Leuven

Jury: Prof. Dr. R. Hauspie (VUB, Promotor), Prof. Dr. K. Hoppenbrouwers (KU Leuven, Promotor), Prof. Dr. T. Cole (UCL Institute of Child Health, London, UK), Dr. Sc. Math. L. Molinari (University Children's Hospital, Zurich, Switzerland), Prof. Dr. M. Thomis (KU Leuven), Prof. Dr. F. de Zegher (KU Leuven), Prof. Dr. D. Maes (VUB), Prof. Dr. L. Triest (VUB, Chair)

ISBN 978-90-72325-07-5

(c) The authors, 2013

Youth Health Care, Department of Public Health and Primary Care, KU Leuven – University of Leuven, Kapucijnenvoer 35, B-3000 Leuven, Belgium

Laboratory of Anthropogenetics, Department of Biology, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium



Vrije Universiteit Brussel



VRIJE UNIVERSITEIT BRUSSEL
FACULTY OF SCIENCE AND BIOENGINEERING SCIENCES
DEPARTMENT OF BIOLOGY

KU LEUVEN BIOMEDICAL SCIENCES GROUP
FACULTY OF MEDICINE
DEPARTMENT OF PUBLIC HEALTH AND PRIMARY CARE

Normal Variation in Human Growth

Mathieu Roelants

Promoters:
Prof Dr. R. Hauspie
Prof. Dr. K. Hoppenbrouwers

A dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Science at the Vrije Universiteit Brussel and Doctor of Biomedical Sciences at the Katholieke Universiteit Leuven

September 2013

Preface/Acknowledgments

The results presented in this work are mostly based on data from the Flemish Growth Study. This research project would not have been possible without the help and support of many people. First of all, more than 18000 children and their parents, some of whom even returned up to 20 times for a follow up measurement. Staff members of Well Baby Clinics and Day Care Centres under supervision of Kind en Gezin, the Centra voor Leerlingenbegeleiding, the Student Health Centre of the Vrije Universiteit Brussel, and Occupational Health Centre attached to the University of Leuven, were responsible for the collection of data. The coordination of this field work was in the capable hands of Sara Daman, at that time attached to the Laboratory of Anthropogenetics. I also want to thank Prof. Dr. Jean Deschepper of the UZ Brussels, whom I could always consult for medical matters beyond my expertise. The Flemish Growth Study was an initiative of an ad-hoc workgroup (you may find their names in the acknowledgments of section 4.1) brought together by Paul Dierickx (then Pharmacia&Upjohn), and was supported by the Flemish Minister of Public Health. I enjoyed the – often enthusiastic – collaboration with all these persons, and truly hope that they are satisfied with the return, given that this has always been a primary concern to me and my colleagues.

There are also many friends and colleagues who consulted me with their interesting but sometimes distracting (at least for my own PhD!) data and projects. I hope you understand that it is not possible to mention all of you, and that it would be inappropriate to mention only some of you. I can however assure you that I enjoyed working with you on these “side projects”.

About half-way through the Flemish growth study, I was introduced to a Norwegian pediatrician who was planning a survey not unlike ours. At that time I could not have guessed that this meeting with Dr. Pétur Júlíusson was the start of a friendship, and a fruitful, pleasant collaboration, that made the work presented in this thesis more complete.

When considering my own development from schoolboy to PhD candidate, I should express my gratitude first and foremost to my mother and my late father.

They supported me to study at the university, and allowed me and my sisters to take opportunities far beyond those they received themselves. My mother must have known about secular trends all along, because each time I had a school health assessment, she was eager to know if I had yet reached 180cm. I also want to thank my longtime friends, without whom I consider it unlikely to have even continued studies beyond secondary school.

As a young graduate in Human Nutrition, I started working at the Centre of Youth Health Care, Department of Public Health and Primary care, University of Leuven, at that time led by the late Prof. Dr. em. Jacqueline Knops, succeeded by Prof. Dr. Paul De Cock and currently by Prof. Dr. Karel Hoppenbrouwers. Jacqueline Knops recognized my preference for working with data and encouraged me to obtain a master degree in statistics. Growth monitoring has always been an important part of community child health, and – given my interests and background – already appealed to me from my days as a student. You can image that when the Flemish Growth Study was approved by the Flemish Government, I was, mildly put, very interested. This project was executed under supervision of Prof. Dr. Roland Hauspie at the Laboratory of Anthropogenetics of the University of Brussels (at that time led by Prof. Dr. em. Charles Susanne), in close collaboration with Prof. Dr. Karel Hoppenbrouwers of the Centre of Youth Health Care. On my very first meeting with Roland I learned that such multidisciplinarity is the essence of Auxology (the study of human growth), and eventually I ended up with both of them supervising this joint PhD. Needless to say I am indebted to all these people, and in particular to Roland and Karel for supervising my thesis, for their valuable advice, for their patience, and for giving me the opportunity to study human growth. I consider both of them true experts in their field, and while I may not have formally attended their lectures, I do hope that I have learned their lessons.

A thesis requires a Jury, and I also would like to express my gratitude towards all members of the Jury for the time and effort they have put into the evaluation of my work.

Finally, many thanks to Ilse, my longtime partner, for her company, patience, and continuing support, and to our children, Elin and Nathan, 160.4 and 146.1cm and growing.

Summary

The growth of children mirrors their health and the environment in which they thrive, and is therefore widely used as proxy for health of individual children and the population to which they belong. The assessment of growth may involve a single measurement of attained size (distance), the timely occurrence of an event (tempo), the rate of growth between two measurements (velocity), or patterns of growth over several years. In this study, we developed cross-sectional references of length, height, weight, Body Mass Index (BMI), head circumference, and pubertal development from a representative sample of Flemish children. Sampling and data collection were up to current standards, and growth distance curves were fitted with the LMS method. Results show an ongoing secular trend in height of 0.8 – 1.2 cm per decade within the Belgian population. Height of Belgian children is roughly comparable with that in other northern European countries, except for the Netherlands. The BMI curve, and consequently the prevalence of overweight, is also largely comparable with that in neighbouring countries. References for pubertal development were estimated from *status presens* data using generalized additive models. Mean age at menarche has not changed over the past 50 years. Menarche and the development of secondary sex characteristics are within ranges that are observed in northern Europe. Our references for BMI and Tanner stages of pubertal development are the first for the Belgian population.

The assessment of changes in size over time requires appropriate instruments based on longitudinal data. We validated conditional growth based on the correlation between measurements converted to standard deviation scores (SDS), which is equivalent to the change in position on a growth distance chart. Correlations between serial measurements of length, weight and head circumference were estimated from longitudinal data of infants and primary schoolchildren. The correlation structure of weight and height SDS was comparable with similar data from the UK and France. References for conditional growth in length and head circumference in infancy were not previously available.

The perspective given by our study is that of the “normal” healthy child, and is therefore a proxy for the specificity or false positive rate of growth assessment.

This rate has been shown to increase when children are frequently assessed. From a systematic analysis of referral patterns we devised and validated a method to estimate the total referral rates of growth monitoring programs from the joint distribution of measurements.

Growth references show how children grow in a particular population at a particular time. The World Health Organization (WHO) *universal standards* are based on the attractive but controversial premise that all children grow alike under optimal feeding and environmental conditions. However, our analysis of growth in Belgian and Norwegian children that complied with these criteria shows that children grow more alike the local reference population. In our sample, as well as in other western populations, we observe shortly after birth a decrease in length and weight relative to the standards. This could be the result of a – in our opinion – too strict application of inclusion criteria by the WHO. Our findings for head circumference were later confirmed in the United Kingdom and in the United States, and indicate persistent population differences of anthropometric characteristics.

The tools that were developed in this study give guidance for the assessment of growth in individual children, and for the development and evaluation of guidelines and programs for growth monitoring.

A dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Science at the Vrije Universiteit Brussel and Doctor of Biomedical Sciences at the Katholieke Universiteit Leuven

September 2013

Samenvatting

De groei van kinderen weerspiegelt hun gezondheid en de omgeving waarin zij opgroeien, en wordt daarom vaak gebruikt als een indicator voor de gezondheid van individuele kinderen en van de populatie waartoe zij behoren. Het beoordelen van de groei kan betrekking hebben op de lengte die op een bepaalde leeftijd wordt bereikt ("de afstand"), het tijdig bereiken van bepaalde ontwikkelingsdoelen ("tempo"), de toename tussen twee metingen ("snelheid"), of groeipatronen over meerdere jaren. In deze studie werden cross-sectionele referentiecurven geschat voor lengte, gestalte, gewicht, Body Mass Index (BMI), hoofdomtrek, en pubertaire ontwikkeling in een representatieve steekproef van Vlaamse kinderen en jongeren. De steekproeftrekking en dataverzameling gebeurden volgens de gangbare normen, en referentiecurven werden geschat met de LMS methode. De resultaten tonen een aanhoudende seculaire toename in gestalte van 0.8 – 1.2 cm per decennium. De gestalte van Belgische kinderen is ruwweg vergelijkbaar met die in andere Noord-Europese landen, met uitzondering van Nederland. De BMI curve, en dus ook de prevalentie van overgewicht, is eveneens vergelijkbaar met die in de ons omringende landen. Referentiegegevens voor de pubertaire ontwikkeling werden geschat aan de hand van *status presens* data. De gemiddelde leeftijd van de menarche bleef de voorbij 50 jaar ongewijzigd. De menarche en de ontwikkeling van secundaire geslachtskenmerken vallen binnen de grenzen die in Noord-Europa worden waargenomen. De referenties voor BMI en de Tanner stadia voor pubertaire ontwikkeling zijn de eerste voor ons land.

Voor het beoordelen van de groei tussen twee meetmomenten zijn specifieke instrumenten, gebaseerd op longitudinale gegevens, noodzakelijk. In dit onderzoek werden referenties voor conditionele groei ontwikkeld op basis van de correlatie tussen standaarddeviatiescores (SDS). Deze methode is equivalent aan de verandering in positie op een groeicurve. De correlaties tussen opeenvolgende metingen van lengte, gestalte, gewicht en hoofdomtrek werden geschat aan de hand van longitudinale meetgegevens bij zuigelingen en schoolkinderen. De correlatie-structuur van gewicht en gestalte is vergelijkbaar met overeenkomstige gegevens uit het Verenigd Koninkrijk en Frankrijk. Referenties voor conditionele groei in lengte en hoofdomtrek waren tot nu toe nog niet beschikbaar.

Deze studie heeft het “normale” gezonde kind als perspectief, en is daarom een maat voor de specificiteit of het percentage vals positieve verwijzingen van groeimonitoring. In het verleden werd aangetoond dat dit percentage gevoelig stijgt wanneer kinderen meerdere keren worden gemeten. Door systematische analyse van verwijspatronen hebben we een methode ontwikkeld om deze toename te voorspellen aan de hand van de gezamenlijke statistische distributie van de metingen.

Referentiecurven tonen hoe kinderen groeien in een bepaalde populatie, in een bepaalde tijd. De universele groeistandaarden van de Wereldgezondheids-organisatie (WGO) zijn gebaseerd op de aantrekkelijke maar ook controversiële veronderstelling dat de groei van alle kinderen vergelijkbaar is indien zij beantwoorden aan optimale voedings- en omgevingscriteria. De groei van Vlaamse en Noorse kinderen die aan deze criteria voldoen komt echter beter overeen met de respectievelijke lokale referentiecurve. In onze steekproef, alsook in andere westerse landen, wordt kort na de geboorte een daling in lengte en gewicht ten opzichte van de standaarden waargenomen. Dit kan het gevolg zijn van de toepassing van – naar onze mening – te strikte inclusiecriteria door de WGO. Onze resultaten voor de hoofdomtrek, die nadien werden bevestigd in het Verenigd Koninkrijk en in de Verenigde Staten, wijzen duidelijk op persisterende verschillen in antropometrische kenmerken van populaties.

De instrumenten die hier werden voorgesteld bieden een referentiekader voor het beoordelen van de groei van een individueel kind, en voor het ontwikkelen en evalueren van richtlijnen voor groeiopvolging.

Proefschrift voorgedragen tot het behalen van de graad van Doctor in de Wetenschappen aan de Vrije Universiteit Brussel en Doctor in de Biomedische Wetenschappen aan de Katholieke Universiteit Leuven.

September 2013

Table of Contents

Preface/Acknowledgments 3
Summary 5
Samenvatting 7
Table of contents 9
Abbreviations 10
Introduction 11
1. Normal growth and development 13
    1.1 The assessment of growth and development 13
    1.2 Growth charts 28
    1.3 Normal growth 38
    1.4 Criteria for overweight 43
    1.5 Growth standards 46
    1.6 Summary and conclusions 51
2. Objectives 53
3. Materials and methods: The Flemish growth study 55
4. Results 59
    4.1 References for growth and pubertal development from birth to 21 years in Flanders, Belgium 61
    4.2 Breastfeeding, growth and growth standards: Performance of the WHO growth standards for monitoring growth of Belgian children 77
    4.3 Growth of Belgian and Norwegian children compared to the WHO growth standards: prevalence below -2 and above +2SD and the effect of breastfeeding 87
    4.4 Conditional growth and correlations between longitudinal measurements of length, height, weight and head circumference in infancy and childhood. 95
    4.5 Referral rates for serial measurements of growth 113
5. Discussion 121
    5.1 A cross-sectional growth distance reference 121
    5.2 References or universal standards 126
    5.3 Monitoring changes in growth over time 128
    5.4 Referral rates of growth monitoring programs 129
    5.5 Some final remarks 130
    5.6 Future perspectives 130
6. Summary and conclusions 132
References 135
Appendix: Flemish Growth Charts 2004 145

Abbreviations

AIC	Akaike Information Criterion
B#	Tanner stage of breast development in girls (# indicates the stage)
BCPE	Box Cox Power Exponential distribution
BIC, **SBC**	Schwartz Bayesian Information Criterion
BMI	Body Mass Index or Quételet Index
CDC	US Centers for Disease Control and Prevention
EDF	Equivalent Degrees of Freedom
G#	Tanner stage of genital development in boys (# indicates the stage)
GAIC	Generalized Akaike Information Criterion
GAMLSS	Generalized Additive Models for Location, Scale and Shape
GWAS	Genome Wide Association Study
IOTF	International Obesity Task Force
LMS	Method to summarize a growth reference curve, **L** = Lambda (Box Cox power), **M** = Mu (median), **S** = Sigma (coefficient of variation)
MGRS	Multicentre Growth Reference Study
NCHS	US National Center for Health Statistics
PH#	Tanner stage of pubic hair development in boys and girls (# indicates the stage)
SD	Standard deviation
SD Score	Standard deviation score
SDS	Standard deviation score
SES	Socio-Economic Status
T#	Testicular volume (discrete volume measured with Prader orchodimeter)
TEM	Technical error of measurement
TS	Turner Syndrome
WHO	World Health Organization

Introduction

One of the most frequent assessments in paediatrics and child care is without doubt that of "growth". The growth of children mirrors their health and the environment in which they thrive, and is therefore widely used as proxy for health of individual children and the population to which they belong. Growth charts are an indispensable tool to assess growth. These charts were commonly tied to a particular population, but in 2006, the World Health Organization (WHO) published universal growth standards from birth to 5 years of age. According to the WHO, all children grow alike, given that they thrive under optimal environmental and feeding conditions (WHO, 2006a). However, comparison of the WHO standards against national samples from developed countries challenges the claim of universality, but as these samples also included formula fed children, they could not discriminate feeding effects from population differences.

In most European countries, growth of infants and young children is measured on a regular basis, and each time compared to a cross-sectional growth chart. The percentiles that delimit normal growth on these charts determine the number of children that is referred at each measurement occasion. Referral rates are however likely to increase when children are assessed more then once. For instance, van Buuren *et al.* (2004a) estimated that the number of children that fell below -2.5 SD at least once before the age of 10 years was about 10 times larger then expected. A systematic analysis of such referral patterns is however still missing. In addition, consecutive measurements are often tracked on a cross-sectional growth chart. This is misleading because centile lines on these charts do not represent individual growth curves, and are not appropriate to monitor changes in growth over time. Several criteria for the dynamics of growth have been proposed, including centile crossing (deflection), growth velocity, and conditional growth, but only few references have been published in the literature. In this research we will investigate criteria for monitoring changes in growth over time, and validate the use of universal growth standards in a single population of Flemish children.

1. Normal growth and development

Growth is a dynamic process determined by a complex interaction of genetic, nutritional and environmental factors. This process is usually characterized by changes in size over time of anthropometric traits. Many of these traits, for instance height, are the intricate sum of various body parts, which in turn display their own pattern of growth. Despite this complexity, most children grow in a predictable way, i.e. within a narrow "normal" range. The growth process is however sensitive to ill health and changes in the physical and psychosocial environment, and poor growth is often the first manifestation of an underlying disorder, inadequate nutrition, or unfavourable environmental conditions. Growth assessment and anthropometry are therefore widely used to monitor health and development in individual children (de Onis *et al.*, 2004) and as indicators for health and environmental conditions within and between populations (WHO, 1986).

In this introductory chapter we will pursue the assessment of growth in individual children, which implies a form of screening. The first section gives an overview of the principles that are at the basis of the assessment of growth and development using one or more measurements. Methods to estimate a growth reference chart are discussed in section 1.2. Finally, sections 1.3 - 1.5 cover the criteria that define *normal* growth, which can be based on a sample of healthy children (for instance, references for length, height, head circumference), on the risk for ill health (for instance criteria for overweight), or on optimal growth (growth standards).

1.1 The assessment of growth and development

The assessment of physical growth is based on the comparison of a measurement (for instance, height, weight) with that observed in a group of healthy peers. Depending on the information at hand, this comparison can usually be reduced to three simple questions: (1) is the child of appropriate size for his age?, (2) is the rate of growth (for instance the change in size since the last visit) appropriate?, and (3) is the pattern of growth over the past several years appropriate? These questions were formulated by Tanner *et al.* (1966a) almost 50 years ago, and, despite a

substantial increase in knowledge on the biology of growth and on methodological aspects of growth assessment itself, they are still pertinent today.

1.1.1 Growth distance

Attained size is a measure of progress or growth since conception. It can be considered as a distance since it indicates that part of the growth cycle that is completed (Tanner *et al.*, 1966a). Several measurements of attained size connected by a line give an estimate of the individual growth distance curve. To assess if attained size is appropriate for a given age and sex, it is usually compared to the corresponding distribution in a group of children with a comparable background, which is called the *reference population*. When the measurement is not overly large or small, the growth distance is considered normal. A *cross-sectional distance chart* shows the distribution of attained size along a covariate (usually age) in a reference population. This distribution is summarized using percentiles or SD-lines (figure 1.1), and the space between these lines is referred to as "channels" or "centile bands". Typical choices are percentiles 3, 10, 25, 50, 75, 90 and 97, or SD-lines -2, -1, 0, 1 and 2, which have an equal spacing for Gaussian distributed traits like height and head circumference (Cole, 1994a). The percentiles 3 and 97 (or SD-lines -2 and +2) usually define the limits of *normal growth* (see section 2.3), but the cut-offs that warrant immediate referral in a public health setting can be more extreme, like -2.67 SD (Cole, 1994a) or -3 SD (Grote *et al.*, 2008). An alternative to dichotomizing the growth distance as normal or abnormal, is to express the position of a child within the distribution as a percentile (i.e. its rank in an ordered group of 100), or standard deviation score (SD-score, SDS, z-score, i.e. the number of standard deviations above or below the mean) (Tanner, 1952; Waterlow *et al.*, 1977; Wright and Royston, 1997). The SDS is more convenient, but requires that the distribution is parametrically defined (Cole, 1993). For a measurement y_t from a Gaussian distributed variable at age t, the SDS is given by $(y_t - \mathrm{MEAN}_t)/\mathrm{SD}_t$.

The statistical distribution of growth distance was first characterized by Quételet in 1835 (Quételet, 1835), but it was Bowditch who created the first growth reference charts using "anthropometric percentiles" as proposed by Galton (Bowditch, 1891). More recent examples of cross-sectional distance references are those from the

Netherlands (Fredriks *et al.*, 2000), UK (Freeman *et al.*, 1995), and USA (Kuczmarski *et al.*, 2002), but in general they are available for the large majority of industrialized countries (de Onis *et al.*, 2004). Belgian reference charts were published by Joossens *et al.* (1977), and Wachholder and Hauspie (1986).

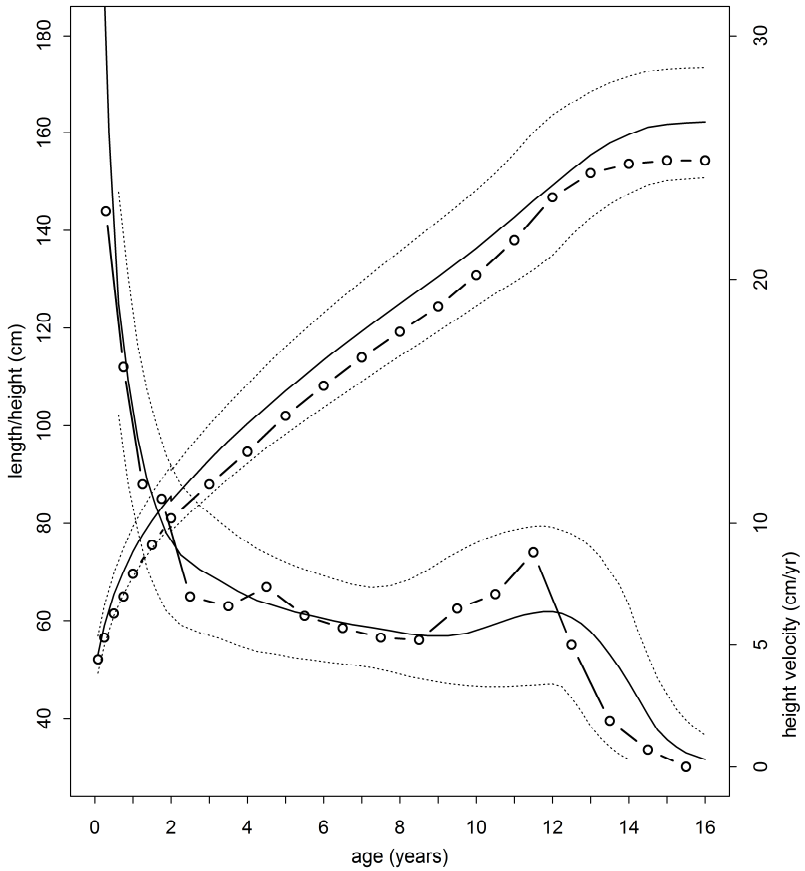


Figure 1.1: Cross-sectional distance and velocity chart for girls (median and percentiles 3 and 97 from Tanner et al., 1966b). The connected dots show raw data of an individual child (data from Hauspie and Wachholder; 1986).

Qualitatively, the individual and cross-sectional distance curves have the same overall shape, but quantitatively they are different. Serial measurements in growth are highly correlated, and therefore children tend to track along the chart (i.e.

maintain their rank), but centile crossing is also common, particular in early infancy (Smith *et al.*, 1976; Mei *et al.*, 2004) and in puberty (Tanner *et al.*, 1966b; Hermanussen et al., 2001). *Infancy* is a transitional period from prenatal growth dominated by maternal factors, to a postnatal growth trajectory determined by the genetic background and environmental factors (Smith *et al.*, 1976). Catch-up or catch-down growth is a common feature observed in the majority of newborns (Smith *et al.*, 1976; Park *et al.*, 1997; Mei *et al.*, 2004). This phenomenon of *canalization* is evident from a low correlation between size at birth and subsequent growth (Smith *et al.*, 1976; Boryslawski, 1988). In *puberty*, the growth rate accelerates, and there is a considerable increase in size over a relatively short period. The duration of the pubertal growth spurt is fairly constant, but the timing of this event is variable (with an SD of approximately 1 year), and the individual growth curves are out of phase (Tanner *et al.*, 1966a). As a consequence, the cross-sectional mean curve, which is based on a mixture of early, average and late maturing children, shows a less intense increase in growth rate of longer duration (i.e. the pubertal growth spurt is flattened out), and the variance increases (proof is given in Cole *et al*. 2008). More importantly, the growth distance curve of an individual child will deviate from the average growth curve during this period, and crossing of centiles is a common and perfectly normal phenomenon: early maturing children will cross centiles upwards at the start of puberty, and downwards at the end of their spurt, and late maturing children will cross centiles downwards first and upwards thereafter. In figure 1.1, the *phase-difference* effect is evident from the steeper slope of the individual distance curve compared to the average distance curve. Differences in tempo may well be present in *childhood*, but have a smaller effect on the shape of the curve, because the increase or decrease in the rate of growth is less variable (Molinari and Gasser, 2004). As a consequence, percentile crossing or decanalization is less frequently observed in this age group (Li *et al.*, 1998; Mei *et al.*, 2004).

1.1.2 Growth velocity

By differencing attained size at two consecutive ages, we get an estimate of the *increment* or *average growth velocity* over the considered interval. Growth velocity is a measure of change rather then status, and is more sensitive to environmental factors operating over a recent period of time, while distance is the result of a

complex interaction between genetic and environmental factors operating over several years (Shuttleworth, 1934; Tanner, 1952; Brook 1983). Increments are usually taken as the absolute increase in size over a specific interval, $(H_t - H_{t-1})$, and the corresponding velocity as rate of growth expressed per unit of time, $(H_t - H_{t-1})/\Delta t$. When properly scaled, increments and velocities are obviously interchangeable, except for their units (e.g. increments over exactly one month in centimetres and monthly growth velocity in cm/month).

Growth velocities are tied to a particular age interval, because the velocity changes with age, and the variance with age and the length of the interval. The dependence on the length of the interval is due to (1) fluctuations in the growth rate, which can be highly irregular over periods of days or weeks (Lampl *et al.*, 1992; Hermanussen, 1998), or even months (Butler et al., 1990; Thalange *et al.*, 1996), (2) seasonal variation (Marshall, 1971; Hauspie and Pagezy, 1989), and (3) measurement error (Bairagi, 1986). The relative contribution of these components becomes smaller with the size of the increment, and with the length of the interval. Empirical studies have repeatedly confirmed that velocities taken over a shorter period have a larger variance and vice versa (Tanner *et al.*, 1966b; Roche and Himes, 1980; Healy *et al.*, 1988b; Neyzi *et al.*, 1993). As a matter of convenience, growth velocities are often taken over a period of one year (0.85 - 1.15 years), which cancels out seasonal variation and ensures increments that largely exceed the measurement errors involved, at least up to puberty (Tanner *et al.*, 1966b; Cole, 1993). Annual velocities are however of little use in infancy, which is marked by rapid growth, particularly sensitive to environmental changes. In infancy, velocities are therefore often based on increments over a period of one or several months, which is justified by the higher growth rates observed during this period (Himes, 1999).

Analogous to a distance reference, a velocity reference shows the distribution of increments or velocities at consecutive ages in a particular reference population. Because velocity references have an additional covariate, which is the length of the time interval, they are often limited to a single interval width (for instance, annual velocity). Some exceptions however exist (Ayatollahi, 2005). Although velocity references allow the assessment of a dynamic aspect of growth, they suffer from the same limitations as distance charts. The variance is larger in early infancy,

when centile crossing is common, and in puberty. The phase-difference effect of pooling data from children that mature at different ages is even more dramatic for growth velocity compared to growth distance (figure 1.1).

Velocities are more variable than distances. While distances are highly correlated ($r > 0.9$) and tend to track along the distance curve, velocities have a low correlation ($r < 0.3$), and appear irregular. This makes sense since, for instance, Tanner *et al.* (1966b) would have observed a normal range of height in adult British males ranging from 140 to 240 cm if their subjects tracked along the 3rd or 97th velocity percentile (Tanner *et al.*, 1966b). The property of uncorrelatedness of successive velocities was used by Healy *et al.* (1988b) to propose referral criteria based on two moderately low velocities since their probabilities may be multiplied. Cole (1997) used it to estimate the approximate risk of falling one or more centile band widths on a cross-sectional distance chart during successive measurement periods (see section 1.1.3). Compared to distance, growth velocity is more efficient (velocities have a smaller standard deviation), but the gain in efficiency is somewhat smaller during the first months of postnatal growth, when the growth rate is more variable, and for velocities over a wide interval. Increments and velocities can however be biased because they do not account for regression to the mean, as discussed in section 1.1.3.

Early estimates of growth velocity were based on differencing the distance curve. This produces a correct estimate of the mean increment curve, but not of the variance, and hence the distribution is not characterized (Tanner, 1951). Proper annual increment references for height and weight from birth to maturity have been published in the UK (Tanner *et al.*, 1966b, Tanner and Buckler, 1997), Sweden (Karlberg *et al.*, 1976), USA (Tanner and Davies, 1985), Belgium (Hauspie and Wachholder, 1986), and Switzerland (Prader et al., 1989). All these references are based on distances taken one year apart, usually with a small allowance (0.85 - 1.15 years) but scaled to exactly one year. Roche and Himes (1980), and Baumgarter *et al.* (1986), created 6 month increment references by interpolating serial data from the Fels study to exact ages. Roche and Himes (1980) did not observe seasonal variation, and created biannual increments, which they considered to be more useful in clinical practice.

Increments over shorter intervals in infancy were generated by Roche *et al.* (1989a, 1989b) and Guo *et al.* (1988) using three parameter growth models to estimate individual distance curves in the Fels study (Roche *et al.*, 1989a, Guo *et al.*, 1988) and in Canadian infants (Roche *et al.*, 1989b). Guo *et al.* (1991) extended this data set and modelled growth up to 24 months, but used an otherwise comparable approach to obtain increments over 2 and 3 month intervals. These studies are particularly interesting because individual growth curves were estimated using a mathematical model, which is a smooth idealization of the growth process, free of measurement errors. Several authors have however shown that the reduction in variance is much larger than expected from the elimination of measurement errors alone (Piwoz *et al.*, 1992; van 't Hof *et al.*, 2000; de Onis *et al.*, 2011). This is the result of using a growth prediction model, which also fails to account for other sources of short-term random variation in growth (Hauspie and Molinari, 2004). The practical use of these references is therefore limited. More recent references for infants that are based on observed increments in growth over one or more months have been published by Boryslawski (1988), Michaelsen *et al.* (1994), Chrzastek-Spruch *et al.* (1994), the Eurogrowth consortium (van't Hof *et al.*, 2000), Ayatollahi (2005), and the WHO (2009). The WHO did not rescale their increments to exact intervals which they felt to be impractical in field circumstances.

Despite the potential merits of velocity references, they have never been used on the same scale as growth distance charts. Most likely this is due to a combination of practical constraints (a relative strict timing of measurements; the need for extra reference charts), difficulties with their interpretation (unfamiliarity with growth velocity; large fluctuations may appear counterintuitive), and the limited availability of increment references, who are largely outnumbered by growth distance charts. In fact, the majority of representative velocity references from the past 50 years were listed in this paragraph.

1.1.3 Conditional growth

Healy (1974) suggested to use regression standards as an alternative for growth velocity, i.e. to predict stature at age t from the stature observed one year before (t –1). Regression standards were described by Tanner (1952) as a means to estimate

the percentile distribution of a variable, given the value of another (correlated) variable: for instance weight for a *given* height, or otherwise said, weight *conditional* on height. A well known example of regression standards are those for height of children conditional on the height of their parents (Tanner *et al.*, 1970). The first conditional references for longitudinal growth were created by Cameron (1980) using a simple linear regression model of height on height one year (0.85 - 1.15 yr.) before:

$$H_t = a_t + b_t H_{t-1} + e_t \quad , \; e_t \sim N(0, \sigma_{e_t}^2) \, .$$

The residual standard deviation (σ_{e_t}) can be used to construct percentile curves for the conditional height $H_{t|t-1}$. Regression coefficients and the residual variances were estimated in yearly age groups from 5 to 16 years. Cameron did not rescale the increment, but added a second term for the deviation from the target interval width of one year (Cameron, 1980).

Berkey *et al.* (1983) used the same model for ages 1 to 6 years, but based on heights and weights estimated at exact ages with a Jenss-Bayley growth curve model. Under the assumption of a bivariate Gaussian distribution of height and log(weight) at different ages, they summarized the dependence of length or log(weight) for any combination of these ages in a correlation matrix. The conditional mean height is given by:

$$H_{t|t-1} = \mu_t + r_{t,t-1}\sigma_t (H_{t-1} - \mu_{t-1}) / \sigma_{t-1} \; ,$$

with a corresponding variance of $\sigma_t^2 (1 - r_{t,t-1}^2)$, where $r_{t,t-1}$ is the correlation between the measurements or variables involved. This system is very flexible since any combination of ages, and even covariates (for instance weight for a given height) is covered, on the condition that the required correlation is available. Berkey *et al.* (1983) acknowledge that their correlations are probably too high because they used a Jenss-Bayley model to estimate height and log(weight), which also eliminated measurement errors and random variation in individual growth (see above). Correction factors for rescaling correlations when measurement errors are known are provided by Berkey *et al.* (1983) and Cole (1997).

A major weakness of the approach by Berkey *et al.* (1983) is the assumption of a multivariate Gaussian distribution, which is unlikely to hold for log(weight). Cole (1994b, 1995) estimated conditional growth based on SD scores, which are Gaussian distributed with zero mean and unit variance, if a properly matched distance reference is used. A side effect of using SD scores is a further simplification of the regression equation for conditional growth since $\mu_t = 0$ and $\sigma_t = \sigma_{t-1} = 1$. Given two measurements H_t and H_{t-1}, with corresponding SD scores SDS_t and SDS_{t-1}, the *conditional distance* SDS at time t is obtained from (Cole, 1994b; Bailey, 1995):

$$\mathrm{SDS}_t = r_{t,t-1}\mathrm{SDS}_{t-1} + e_t \quad , \quad e_t \sim N(0, 1 - r_{t,t-1}^2) \, .$$

Percentiles for the conditional SDS are given by:

$$Q_{t|t-1,\alpha} = r_{t,t-1}\mathrm{SDS}_{t-1} + z_\alpha \sqrt{(1 - r_{t,t-1}^2)} \quad ,$$

where z_α is the normal equivalent deviate for the proportion α. The difference $\mathrm{SDS}_t - r_{t,t-1}\mathrm{SDS}_{t-1}$ is a *conditional velocity* with zero mean and variance $(1 - r_{t,t-1}^2)$. The SDS of the conditional velocity is thus obtained from (Cole, 1994b):

$$\mathrm{SDS}_v = (\mathrm{SDS}_t - r_{t,t-1}\mathrm{SDS}_{t-1}) \Big/ \sqrt{(1 - r_{t,t-1}^2)} \, .$$

When a poorly matched distance reference is used, the condition of zero mean and unit variance of the SD scores does not hold, and the models from Cameron (1980) or Berkey *et al.* (1983) should be used, with appropriate estimates for the mean and variance of the SD scores, and corresponding regression coefficients or correlations (Wright *et al.*, 1994). The advantage of working on a SDS scale is however largely lost.

Compared to (unconditional) increments or velocities, conditional growth is under certain assumptions more efficient and unbiased. The *efficiency*, expressed as the ratio of the variances of conditional to unconditional velocity, depends on the correlation and is given by $\mathrm{R} = (1 + r)/2$, assuming equal variances of the growth distance at ages t and t–1 (Cameron, 1980). Since distances are highly correlated, the gain in efficiency is relatively small compared to unconditional velocity, but

larger compared to distance (Cameron, 1980; Cole, 1993). A more important property of conditional growth is that it is *unbiased*, because it corrects for regression to the mean. On the SD scale, this bias is proportional to the correlation and the initial distance: the bias is larger when the correlation is small, and it is larger in the tails of the distribution (Cole, 1993). For this reason, conditional increments are more efficient and less biased in early infancy, when successive distances are less well correlated. The effect of regression to the mean is however small, with a corresponding small gain in efficiency (5% or less) throughout most of childhood (Cole, 1993).

Correlation models

The assessment of conditional growth on the SD scale only requires a suitable growth distance reference to convert measurements to SD scores, and the correlation matrix. In principle, we can estimate conditional growth over any interval covered by the correlation matrix. Correlation coefficients are usually estimated within age groups (see for instance Cameron, 1980) or between exact ages (for instance Berkey *et al.*, 1983; and Boryslawski, 1988). However, the correlation coefficient for any combination of (exact) ages within the boundaries of the (discrete) correlation matrix can be obtained by interpolation or with a suitable statistical model (Cole, 1993). Cole (1995, 1998) fitted a 6 term fractional polynomial to the Fisher z transformed correlations for weight SDS between birth and 24 months. Wade and Ades (1998) estimated correlations in length from birth to 8 years of age with an extended autoregression model. Argyle *et al.* (2008) used a 2 parameter model with interesting properties, $r_{t_1,t_2} = \left[(\tau + t_1)/(\tau + t_2)\right]^{\lambda}$, but the approximation was not always satisfactory.

Annual correlations for height in childhood were published by Bailey (1994), based on data from Tanner *et al.* (1966a), and by Cole (1994b, 1997) based on longitudinal data from French children. Correlations estimated from age-grouped data on the measurement scale require adjustment for the change in mean and variance within age groups (Goldstein, 1981). For variables with a skewed distribution, like weight, the correlation should be estimated from the SDS transformed variable, and not on the measurement scale (Cole, 1993; Pan and Goldstein, 1997). Correlation matrices based on monthly or quarterly weight SD

scores in infancy are available in Cole (1995, 1998) and Wright *et al.* (1994). The dependency on a particular population is not well known. Cole (1994b) showed that annual correlations between successive heights in French children differed little from those in their UK peers, but the correlation may be sensitive to sample selection (Argyle *et al.*, 2008) and measurement errors (Cole, 1994b).

Conditional growth in practice

According to the WHO, conditional references never have caught on in clinical practice (WHO, 2009). This is attributed to the same practical constraints that inhibited the use of velocity references, and to the extra level of complexity involved in deriving and implementing such references. Cameron (1980) needed a minimum of 15 different charts to cover longitudinal growth in both sexes from the age of 5 to 16 years: a single chart combining both sexes and ages 5 to 9 years, and yearly charts by sex from 9 to 16 years, including a nomogram to correct for small deviations from the yearly interval. This is, mildly put, cumbersome compared to the two annual increment curves (one for each sex) provided by Tanner *et al.* (1966b). Berkey *et al.* (1983) created a table that shows the probability of crossing one or more centiles between 3 months and 0.5, 1, and 2 years of age, and Bailey (1994) a table that shows the conditional distribution of height after one year of growth in children initially on the 1st, 2nd or 3rd percentile. Wright *et al.* (1994) used the difference between observed and predicted SD scores for weight in infancy to define a "thrive index" (SDS_{obs} - SDS_{pred}), assuming a constant error SD of 0.9, and created a table with the median and lower limit (set at the 5th percentile) of gain in weight between 6 weeks and 18 months. Wright *et al.* (1998) used the thrive index to create a chart with additional percentiles, such that a fall of more than 2 channel widths occurs in 5% of children. Thrive lines were unequally spaced to account for regression to the mean. A comparable chart with equally spaced lines was also developed in Sheffield to detect abnormal weight gain or weight loss in infants over 2 or 8 week intervals (Emery *et al.*, 1985). Cole superimposed conditional annual height velocity on a distance chart (Cole, 1994b); median conditional weight velocity over successive 4 week periods on a SDS diagram with contour lines of constant weight (Cole 1995); and the 5th and 95th percentile of conditional weight velocity on a distance chart (Cole, 1998). In the latter publication Cole also

suggested to create transparent chart overlays that show conditional growth over other intervals.

It is clear that a two-dimensional representation fails to take full advantage of conditional growth. For this reason, and those given above, the WHO decided not to pursue this, and only published unconditional increments (WHO, 2009). The implementation in computer applications is however no more complicated than adding the required correlation matrix or a suitable prediction algorithm (Berkey *et al.*, 1983; Cole, 1998). Cole (1997) also used a related approach to estimate the proportion of children that fall one or more centile spaces on a cross-sectional distance chart: since the mean change in SDS in the population is zero, and the corresponding variance $2(1 - r)$, the desired proportion can be estimated from the correlation alone. However, this approximation does not correct for regression to the mean because the initial SDS is not taken into account (Cole, 1997).

Conditioning on multiple previous measurements

In principle, conditional growth can be extended to include more previous measurements (and other covariates, but this will not be pursued here). Berkey *et al.* (1983) used partial correlations to condition on two or more measurements, while Thompson and Fatti (1997) proposed a multivariate generalization of conditional growth. Others partition the variance in between- and within-subject components using Bayesian updating (Fatti *et al.*, 1998), multilevel models linear in the parameters (Pan and Goldstein, 1997; Royston, 1995), and nonlinear multilevel models (Stirnemann *et al.*, 2012). Except for Berkey *et al.*, (1983) and Pan and Goldstein (1997), all authors applied their method to fetal growth, which is marked by high velocities and sparse, irregular spaced measurements. The practical application often relied on approximation algorithms for the covariance, which may have significant consequences for the estimated limits of conditional growth. All methods listed so far assume a Gaussian distribution of the conditional growth, but recently several semi-parametric estimators have been proposed using regression quantiles (Wei and He, 2006; Mayr *et al.*, 2012; Chen and Müller, 2012). They lack the convenience of producing a z-score, but could be useful if no suitable distributional model exists.

Models that are based on multiple measurements assume that children somehow grow along a personal path that can be predicted. According to results from Hermanussen *et al.* (2002), many such paths are plausible. Cole (1994b) showed that there is little to be gained from conditioning on more then one measurement in childhood, but growth is negatively correlated with height two years earlier in early and late puberty.

1.1.4 Longitudinal growth patterns

A problem that was briefly mentioned, but otherwise carefully avoided in the previous sections is the assessment of growth at adolescence. In infancy and childhood, conditioning on multiple measurements is of little use, and patterns of growth over several years are not informative about the normality of growth (see previous section). In adolescence, on the contrary, increments and conditional growth are of little use because it is not possible to distinguish, for instance, a delayed growth spurt in an early maturing child, from a normal timed spurt in a late maturing child. The pubertal growth spurt is characterized by the timing (tempo), duration and intensity (velocity), and can only be assessed when (many) more then two measurements are available. Tanner *et al.* (1966b) showed that the duration of the pubertal spurt is fairly constant. Adjusting chronological age for age at peak velocity would allow to assess growth the same way as in infancy and childhood. Unfortunately no suitable marker for the tempo of growth has yet been identified, apart from the age at peak velocity itself. Skeletal maturity (bone-age) and pubertal development are better correlated with tempo than chronological age, but still not sufficiently close (the standard deviation of the difference is reduced by approximately 50%; Tanner *et al.*, 1966b). Tanner *et al.* (1966b) determined age at peak velocity by visual inspection of velocity curves, but nowadays this age can be estimated more objectively using a mathematical growth model (see Hauspie and Molinari, 2004). To produce a reliable estimate of the growth process, these models however require a relative high density of measurements, including near adult height. Their use for the assessment of growth in a clinical context is therefore limited.

For these reasons, the clinical assessment of growth at adolescence is, even today, based on a qualitative inspection of the growth curve. Informal guidance is given

by tempo-conditioned growth charts that show the distance and velocity curves of typical early, average and late maturing children (Tanner *et al.*, 1966a). Tempo-conditioned growth charts have been developed for children living in the US (Bayley, 1956; Tanner and Davies, 1985; Berkey *et al.*, 1993), UK (Tanner *et al.*, 1966a, 1966b; adapted in Tanner and Whitehouse, 1976; Tanner and Buckler, 1997), Sweden (Karlberg *et al.*, 1976), Belgium (Wachholder and Hauspie, 1986; Hauspie and Wachholder, 1986), Ireland (Hoey *et al.*, 1987), and Switzerland (Prader *et al.*, 1989).

The charts by Tanner *et al.* (1966a, 1966b), Tanner and Davies (1985), and Hoey *et al.* (1987) combine estimates of average growth from national cross-sectional studies, with patterns of growth observed in smaller longitudinal studies. All charts listed above were created by aligning velocity curves around a common age at peak velocity, except Hauspie and Wachholder (1986) who used mean-constant curves based on Preece-Baines model I to estimate typical growth in early average and late maturing children. Tempo-conditioned growth charts are essential for clinical follow-up of growth during puberty, but their use in a non-specialist setting is limited.

1.1.5 Pubertal development

Pubertal development is measured by the occurrence of menarche in girls and the appearance of secondary sex characteristics in boys and girls. These are markers of sexual maturation, which is closely related but distinct from physical growth. Maturation is the acquirement of functional capacity (c.q. reproductive capacity), and is generally observed as a progression of qualitative or quantitative changes in function or appearance, while growth refers to changes in size. For a standardized assessment of pubertal maturation, Tanner assigned the development of secondary sex characteristics (breast development in girls, genital development in boys, and pubic hair in both sexes) to 5 distinct stages (Marshall and Tanner, 1969, 1970). These stages are scored by clinical examination using photographs or drawings and a description of typical features of each stage as a reference.

The assessment of normal pubertal development is an assessment of the current status (*status presens* or *status quo*), since the transition from one stage to the next

cannot be observed. Current status data however indicate that this event has occurred. This is illustrated in figure 1.2 for the occurrence of menarche. The bars on this chart show the proportion of post-menarcheal girls in each age group, which corresponds to the empirical cumulative distribution of the age at menarche. If the distribution of this age is Gaussian, the cumulative Gaussian distribution curve is a natural estimator for these data, i.e. *probit regression*. The age where menarche has occurred in half of the girls is the median age at menarche (or mean age, in case of a Gaussian distribution), the age where 25% of the girls have reached menarche the 25th percentile etc... (figure 1.2). On a growth reference chart, the percentiles of the age distribution of menarche and secondary sex characteristics are usually shown as a line (parallel to the age axis) with markers that coincide with a chosen set of age percentiles (e.g. percentiles 3, 10, 25, 50, 75, 90, 97).

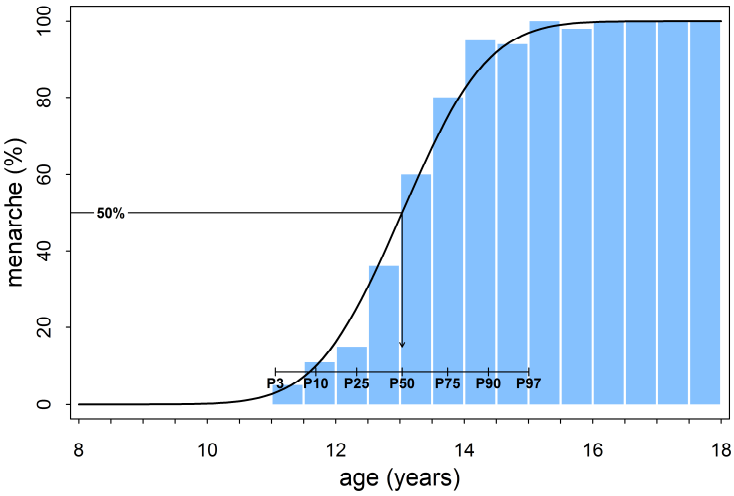


Figure 1.2: Occurrence of menarche in girls: proportion in discrete age groups (bars) and probit estimate (line). The arrow indicates median age at menarche (13.04 years); other age percentiles are shown on the horizontal line (data from Vercauteren and Susanne, 1985).

The distribution of ages is estimated for each variable and each stage separately, using probit analysis of grouped current status data (Finney, 1971), or from ungrouped current status data with generalized linear models (Nelder and Wedderburn, 1972). Age at menarche can also be estimated from a mixture of

current status and recall data using maximum likelihood estimators (Marshall and Tanner, 1969; Swan 1969; Atwood and Taube, 1976). All these estimators assume a Gaussian distribution of ages, which does not always hold. In this case, percentiles can be estimated with a generalized additive model (Mul et al., 2001). Alternatively, the age distribution can be estimated with non-parametric or parametric survival analysis of a combination of current status data and recall data.

References for the age distribution of pubertal status scores are cross-sectional, and give no indication about the speed of progression to a subsequent state. Estimating the timing between consecutive stages is generally problematic because the transition from one stage to the next cannot be directly observed, and requires longitudinal measurements, preferably taken close in time. Marshall and Tanner (1969, 1970) provide estimates assuming a log-Gaussian distribution of transition times. More recently, the timing and mean ages at transition were estimated by fitting logistic functions to individual yearly longitudinal data (Marceau *et al.*, 2011). Clinical validation of such references is however important to verify the extent to which they are affected by the modelling procedure.

1.2 Growth charts

Growth charts show the distribution of an anthropometric variable along a covariate (usually age) in a group of children. There is a strong age dependence of anthropometric measurements, and growth charts provide the required context for growth assessment. There are two main types of growth charts: growth *reference* charts are estimated from a representative sample of the target population, and growth *standards* from a group of children assumed to grow optimally. The difference between references and standards is in essence a difference in inclusion criteria, but curve estimation is otherwise identical. Sampling for references and standards is discussed later. In this section we review methods commonly used for estimating postnatal growth charts.

1.2.1 Curve estimation

Growth charts show the percentile distribution conditional on age, and that is precisely how they were created for almost a century: empirical quantiles calculated in contiguous age groups, and smoothed across age. This method is inefficient (extreme empirical percentiles have a relative large variance and can be highly irregular), but illustrative for the setup and interpretation of a growth chart. Efficiency is however important because we are mainly interested in the tails of the distribution, i.e. where data are scarce. As growth charts typically cover a wide age range, the sample size required to achieve a certain precision of the outer percentiles should also be multiplied by the number of age groups. Improvements over the past decades include the use of a statistical distribution for the measurement conditional on age, and the treatment of age as a continuous covariate. Some of these methods are given below. They have typically evolved from *"bin and smooth"* methods, based on age grouping, to methods that treat age as a continuous covariate. The latter choice is preferred, because age grouping is arbitrary, and a different grouping may result in slightly different estimates. When the average value changes within age groups, unbiased parameter estimates require a uniform distribution of ages, and variance estimates should be downscaled (Healy, 1962).

Mean and SD models

Observing that height was approximate Gaussian distributed by age (a small deviation in infancy and puberty was ignored), Tanner *et al.* (1966a; 1966b) obtained percentile curves from smoothed estimates of the mean (M) and standard deviation (SD) within age groups:

$$Q_\alpha(t) = M(t) + K \cdot SD(t),$$

where, for reasonably large n, the multiplier K is the normal equivalent deviate that corresponds to the percentile 100α, and the SD is estimated with an appropriate correction for age grouping (Healy, 1962). Distributional models have a number of important advantages: (1) The precision of the percentiles is higher, or alternatively, smaller sample sizes are required to obtain the same precision. The

gain in efficiency is moderate for the median (the standard deviation of the empirical quantile estimator is 1.6 times larger), but increases for more distal percentiles (2.3 times larger at the third percentile) (Healy, 1974). (2) Measurements can be converted to exact z-scores. (3) The whole distribution is characterized at once, rather than a discrete set of percentiles. This property also precludes sudden changes in the relative distance between centiles or even centile crossing. (4) Extreme percentiles (for instance -2.5 or -3 SD) can be calculated from the distribution. However, this should be done cautiously, because it relies heavily on the distributional assumption (notably the absence of skewness and kurtosis), and the precision of the SD. The precision of a percentile also decreases with its corresponding normal equivalent deviate (Healy, 1974).

Despite their convenience, Gaussian *mean and SD* models are not suitable for many anthropometric measurements, because the requirement of a Gaussian distribution is not met. Some authors have used other statistical distributions (for review, see Borghi *et al.*, 2006), but a more popular approach is to transform the measurements to an approximate Gaussian distribution (Wright and Royston, 1997). One such method, which corrects skewness, but not kurtosis, is the 3 parameter Box-Cox family of power transformations:

$$y^{(\lambda)} = (y^{\lambda} - 1)/\lambda \text{ , when } \lambda \neq 0; \text{ and } y^{(\lambda)} = \log(y) \text{ , when } \lambda = 0,$$

which is defined for $y > 0$. Realizing that a single power transformation is too inflexible for the complex shape of growth curves over wide age ranges, van 't Hof *et al.* (1985) calculated smoothed optimal values of λ within age groups to transform measurements conditional on age.

LMS method

Arguably the most popular method is the LMS method proposed by Cole (1988) and Cole and Green (1992). Cole (1988) used the Box-Cox transformation for an alternatively scaled variable (y/μ):

$$y^* = ((y/\mu)^{\lambda} - 1)/\lambda \text{ , when } \lambda \neq 0; \text{ and } y^* = \log(y/\mu) \text{ , when } \lambda = 0,$$

where μ is the median of y, and σ, the SD of y^* is (approximately) the coefficient of variation (CV) of y. The maximum likelihood estimator of λ is the value which minimizes the SD of y^* (or CV of y). Cole (1988) estimated λ, μ, and σ within age groups, which were smoothed across age to obtain $L(t)$, $M(t)$, and $S(t)$ or "LMS"-curves. Given a measurement y at time t, the corresponding SDS can be obtained by:

$$z = [(y / M(t))^{L(t)} - 1] \big/ L(t)S(t) \text{ , when } L(t) \neq 0,$$
and
$$z = \log(y / M(t)) \big/ S(t) \text{ , when } L(t) = 0.$$

Percentiles are obtained from the relation:

$$C_{100\alpha}(t) = M(t)[1 + L(t)S(t)z_{\alpha}]^{1/L(t)} \text{, when } L(t) \neq 0,$$
and
$$C_{100\alpha}(t) = M(t)\exp(S(t)z_{\alpha}) \text{ , when } L(t) = 0,$$

where z_{α} is the normal equivalent deviate that corresponds to the percentile 100α. In a follow-up paper, Cole and Green (1992) estimated the $L(t)$, $M(t)$, and $S(t)$ curves over all ages t using cubic splines by maximizing the penalized likelihood. The degree of smoothing is determined by a smoothing parameter (one for each curve), or more conveniently, by the *equivalent degrees of freedom* (EDF) (Cole and Green, 1992). Two EDF corresponds to a straight line (an infinitely smooth curve), and larger EDF allow more complex shapes for the corresponding parameter curve. EDFs are strictly not defined below two, but by convention one EDF is taken for a constant estimated from the data, and zero EDF when this constant is user provided (Pan and Cole, 2004).

The LMS method is attractive because it has all the advantages of *mean and SD models*, but is more widely applicable by removing skewness from the data. The LMS curves have a biological meaning (skewness, median and CV) that is reasonably well understood (figure 1.3). The LMS method is however not without limitations: it is only defined for strict positive values of y (which suits anthropometric measurements of size, but growth velocities can take negative

values), and it does not account for kurtosis. The effect of kurtosis is usually small for percentiles that are not too extreme (Cole *et al.*, 1998).

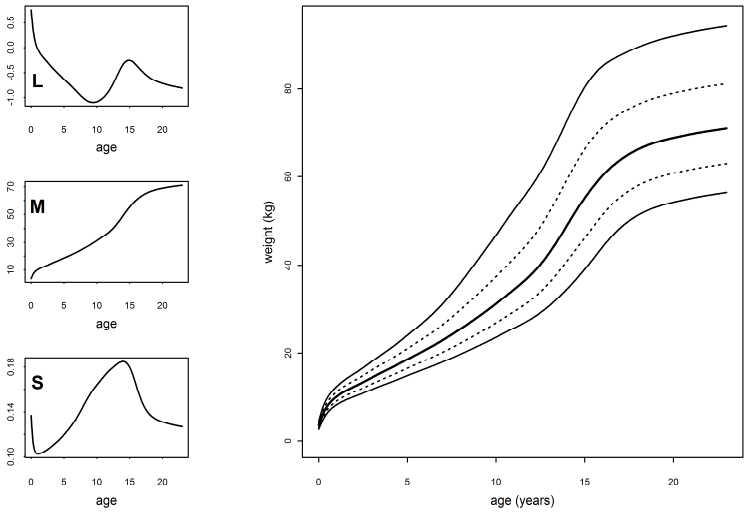


Figure 1.3: Smooth percentile curve of weight for age in boys (median and ±1 and ± 2 SD; right panel), and corresponding curves of the Box-Cox power (L), median (M), and coefficient of variation of the transformed variable (S) (panels on the left) (data from Freeman *et al.*, 1995).

GAMLSS: Generalized additive models for location, scale and shape

GAMLSS (Rigby and Stasinopoulos 2005) are a more general class of models, of which the LMS method is a special case. The GAMLSS framework includes several skew, kurtotic, continuous or discrete measurement distributions. Distribution parameters (mean, variance, and one or more shape parameters, e.g. skewness and kurtosis) can be modelled as a parametric or additive (non-parametric) function of one or more covariates and random effect terms. GAMLSS are fitted by (penalized) likelihood. The LMS method by Cole and Green (1992), for instance, is a GAMLSS with a Box-Cox normal (Gaussian) distribution, whose three parameters (mean, variance, skewness) are modelled as a function of a single covariate using cubic splines. Other distributions of interest are the Box-Cox *t*-distribution for leptokurtotic data (Rigby and Stasinopoulos, 2006), and the Box-

Cox power exponential (BCPE) distribution for both platy- and leptokurtotic data (Rigby and Stasinopoulos, 2004). The LMS method is a limiting case of the Box-Cox *t*-distribution (with infinite degrees of freedom) and a special case of the BCPE distribution (with $\tau = 2$). Based on recommendations from a WHO advisory committee, the WHO used the BCPE distribution for their growth standards, because it has the ability to remove skewness and kurtosis when necessary (Borghi *et al.*, 2006). However, only the special case, $\tau = 2$, where the power exponential and Gaussian distributions coincide, was actually used, which corresponds to the LMS method (WHO, 2006a).

Other estimators

Cole (1993), Wright and Royston (1997) and the WHO (Borghi *et al.*, 2006) reviewed more than 30 methods for curve estimation, of which only three (distributional) methods were discussed in this section. Notably absent are non-parametric estimators, which have been studied intensively in the literature. Some limitations of distribution free methods have been resolved, for instance, communality of quantile curves (Healy *et al.*, 1988a; He, 1997; Heagerty and Pepe, 1999), but distributional methods remain in general more straightforward to implement.

Longitudinal data

The majority of growth curve estimators assume a set of cross-sectional (i.e. independent) observations. However, cross-sectional curves are sometimes estimated from longitudinal data, i.e. including multiple (usually correlated) measurements from the same subject. A possible approach is to select only one observation per subject and to discard the rest, which is highly inefficient, or alternatively, to incorporate the correlation between measurements. Wade and Ades (1998) showed that the latter method has little effect on the estimated percentiles. Both discarding observations and ignoring the correlation are however liable to bias when the timing and frequency of measurements is related to the outcome (Wade and Kurmanavicus, 2009). This situation could occur, for instance, in fetal ultrasound measurements, but is considered unlikely in studies of post natal growth in healthy children (Borghi *et al.*, 2006).

Edge effects, non-uniform smoothing, and age transformations

Edge effects arise from a relative lower precision near the lower and upper boundaries of the covariate, and non-uniform smoothing could result from sparse observations at certain ages (Cole and Green, 1992). Cole and Green (1992) point out that these effects are quite general, and more a problem of sampling. A possible solution for edge effects is to extend the age range, such that boundary effects are small within the range of interest. The WHO, for instance, used ages up to 71 months for curve fitting, but only published growth standards up to 60 months (Borghi *et al.*, 2006). Extension of the age range is however not always feasible (for instance at birth), in which case the precision can be increased by local oversampling. At birth, the WHO used four times the original sample to account for a left boundary effect (Borghi *et al.*, 2006). More in general, Cole and Green (1992) suggest over-sampling at the extremes, and at ages where the LMS curves change rapidly (i.e. infancy and puberty). A third possibility is transforming the age scale prior to curve fitting. Age transformation increases the local density of observations, and may considerably improve the curve fit (Cole *et al.*, 1998). Using log(age), for instance, may help to model asymptotic behaviour at the right boundary. For monotone increasing or decreasing traits, Cole *et al.* (1998) also proposed a two-stage fitting procedure where the age is "rescaled" such that median curve is approximately linear.

1.2.2 Goodness of fit

Most postnatal growth curves over wide age ranges are non-parametric (smooth) in age, and model selection cannot rely on statistical inference (Wright and Royston, 1997; Royston and Wright, 2000). Smoothing is a trade-off between bias and variance of the estimate: smoother curves may discard true local features, while rougher curves may display features that are the result of sampling errors. Choosing the degree of smoothing is a subjective procedure, guided by several measures for *goodness of fit*. Informal, though powerful methods are: (1) inspection of the shape of the reference curve (van Buuren and Fredriks, 2001), or its decomposition in L, M, and S curves (Cole *et al.*, 1998); (2) a comparison of empirical and fitted percentiles; (3) a plot of the fitted curve with the data (van Buuren and Fredriks, 2001); and (4) a plot of quantile residuals (z-scores) by age

(Wright and Royston, 1997). Growth curves fitted with a distributional assumption (e.g., the Gaussian distribution in case of the LMS method) generate z-scores which are used as a starting point for some methods given below.

Deviance

As a rough guideline, Cole and Green (1992) suggested that the change in deviance (d = -2 log(penalized likelihood)) follows a χ^2_ν distribution with ν (= difference in e.d.f.) degrees of freedom. This test is analogous to a likelihood ratio test, but has no formal statistical justification (Cole and Green, 1992; Wright and Royston, 1997), and tends to select noisy overfitted models for large sample sizes (Cole et al., 1998; Pan and Cole, 2004). Rigby and Stasinopoulos (2005) selected GAMLSS models by minimizing the Generalized Akaike Information Criterion (GAIC(#) = deviance + #edf). Special cases are # = 2 (AIC, Akaike Information Criterion), and # = log(n), the Schwartz Bayesian Criterion (SBC). Higher values of the hyperparameter # will result in more parsimonious models.

Tests for a Gaussian distribution (normality tests)

Cole (1988) originally used the Shapiro-Wilk (W) statistic as an omnibus test for the Gaussian distribution of z-scores within age groups. Royston (in the discussion of Cole, 1988) suggested a composite p-value ($-2\Sigma\log(p_i) \sim \chi^2_{2G}$), to test for a uniform distribution of p-values over G age groups. The Shapiro-Wilk test is however insensitive to symmetric non-Gaussian alternatives (Royston, in the discussion of Cole, 1988), and does not indicate *how* the z-scores differ from the Gaussian distribution. The test is also conservative because it does not account for the transformation parameter (Cole, 1988) and model dimensions (Royston and Wright, 2000). However, due to the large sample sizes involved, the test is overpowered, and departure from the Gaussian distribution is easily obtained. If the degree of smoothing is guided by statistical inference for the Gaussian distribution, the model will usually be overfitted (van Buuren and Fredriks, 2001). Global tests for a Gaussian distribution of all the residuals are also insensitive to age varying skewness (Royston and Wright, 2000).

Gaussian quantile plot and worm plot

The Shapiro-Wilk (W) test and related Shapiro-Francia (W') test for large sample sizes, are tests for the correlation between empirical quantiles and the corresponding expected values from the Gaussian distribution. This relation is visualized in a Gaussian quantile plot, and the shape of this plot (or a detrended alternative) is used to interpret the nature of departures from the Gaussian distribution: the offset depends on the mean; the slope on the variance; a quadratic (parabolic) shape results from skewness, and a cubic (sigmoidal) shape from kurtosis. The presence of any of these features in the residuals indicates an inadequate fit of the corresponding moment of the standard Gaussian distribution. Recognizing that global Gaussian quantile plots are insensitive to age specific departures from the Gaussian distribution, van Buuren and Fredriks (2001) plotted a series of detrended Gaussian quantile plots from equal sized contiguous age groups in a grid (figure 1.4). The variety of shapes that the quantile plots can take, gave this procedure its name "worm plot" (van Buuren and Fredriks, 2001).

Q-tests: low order moments of the Gaussian distribution

Royston and Wright (2000) proposed so-called Q statistics for the low order moments of the Gaussian distribution (mean, variance, skewness, kurtosis). Every Q test is based on the sum of the corresponding test statistic in G age groups, which is compared to a χ^2 distribution with G–1 (mean, variance) or G (skewness, kurtosis) degrees of freedom adjusted for the dimension of the corresponding model. The choice for a particular number of age groups is however arbitrary and should ideally depend on the sample size (Royston and Wright, 2000). Pan and Cole (2004) considered Q-tests over a range of different age groupings, and presented the results graphically.

Grid test

Healy et al. (1988a) counted observations between percentiles in two to four contiguous age groups, and compared these counts with the expected frequency (e.g. 7% between percentiles 3 and 10, 25% between the percentiles 25 and 50, ...). Royston and Wright (2000) formalized this comparison by using a χ^2 test with $G*C$

degrees of freedom (for G age groups and C percentiles). The test is however conservative, because the degrees of freedom of the null distribution depend in a complex manner on the model parameters, sample size and dimension of the grid (Royston and Wright, 2000). The grid test is nevertheless often used informally, because it is very intuitive and closely resembles the purpose of a growth chart (Borghi *et al.*, 2006).

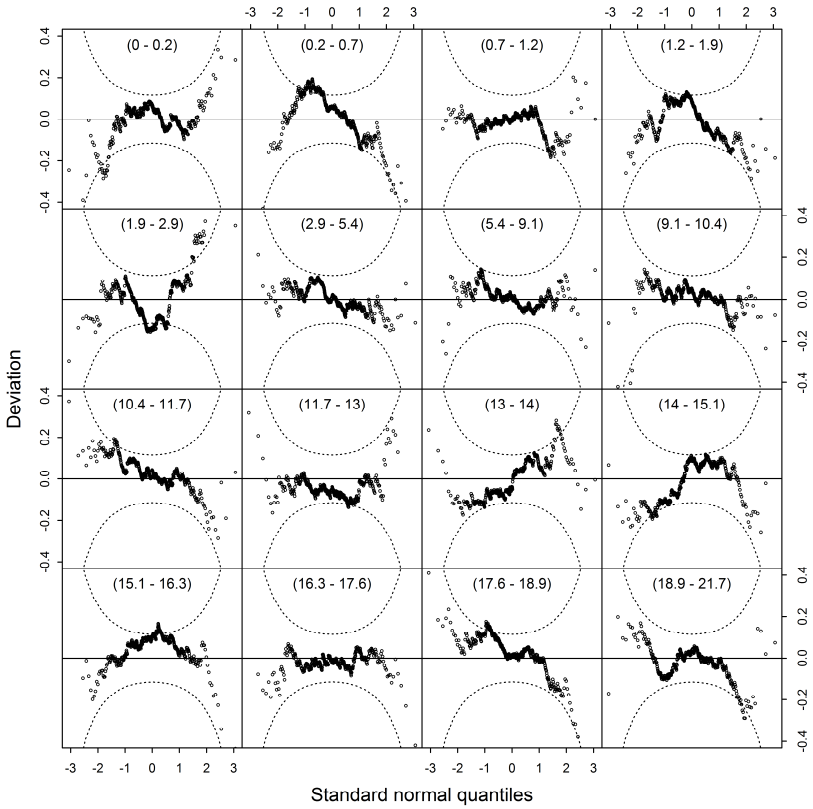


Figure 1.4: Wormplot (detrended Gaussian quantile plots) of LMS residuals from a model for height. Adapted from van Buuren and Fredriks (2001).

The goodness of fit tests in this paragraph had a focus on methods that generate Gaussian distributed residuals. However, some methods are quite general and can be used with any parametric or non-parametric curve estimation procedure. For instance, informal graphical methods and the grid test do not depend on a particular

distribution; Gaussian quantile plots and corresponding worm plots can be generalized to any statistical distribution; and the Kolmogorov-Smironov test is a general, though less powerful, alternative for tests for a Gaussian distribution. Some methods may have their own (inferential) criteria. A common strategy for LMS and GAMLSS models is to use the global deviance for initial model selection; and a combination of Q-tests, worm plot, and observed and expected frequencies for model validation (Pan and Cole, 2004; Borghi *et al.*, 2006).

1.3 Normal growth

The concept of *normal growth* is closely related to that of *normal ranges* or *reference ranges* for biochemical parameters in clinical chemistry. A reference range is based on the distribution of a parameter in healthy individuals, and is usually defined by the upper and lower boundaries of this distribution. Within these limits a healthy state is plausible and cannot be excluded. The definition of normal growth is built on the same principles.

In clinical chemistry the expression *normal ranges* has been largely abandoned in favour of *reference ranges* or *reference intervals*, because of several conceptual problems with the notion of "normal": (i) normal is not unambiguously defined, and can have several meanings from a statistical (Gaussian distribution, most representative, most common), epidemiological (most suited for survival), or cultural (conventional, ideal) perspective; (ii) normal implies abnormal, which is not necessarily true; (iii) the term normal suggests that the reference population is healthy, which is assumed but actually unknown; and (iv) an important flaw is that 5% is considered as abnormal (in case of a 95% reference interval), although the complete distribution is estimated from healthy individuals (Ceriotti *et al.*, 2009; Healy 1995; PetitClerc and Solberg, 1987; Wright and Royston, 1999). Adopting this alternative terminology for the limits of normal growth, may however create confusion with the term growth references, which is commonly used for charts, curves, tables, and models that describe growth, irrespective of the choice for a particular cut-off. We will therefore adhere to the definition of normal as what is commonly observed in a large majority of healthy children, but obviously some of the limitations outlined above still apply.

A formal definition of a *p*% reference interval is a lower and upper limit that encloses the central *p*% of a sample drawn from a specified target population (Royston, 2005). A reference interval is in general symmetric, such that 100*(1 – *p*)/2 percent of the values lie below the lower limit and the same proportion above the upper limit. A typical value of *p*% in clinical settings is 95%, which is usually approximated by the 3rd and 97th percentiles (94% coverage) or ± 2SD (95.4% coverage) cut-offs.

Properties of normal ranges

The properties of a normal range as a screening test for disease are given by the sensitivity, specificity, and predictive values of a positive or negative test (figure 1.5). If we define, for instance, height outside the normal range as a positive test (T+) and ill health as the diseased state (D+), then the sensitivity is the probability that disease will be detected by the test: sensitivity = P(T+ | D+). Specificity is the proportion healthy children identified as such: specificity = P(T– | D–). The predictive value of a positive test is the proportion of children with a positive test who actually have the disease, PPV = P(D+ | T+), and the predictive value of a negative test the proportion of children with a negative test who are indeed healthy, NPV = P(D– | T–). Sensitivity and specificity are properties of the test that do not depend on the prevalence of the disease (figure 1.5), but the PPV increases and the NPV decreases when the disease prevalence is higher in the population tested. A consequence of this property is that the predictive value may differ when a test is applied in a clinical context compared to a preventive health setting (Healy, 1995).

The choice for a particular cut-off is a trade-off between maximizing the detection of disease (high sensitivity), and minimizing the amount of erroneously referred children that are actually healthy (high specificity), usually against some gold standard for presence or absence of disease. A problem inherent to growth monitoring is the lack of an unambiguously defined gold standard, since abnormal growth (for instance, short stature) is not a disease as such, but indicative for an underlying disorder or condition that affects the growth process. The European Society for Pediatric Endocrinology listed more than 160 conditions, some very rare, that have a direct or indirect effect on growth (Wit *et al.*, 2008), but as a rule,

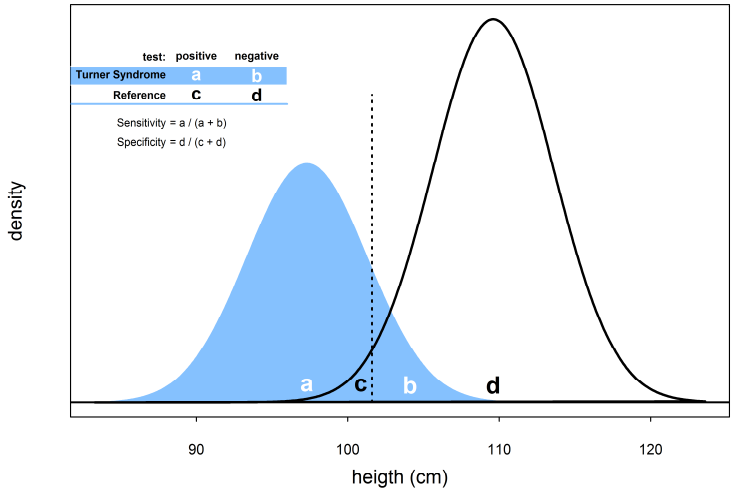


Figure 1.5: Distribution of height in a reference population of 5 year old girls (Prader *et al.*, 1989), and distribution of heights in girls with Turner Syndrome of the same age (TS; shaded area; Ranke *et al.*, 1983). The distribution for TS is disproportionate (occurs in < 1/2000 females). The dotted line shows a cut-off at -2SD in normal girls (specificity for short stature = 97.7%). The sensitivity depends only on the distribution of height in TS; the specificity only on the reference distribution.

any disease episode may result in (often reversible) growth faltering. Abnormal growth is therefore a highly unspecific criterion in the analytical sense, i.e. not indicative for a *particular* disorder, and further clinical assessment is warranted. A second problem is that "cases" are hard to define. Some authors report sensitivities of growth monitoring to detect a particular disease (which is moderate or low at best, if a reasonable specificity is maintained; Hall, 2000), but many conditions lack an unambiguously defined gold standard. Moreover, screening for a particular disorder is generally not the aim of growth monitoring. This puts growth monitoring in an awkward position: widely used as a sensitive indicator of health, but unlikely to meet formal criteria for a screening test (Hall, 2000; Garner *et al.*, 2000), and highlights the importance of maintaining a high specificity, i.e. a conservative approach. The specificity of a single measurement is determined by the choice for a particular cut-off percentile (e.g. 97% for the 3rd percentile), on the condition that a well-matched reference chart is used. The 3rd and 97th percentile have long time been the outer limits on growth charts, due to practical and statistical constraints (Cole, 1994a). However, on recent growth charts, limits that

warrant immediate referral in a public health setting are often more extreme, like the ± 2.67 SD (Cole, 1994a) or ± 3 SD (Grote *et al.*, 2008). When children are measured on multiple occasions and each of these measurements is assessed independently, the specificity is however likely to increase. For instance, van Buuren *et al.* (2004a) and Grote *et al.* (2007) investigated referral patterns in routinely monitored Dutch children, and concluded that the number of children that would be referred is much higher (up to 10 times larger) than theoretically predicted. For completeness, it should be mentioned that correction for parental height improves the outcome of growth monitoring in terms of sensitivity and specificity (van Buuren *et al.*, 2004b), but this will not be pursued in this research.

Selection of a reference sample

Reference ranges are tied to a particular target population, and their validity depends on the selection of subjects, the analytical variability (measurement error), and the statistical methods used to estimate the distribution (Ceriotti *et al.*, 2009). The consequences of an ill-matched reference for the specificity were already highlighted above. Reference ranges are usually estimated from a representative sample of healthy subjects. As the WHO definition of 'healthy' is not a realistic starting point, a more pragmatic approach is to include only subjects without visible signs of diseases that are related to the measurement. The selection of subjects should also account for all variables that have an effect on the measurement, by either considering them as a covariate (e.g. age, sex), or by proportionate inclusion in the reference sample (Ceriotti *et al.*, 2009).

Known determinants of growth include the biological inheritance (genetic background), and the proximate environmental factors: diet and nutrition, infection, psychosocial stress, pollution, food contamination, and hypoxia. These factors are often conditioned by socio-economic status or affluence, culture, behaviour, geographical region, and history (Ulijaszek, 2006). Based on parent-offspring and twin studies, it is estimated that the heritability accounts for 60 to 90% of the observed variation in height (Hauspie *et al.*, 1994; Thomis and Towne, 2006). The heritability of growth tempo and sexual maturation is in the same range, but the genetic factors involved may be quite different. Weight is more susceptible to environmental factors and has a somewhat lower heritability of 20 - 90% (Thomis

and Towne, 2006). Genome-wide association studies (GWAS) have identified 180 loci that are associated with height, but these accounted for only 10% of the variance (Lango Allen *et al.*, 2010). Testing all single-nucleotide polymorphisms together increased the explained variance to 45% (Yang *et al.*, 2010), but a significant part is still missing (Maher, 2008; Aulchenko *et al.*, 2009). The remaining factors in the list above are environmental factors, which contribute ~ 20% of variance for height. Proximate factors operate on the individual level, and the conditioning (distal) factors through a differential exposure in children (Ulijaszek, 2006).

In practice, a target population is often defined by a nation, but the reference sample is drawn from the native population only (e.g. Freeman *et al.*, 1995; Fredriks *et al.*, 2000). This restriction is imposed to avoid a high degree of heterogeneity (see for instance Fredriks *et al.*, 2003; and Fredriks *et al.*, 2004). A representative sample is obtained by the inclusion of subjects proportional to census data regarding determinants of growth (e.g. Fredriks *et al.*, 2000), or by assigning appropriate weights (e.g. Kuczmarski *et al.*, 2002). It is however not feasible to account for all determinants, and the representativeness is usually judged against the strong but distal socio-economic and geographical factors only. Samples drawn from populations with a diverse background (for instance the United States) may also aim at a proportional inclusion of different ethnicities (Kuczmarski *et al.*, 2002). The environmental factor 'history', synonymous for secular trends, is explicitly accounted for by using recent data. A consequence of a representative reference is that the specificity of a cut-off is defined for the total population, but it may be quite different for strata within the population if, for instance, growth depends on the socio-economic background or region (e.g. Fredriks *et al.*, 2000).

In this section, we have defined normal limits by their specificity, i.e. the distribution in healthy children, which is a common approach for most anthropometric measurements. Different approaches are however possible (and sometimes necessary), particularly for a target condition with a high prevalence, or when substantial health risks are involved. In such cases, normal limits can be determined by the sensitivity (i.e. based on the "ill" population) or by the risk for ill

health (section 1.4). A second alternative is to use a prescriptive approach, i.e. show how children should grow, an example of which is discussed in section 1.5.

1.4 Criteria for overweight

The principles of normal growth that were outlined in the previous sections do not apply for weight beyond infancy for two reasons: (1) weight is correlated with height, and (2) the reference population includes a large but unknown proportion of subjects that are overweight, i.e. the "unhealthy" part of the population that we want to identify. The first problem requires a different parameter; the second problem requires an alternative definition of "normal" limits.

Body Mass Index

In infancy, weight-for-age is used as a sensitive indicator for health and nutrition, but in childhood and adolescence the focus has largely shifted towards the detection of overweight (excess body *weight*), as an indicator of adiposity (excess body *fat*). Weight is correlated with body fat, but also with height, while height itself is weakly correlated with body fat (Bellizzi and Dietz, 1999). Weight adjusted for height would therefore be a more sensible indicator for body fatness For precisely the same reason, the Body Mass Index (BMI, weight in kilograms divided by the square of height in meters, kg/m^2) or Quételet Index, is widely used as an index of adiposity in adults (Keys et al., 1972; Garrow and Webster, 1985). The BMI also correlates with fat mass in children (Dietz and Bellizzi, 1999; Mei *et al.*, 2002), but is not completely independent of height in adolescence (Cole, 1986). The BMI is however generally accepted as the best overall compromise in all age groups, except in infancy, where the consequences of a low or high BMI are not well known (Dietz and Robinson, 1998).

Criteria for overweight

In adults, overweight and obesity are defined by absolute thresholds of 25 and 30 kg/m^2 (Garrow and Webster, 1985). However, in children and adolescents there is also a strong dependence on age, such that a BMI reference chart is needed (figure

1.6). BMI reference charts differ from the growth charts discussed so far, because the BMI is an anthropometric *index*, which is different from anthropometric *size*. While most growth curves increase monotonically with age, the BMI curve will rise and fall (figure 1.6). A second important difference with regular growth charts is the selection of a suitable reference population and choice of cut-off. Ideally cut-offs should be based on health risks, but this information is currently missing, and the consequences of overweight may only appear after many years. In the summary of a childhood obesity workshop of the International Obesity Task Force (IOTF), Bellizzi and Dietz (1999) listed several alternatives:

(1) *Fixed percentile cut-offs* are used in the United States (the 85th percentile for overweight and the 95th percentile for obesity; Barlow and Dietz, 1998), in the UK (+1.3 SD for overweight and +2 SD for obesity; Power *et al.*, 1997), and by the WHO (+1 SD for overweight and +2 SD for obesity on the international reference chart; de Onis et al., 2007b). Fixed cut-offs are generally based on the prevalence and observed health risks in a particular population at a particular time (Barlow and Dietz, 1998; Power *et al.*, 1997), and should not be simply applied to other populations. Also, when the reference chart needs updating, fixed cut-offs would return the prevalence to the level defined by the chosen percentile (e.g. 15% overweight in the US). Therefore, the CDC2000 growth charts did not include recent data on the BMI for children above 10 years of age (Kuczmarski *et al.*, 2002).

(2) *Normal limits based on a population with low levels of overweight*. This is the prescriptive approach that is discussed in the section on growth standards. The selection of a reasonable large reference population with a low prevalence of overweight is however problematic. Cole and Roede (1999) explored this idea by estimating the BMI for age curve based on the Dutch third national growth study from 1980.

(3) *A percentile cut-off chosen by an objective criterion*. The percentile that corresponds to a BMI of 25 and 30 kg/m^2 at the age of 18 years was proposed as a reasonable choice, consistent with the adult cut-offs that are used from that age onwards (figure 1.6). This approach assumes that there are no cohort effects in the sample, and that children show a tendency to track along the BMI chart (Bellizzi and Dietz, 1999).

International criteria for overweight and obesity

Cole et al. (2000) derived percentiles that correspond to a BMI of 25 and 30 kg/m^2 at 18 years of age from six large national representative data sets. The national curves for overweight and obesity were averaged by age to create an international criterion for overweight and obesity from 2 to 18 years of age. These curves, labelled IOTF-25 and IOTF-30 in figure 1.6, have been widely used ever since to obtain international comparable estimates of overweight and obesity (Lobstein *et al.*, 2004). Cole *et al.* (2007) also calculated international criteria for thinness based on WHO adult cut-offs of 18.5, 17, and 16 kg/m^2. Recently, Cole and Lobstein (2012) published LMS reference values that allow (with minor differences) to calculate these, and any other percentile from this composite data set.

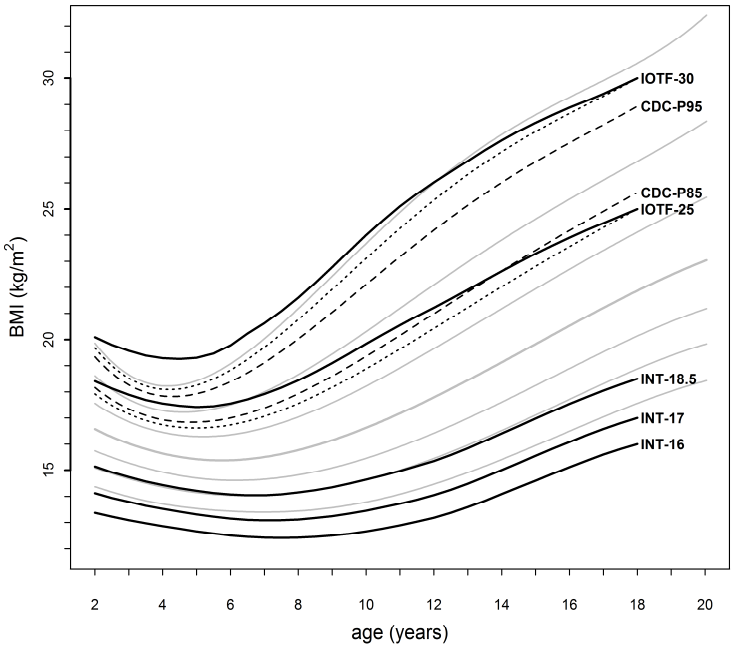


Figure 1.6: Different cut-offs for overweight and thinness in children plotted with the CDC-2000 BMI chart for boys in the background (Kuczmarski *et al.*, 2002): CDC percentiles 85 and 95 (dashed lines); international criteria for overweight and obesity (IOTF-25 and 30; Cole *et al.*, 2000), and thinness (INT-16, 17, 18.5; Cole *et al.*, 2007); and CDC percentiles that correspond to a BMI of 25 and 30 at 18 years of age (dotted lines).

1.5 Growth standards

The publication of the World Health Organization child growth standards in 2006 (WHO, 2006a) revived a debate that started decades ago (Ziegler and Nelson, 2012). Historically, this debate originated from the need for an appropriate (international) tool to assess the nutritional status of a group, and to assess growth and nutritional needs in individual children living in deprived areas of the world. This led first to the adoption of the existing Harvard and National Center for Health Statistics (NCHS) growth references by the WHO (Dibley et al., 1987), and finally to the creation of the WHO growth standards (WHO, 2006a).

Standards and references

There is an important difference between *growth references* and *growth standards*, but this difference is subject to a lot of confusion. In principle, growth standards show how growth *should be* (i.e. a prescriptive norm; a norm to aspire to), and references show growth *that is* (i.e. actual growth). According to the WHO, references should not be used to infer any meaning when comparing measurements, while a standard involves a value judgment (WHO, 1986). In this respect, growth references are actually used as standards, viz. section 1.3, where a reference interval was defined as *normative*. Because this use of references cannot be prevented, the WHO recommends that references should resemble, as far as possible, true standards (WHO, 1986). The major drawback of a true standard is the operational definition of the kind of growth it should represent: good growth, maximum growth, optimal growth,... . Ideally, growth standards should reflect growth that is associated with the greatest amount of health, but given that no suitable positive indicator of health exists (maybe apart from growth itself), this is often translated in minimum morbidity and mortality (Goldstein and Tanner, 1980; Waterlow, 1980).

Elite standards

In the developing world, growth is often very different in socio-economically deprived children compared to their more privileged peers. Under the assumption

that both groups share a common gene pool[1], these differences are entirely due to the environment, and so called *elite standards* would reflect achievable growth under optimal conditions (Falkner et al., 1972). Goldstein and Tanner (1980) argue against the use of elite standards, because they feel that individual children should be assessed within their environment for two reasons: (i) elite standards measure personal circumstances and environmental conditions at the same time. The target for an intervention is however different, since, in the latter case, the environment should be altered, not the individual; (ii) within a given environment, maximum growth is not necessarily a guarantee for minimum morbidity and mortality. Recent examples that support this are the morbidity associated with rapid catch up growth after intrauterine growth restriction (Ong *et al.*, 2000), and the high levels of overweight in privileged children from developing countries (Lobstein *et al.*, 2004). Elite standards may also generate political opposition, because they emphasize inequalities in health and wealth (Cole, 1993). The same is however true for "double standards", which indicate a level of growth that is good enough for one group but not for the other (Waterlow, 1980). Differences between subpopulations are usually less prominent in developed countries, which often have their own national (normative) reference. The screening value of a reference derived from a population with a large growth deficit is however questionable (de Onis and Yip, 1996).

International references

Assuming that the variation within a population is largely determined by genetic factors, and the variation between populations by environmental factors, a growth reference from a developed country could serve as a proxy for achievable growth of children in any country. Several studies have confirmed that between country differences in length of preschool children are largely (Habicht, *et al.*, 1974; Graitcer and Gentry, 1981), but not entirely (Eusebio and Nubé, 1981) driven by environmental constraints. However, from puberty onwards (from childhood onwards in south Asian populations) marked differences between populations start to appear (Ulijaszek, 2001; Haas and Campirano, 2006). For instance, average final

[1] This assumption is not always justified; in some countries ethnicity and socio-economic status coincide (Ulijaszek, 2001).

height of young adult males is 184.0 cm in the Netherlands (Fredriks *et al.*, 2000), and 176.8 cm in the US (Kuczmarski *et al.*, 2002). Differences of the same magnitude are also observed within countries: young adults of Turkish or Moroccan origin, living in the Netherlands, are on average 9 - 10 cm shorter than those of Dutch origin (Fredriks *et al.*, 2003; Fredriks *et al.*, 2004). For weight, entirely different results are noted in all age groups: (over)weight is inversely related with socio-economic status (SES) in developed countries, but directly proportional tot SES in developing countries (Gupta *et al.*, 2012; Lobstein *et al.*, 2004).

According to the WHO, the advantages of a global reference that allows international comparisons without losing the usefulness of local application, outweigh the disadvantage of not taking into account minor racial and ethnic differences (WHO, 1986). For this reason, they released a sexes combined version of the Harvard growth charts in 1966 (based on US data from 1930 – 1956), and from 1978 onwards a normalized version of the NCHS data set developed by the CDC (Dibley *et al.*, 1987). The WHO acknowledges the existence of ethnic differences between groups, just as there are differences between individuals, but for practical purposes they were not considered large enough to invalidate the general use of the NCHS population both as reference and as a standard [*sic*] (WHO, 1986). The choice for the NCHS data set was one of convenience as the data set did not comply with all the requirements that were forwarded by the WHO, but it was considered the most statistically complete (Waterlow *et al.*, 1977; Dibley *et al.*, 1987). The infants in the NCHS data set were recruited in the Fels study in 1929 – 1975. These children were not representative of the US population, and most were fed artificial milks. From two years onwards, data were representative of the US population 1960 – 1975. A review of the NCHS growth reference by the WHO showed that the majority of breastfed children, in both the developed and developing world, deviated negatively from the international reference curve (WHO, 1995a). In this study, and also in a follow-up study (WHO, 2000), between population differences were considered small (a range of approximately 0.5 to 1 SD included most, but not all, ethnicities).

WHO Multicentre Growth Reference Study (MGRS)

Recognizing the shortcomings of WHO/NCHS reference, the WHO decided to create a new growth chart that would reflect how children should grow, i.e. a standard. This study was conducted from 1997 - 2003 in six study sites in Accra (Ghana), Davis (USA), Muscat (Oman), Oslo (Norway), Pelotas (Brazil), and South Delhi (India). Study sites were selected because of their favourable socio-economic climate, altitude below 1500m, low mobility, high prevalence of breastfeeding (>20%), and lactation support. The standards are based on two distinct samples: a longitudinal sample of infants followed from birth to 24 months, and a cross-sectional sample of children from 24 to 71 months. Individual inclusion criteria for the longitudinal component were: single term birth, absence of significant morbidity, no environmental or economic constraints[2], no maternal smoking, exclusive or predominant breastfeeding for at least 4 months, introduction of complementary feeding by 6 months of age, partial breastfeeding to at least 12 months of age. Comparable inclusion criteria were used for the cross-sectional sample, except for the feeding requirements, where a minimum of 3 months of any breastfeeding was allowed (WHO, 2006b).

For the longitudinal part of the MGRS, 1743 newborns were included in the study, but only data from 882 (50.6% of included) infants, who complied fully with the maternal smoking and infant feeding requirements and completed the 24 month follow up, were used for curve fitting. The other half of the longitudinal sample contributed only the measurement at birth to account for left edge effects. The cross-sectional part is based on a single measurement from 6697 children (WHO, 2006b). Weights outside the ±3SD limits of weight for length (-3/+2 SD of weight for height in the cross-sectional sample; 1.4% in boys and 1.1% in girls) were excluded as they were considered unhealthy (WHO, 2006b). Growth curves were estimated with the LMS method (within the GAMLSS framework; see section 1.2). The WHO created both distance (0 - 60 months) and velocity (0 - 24 months) charts for a wide variety of anthropometric indicators, together with references for 5 developmental (motor) milestones.

[2] These criteria were pretty relaxed according to European standards: < 10 years of parental education was accepted (11%), and 41% of mothers and 36% of fathers completed less then 15 years. No Norwegian children and < 1% of US children were excluded for this criterion (WHO, 2006b).

A comparison of growth in length and height showed that site means deviated no more the 0.5SD from the MGRS mean after birth, but Omani, and to a lesser extend Indian children, were consistently below the overall mean; and Norwegian children were larger at birth and remained above average during the first 12 months (WHO, 2006c). Site differences for other traits were not presented, despite previous reports that showed a significant lower weight for age in Indian and Chinese children (WHO, 2000). Compared to the "old" WHO/NCHS reference, the breastfed MGRS infants are considerably taller and heavier during the first 6 months, and a little smaller and lighter thereafter (de Onis *et al.*, 2006). Consequently, prevalence estimates of wasting and stunting in 0 - 6 month old children are likely to increase when based on the same percentile criterion, but remain comparable thereafter (de Onis *et al.*, 2006; Kerac *et al.*, 2011). Similar trends are observed when the WHO weight and length standards are compared to the more recent growth references from the CDC (de Onis *et al.*, 2007a), the Netherlands (van Buuren and van Wouwe, 2008), France and the UK (Rolland-Cachera and Péneau, 2011), and the Euro-growth consortium (Ziegler and Nelson, 2012). All these references are however (and by definition) based on both breastfed and formula fed infants.

The high growth rates during the first 6 months were a concern to Binns and Lee (2007), who attributed this to the strict inclusion and exclusion criteria, which kept only high achievers in the sample. Bins and Lee (2007) felt that the standards therefore presented a target that was out of reach for the large majority of infants, and would impede rather than promote breastfeeding. The selection process was of course a crucial part of creating a standard, but the WHO also discarded all data from children that did not comply with the feeding requirements, including measurements that were taken before the breach of protocol with respect to feeding requirements (WHO, 2006b). An analysis by van Buuren (2010) showed that this could be problematic, if the breech of protocol was the result of (relative) poor growth. Based on longitudinal data of breastfed Dutch children, van Buuren (2010) showed that there was an inverse relation between the duration of breastfeeding and growth: children who grew better were breastfed exclusively for a longer duration and children who grew less were more likely to receive supplementary feeding. To date, the WHO did not publish any data on (growth in drop-out).

1.6 Summary and conclusions

The assessment of growth and development may involve a single measurement of attained size (distance), the timely occurrence of an event (tempo), the rate of growth between two measurements (velocity), or the patterns of growth over several years. These aspects imply an increasing level of complexity, and depend on appropriate assessment criteria, preferably based on a suitable, well-matched growth reference. Percentile charts of attained size or growth distance are now a common instrument, and research over the past decades often focused on methods to estimate these curves. Current curve fitting procedures usually combine parametric and non-parametric methods, for the distribution of size and how it changes with age, respectively. Assessing changes in size over time, i.e. the dynamics of growth, requires appropriate instruments. References for growth velocity are usually the norm for paper charts, but they are tied to a particular age and interval. Electronic growth charts are more flexible, and allow the use of alternative methods like conditional growth, which has the added benefit that it is – under certain assumptions – unbiased and more efficient compared to (unconditional) growth velocity. Patterns of growth over several years are the most complex aspect of growth assessment and particularly relevant during adolescence.

Renewed interest in growth monitoring – particularly its effectiveness – in the UK and more recently in The Netherlands highlighted the importance of maintaining a high specificity or low false-positive rate. The specificity of a single measurement depends on the cut-off percentile for referral, on the condition that a well-matched reference chart is used. However, it has been shown in registration data that the actual number of referrals increases when children are measured on more than one occasion. A systematic description of this increase is however still missing.

Finally, the publication of universal growth standards by the WHO revived a debate that started decades ago, but which was always held *in absentia* of an actual standard. Standards show optimal growth and references show actual growth. The concept behind the *universal standard*, that all children grow alike under optimal conditions, is attractive, but controversial as it refutes the paradigm of variation across populations. The validity of the universal growth standards in developed populations is currently unknown.

2. Objectives

2.1 To estimate references for length, height, weight, BMI, head circumference, and pubertal development, from a representative sample of healthy children, and up to current methodological standards (section 4.1)

Results from this analysis will be compared within the population, to estimate secular trends; and with neighbouring countries, to document population differences.

2.2 To compare growth of infants, breastfed according to WHO/MGRS feeding criteria, with the World Health Organization universal growth standards (sections 4.2 and 4.3)

With this analysis we want to validate the claim that growth of infants and young children is population-independent, and consequently, that the universal standards are appropriate for monitoring growth of all children.

2.3 To establish a reference for monitoring changes in growth over time (section 4.4)

References and criteria for normal changes in growth will be based on the concept of conditional change in SD score. The required correlation structure of length, weight and head circumference will be estimated from longitudinal data of infants and young children, and summarized with a correlation model. Results are validated, and compared with similar data for weight from the UK.

2.4 To estimate referral rates when children are measured on multiple occasions (section 4.5)

This objective includes a systematic analysis of referral patterns in longitudinal data, and a method to predict referral rates from the joint distribution of measurements.

3. Material and Methods: The Flemish Growth Study

Data were collected in a mixed longitudinal study on growth and pubertal development from birth to maturity in Flanders ("The Flemish growth study"). This study was an initiative of several stakeholders, notably, the universities of Brussels (Human Biology), and Leuven (Public Health), Well Baby Clinics (Kind en Gezin), School Health Care (CLB, Centra voor Leerlingenbegeleiding), and pediatric endocrinologists (BSGPE and UZ Brussels), and supported by the Flemish Minister of Public Health. The main aims of this study were (i) to update existing growth charts with contemporary references that allow the conversion of measurements to SD scores, (ii) to provide previously unavailable references for BMI and pubertal development; (iii) to provide references for the assessment of growth dynamics; and (iv) to analyze secular trends and determinants of growth, with special emphasis on sociodemographic differences and lifestyle indicators.

An overview of the study design, sample, and data collection is listed in table 3.1. A regional representative sample of 18050 children was recruited for the study, but the *cross-sectional reference sample* is limited to 15989 healthy children of Belgian origin (7859 boys and 8130 girls), based on the inclusion criteria listed in table 3.1. Children with a disease or on medication that could affect growth were evaluated case by case by a pediatrician (Prof. J. De Schepper). More detailed information on the number of excluded children is provided in section 4.1 (see also Roelants et al., 2008). Children were recruited by staff members of well baby clinics, daycare centres, centres for pupils guidance, a student health centre (Vrije Universiteit Brussel) and a centre for occupational medicine (IDEWE Leuven). A randomized selection of children, stratified for the factors listed in section 4.1 was ensured with a latin square design. Schoolchildren were recruited by class, preschool children were recruited individually. Due to the recruitment procedure, the actual response is unknown, but assumed to be at least 80% in schoolchildren.

The *longitudinal reference sample* includes:

a) 862 term born healthy infants (440 boys and 422 girls) of Belgian origin that were measured each month from birth to 12 months, and every three months

from 15 to 36 months of age. Children were measured at the exact target age, with a small allowance of 7 days for monthly and 14 days for quarterly measurements. These children were recruited together with the cross-sectional sample (to which they contributed a single randomly selected measurement between 0 and 12 months) but the longitudinal follow up continued until June 2006. Follow up and drop-out from the longitudinal study is discussed in section 4.4.

b) a single follow-up measurement after one (0.85 – 1.15) year in 5211 children (2581 boys and 2630 girls) 3 - 18 years of age from the reference sample.

Table 3.1: The Flemish growth study at a glance

Total number in study:	N = 18050
Recruitment:	September 2002 – June 2003, recruitment in well baby clinics and day-care centres (0 – 3 years), kindergartens and schools (3 – 18+ years), student health centre and centre for occupational medicine (18 – 21+), stratified by age, sex, region, month, am/pm, type of school/childcare (proportional to census data)
Inclusion criteria for reference sample:	Belgian origin; no known disease or medication that affects growth; term birth (37 - 42 weeks of gestation; age group 0 - 3 years only).
Cross-sectional reference sample:	N = 15989 (7859 boys and 8130 girls), 0 – 21+ years of age; including one measurement from each child in the longitudinal reference sample
Longitudinal reference sample:	(**a**) Longitudinal follow up at the exact ages of 1 – 12 months (± 7 days), and every 3 months (± 14 days) from 15 to 36 months of 862 children (440 boys and 422 girls) (12595 recorded visits; on average 14.6 per child; but 1115 visits were outside the allowed age-range); (**b**) A single follow up measurement after one (0.85 - 1.15) year in 5211 children (2581 boys and 2630 girls) 3 – 18 years of age from the reference sample (excluding 705 observations which did not belong to the reference sample or were outside time limits).
Anthropometry:	Length (≤ 2 years of age), height (≥ 2 years of age), weight, BMI (derived), head- waist- and mid-upper arm circumference. By staff members of participating centres, using standardized procedures and equipment.
Pubertal development:	Tanner staging of secondary sex characteristics (7 - 18 years of age), testicular volume with Prader orchidometer; menarche by anamnesis.
Questionnaire:	(**a**) Demographic indicators (age, sex, family composition, study program, childcare); (**b**) Social indicators (parental education and job status); (**c**) Origin of parents and grandparents; (**d**) Age, height and weight of parents; (**e**) Pregnancy and birth outcome (gestational age, size at birth; complications); (**f**) Infant feeding type (breast, formula, solids); (**g**) Health and disease (chronic disease, genetic and congenital anomalies); (**h**) Lifestyle (nutrition, physical activity, smoking and alcohol use); (**i**) Menarche and oral contraceptive use (girls).

Anthropometric measurements were taken by staff members of the participating centres. Staff members were carefully instructed on standardized measurement procedures, and all equipment was calibrated or replaced whenever this was deemed necessary. Pubertal development was assessed by a school doctor using Tanner staging of secondary sex characteristics. Illustrated reference cards with a description of each stage and a corresponding photograph (courtesy of F. Wafelbakker and J. van Wieringen) were developed for standardized assessment. Background data were collected by interview and parental questionnaire. Growth reference curves were estimated with the LMS method with penalized likelihood (Cole and Green, 1992), and validated using multiple criteria as specified in section 4.1. Minor adaptations were needed to accommodate a discontinuity in the transition from length to height (section 4.1), and to estimate annual velocity curves (data not shown). References for pubertal development were estimated from current status data using generalized additive models.

The study protocol, questionnaires and informed consent were approved by the ethics committee of the Vrije Universiteit Brussel. Written consent was obtained from a parent or legal guardian of each participant, or from the subject when they were of legal age. All documents were anonymized before transfer and further processing by the coordinating research team.

The overview presented here is brief at best, but it considers the study as a whole. Detailed information on the study design, and auxological and statistical methods of the cross-sectional study part is provided in section 4.1, while longitudinal aspects are covered in sections 4.4 and 4.5. An English version of the reference charts for growth and pubertal development can be found in the Appendix.

4. Results

4.1 References for growth and pubertal development from birth to 21 years in Flanders, Belgium. Roelants M, Hauspie R, Hoppenbrouwers K. Ann Hum Biol. 2009; 36(6): 680-94.

4.2 Breastfeeding, growth and growth standards: Performance of the WHO growth standards for monitoring growth of Belgian children. Roelants M, Hauspie R, Hoppenbrouwers K. Ann Hum Biol. 2010; 37(1): 2-9.

4.3 Growth of Belgian and Norwegian children compared to the WHO growth standards: prevalence below -2 and above +2 SD and the effect of breastfeeding. Júlíusson PB, Roelants M, Hoppenbrouwers K, Hauspie R, Bjerknes R. Arch Dis Child. 2011; 96(10): 916-21.

4.4 Conditional growth and correlations between longitudinal measurements of length, height, weight and head circumference in infancy and childhood. Roelants M, Hoppenbrouwers K., Hauspie R. *Submitted*, 2013

4.5 Referral rates for serial measurements of growth. Roelants M, Hauspie R, Hoppenbrouwers K. *Submitted*, 2013

4.1

References for growth and pubertal development from birth to 21 years in Flanders, Belgium

Roelants M, Hauspie R, Hoppenbrouwers K.

Annals of Human Biology, 36(6): 680-94, 2009

Annals of Human Biology, November–December 2009; 36(6): 680–694

ORIGINAL ARTICLE

References for growth and pubertal development from birth to 21 years in Flanders, Belgium

M. ROELANTS[1,2]*, R. HAUSPIE[1] & K. HOPPENBROUWERS[2]

[1]*Laboratory of Anthropogenetics, Vrije Universiteit Brussel (VUB), Belgium, and* [2]*Department of Youth Health Care, Katholieke Universiteit Leuven, Belgium*

(Received 27 November 2008; accepted 19 May 2009)

Abstract
Background: Due to the secular trend in length and height, growth references need to be updated regularly. Reference charts that were until recently used in Belgium are based on samples collected more than 30 years ago, and references for body mass index (BMI) and pubertal development have not been established before.
Aims: To establish contemporary cross-sectional reference charts for height, weight, BMI, head circumference, and pubertal development from birth to 21 years of age, based on a representative sample of children from Flanders, Belgium.
Subjects and methods: 15 989 healthy subjects of Belgian origin, 0–25 years of age, were measured in 2002–2004. Growth curves were fitted with the LMS method, and percentiles for the pubertal development were estimated with generalized additive models on status quo data from 8690 subjects aged 6–22 years of age.
Results: A positive secular trend in height and weight is observed in children above 5 years of age. Adult median height has increased by 1.2 cm/decade in boys and 0.8 cm/decade in girls; median weight by 0.9 kg/decade in boys, and 1.0 kg/decade in girls, and the weight distribution became more skewed. The BMI curve is comparable to that of other populations, except for higher percentiles. This reflects the increasing prevalence of overweight and obesity. Median age at menarche (13.0 years) has not advanced any more over the past 50 years. Median ages at menarche and B2 in girls and G2 or T4 in boys are comparable to other West European estimates, but approximately 10% enter G2/T4 before 9 years of age.
Conclusions: The ongoing secular trend in height and weight makes growth charts previously used in Belgium obsolete. New representative charts for growth and pubertal development are introduced. For weight monitoring, it is advised that the now-available BMI growth charts are used.

Keywords: *Auxology, growth reference charts, pubertal development*

Introduction

The secular trend in length and height has the important side effect that growth references need to be updated regularly in order to guarantee their sensitivity to detect short stature

Correspondence: M. Roelants Laboratory of Anthropogenetics, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium. E-mail: mathieu.roelants@vub.ac.be

ISSN 0301-4460 print/ISSN 1464-5033 online © 2009 Informa UK Ltd.
DOI: 10.3109/03014460903049074

and their specificity to defect overgrowth. The first data on height and weight in the Belgian population were published by Quetelet in 1831 (Quetelet,1831), and subsequent surveys showed an ongoing increase in height at all ages beyond infancy ever since (Galet 1931; Graffar and Corbier 1965; Joossens et al. 1977; Beckers and Pleysier 1983; Wachholder and Hauspie 1986; Hauspie and Wachholder 1986). Hauspie et al. estimated a secular increase in median adult stature between 1960 and 1980 of 1.2 cm/decade in Belgian males, and 1.1 cm/decade in females (Hauspie et al. 1996,1997). Reference charts that were until recently used in Belgium were based on non-representative (Joossens et al. 1977), or small (Graffar and Corbier 1965; Wachholder and Hauspie 1986) samples collected during the 1960s and 1970s. References for body mass index (BMI) and pubertal development for the Belgian population have never been established until now. The Flemish Growth Study is the first large-scale growth survey, representative for the northern part of Belgium (Flanders), and aims at characterizing growth in length and height, weight, BMI, growth velocity, head, arm and waist circumference, and pubertal development from birth to maturity, and at the study of a number of determinants of growth. In this paper, we present cross-sectional reference charts for height, weight, BMI, head circumference, and pubertal development from birth to 21 years of age.

Materials and methods

Subjects

A cross-sectional sample of 18 050 subjects, 0–25 years of age, was recruited and measured between September 2002 and June 2004. Additional information was obtained by questionnaire. Children of non-Belgian origin (n=1335), children with known growth disorders, severe chronic disease, or on medication that might affect growth (n=317), and 228 subjects with an uncertain origin or health status were excluded from the analysis. Additionally, in the age group 0–3 years, 181 children born before 37 weeks were excluded, but low birth weight was not an exclusion criterion. Belgian origin was defined as having at least one parent of Belgian origin, and the other parent of Belgian or North Western European origin (France, Luxembourg, Germany, the Netherlands, Austria, the UK, Denmark, Sweden or Norway). Based on these criteria, 15 989 subjects (7859 boys and 8130 girls) were eligible for the analysis.

Subjects were recruited in four different settings, according to age. Subjects of 0–3 years of age were recruited from September 2002 to July 2004 through a random selection of 49 Well Baby Clinics and 25 Day Care Centres accredited by Kind en Gezin (Child and Family). Subjects aged 2.5–20 years were recruited between January 2002 and June 2004 through 59 School Health Services (Centres for Pupil's Guidance). A representative sample of schoolchildren was obtained by allocating classes of pupils (the sampling unit) to centres, stratified by region, and randomized for season of the year, study level and grade through a Latin square design. The allocation was proportional to census data from the ministry of education (Vlaams Ministerie van Onderwijs 2000). Half of the children were measured before 12:00 am, the other half in the afternoon. Additional subjects aged 17–25 years were recruited among university freshmen by the Student Health Centre of the Vrije Universiteit Brussel, and among young employees who did not attend higher education, by a Centre for Occupational Medicine. Students originated from over the whole target region, but all attended the Vrije Universiteit Brussel, and employees were recruited in a single Centre for Occupational Medicine (IDEWE, Leuven).

The questionnaire contained items on demography, social status, parental height and weight, pregnancy and birth, health, and lifestyle, including nutrition and physical activity. Written informed consent was obtained from all participants or their legal guardian. The study protocol, questionnaires and informed consent were approved by the ethics committee of the Vrije Universiteit Brussel.

Measurement techniques

Biometry. Infants up to 2 years of age were measured in the supine position on a wooden Hyssna Limfog AB 4093 measuring board (0–100 cm) with fixed headrest and sliding footboard. The child was positioned in line with the centre of the board, with hips and shoulders perpendicular to the central axis, the head in the Frankfurt plane, the feet pointing upwards and the heels against the footboard. From 2 years onwards, standing height was measured barefoot in front of a wall-mounted Hyssna Limfog 4205 stadiometer, with heels, buttocks and shoulders against the wall, arms hanging loosely on the sides, and the head in the Frankfurt plane. The child was encouraged to stand straight and to breathe normally. Length and height were recorded to the last completed millimetre. Infants and toddlers up to 20 kg were weighed naked on a Tanita 1583 digital baby scale, older and heavier children were weighed in light underwear with a Seca 840 electronic scale. Weight was recorded to the nearest 0.01 kg (up to 10 kg), 0.02 kg (up to 20 kg) or to the nearest 0.1 kg otherwise. In subjects of 2 years and older, BMI was obtained by dividing weight in kilogram by the square of height in metres (kg/m^2). Head circumference was measured with a non-stretchable glass fibre tape measurer (KAWE 43971). The tape measurer was placed in a straight line from the supra-orbital position to the largest protuberance at the back of the head. Caution was taken to clear the auricles before measuring. The head circumference was recorded to the last completed millimetre.

All measurements were taken by trained nurses, assistants and physicians working in Well Baby Clinics, Day Care Centres or Health Centres, using newly installed and calibrated equipment. A standardized measuring technique was ensured by instruction sessions, illustrated cards, and a video developed for the purpose of this survey.

Pubertal development. Pubic hair (PH) in girls and boys, and genital development in boys (G) were scored visually on the Tanner scale (Tanner 1962). Breast development in girls (B) was scored both visually and by palpation to detect the budding in stage 2, and to discriminate mammary gland tissue from fat tissue. Testicular volume was measured by palpation using the Prader orchidometer with discrete volumes 1–6, 8, 10, 12, 15, 20 and 25 mL. Menarche was recorded as having occurred or still absent (status quo method). The medical staff involved in the assessment of pubertal development used instruction cards explaining the various stages of pubertal development, and illustrating these stages with photographs kindly made available by van Wieringen (Roede and van Wieringen 1985).

Statistical analysis

Length, height, weight, BMI and head circumference centiles were obtained for boys and girls separately using the LMS method (Cole and Green 1992; Cole et al. 1998). The LMS method assumes that the variable of interest (e.g. height, weight) has a Gaussian distribution after a Box–Cox power transform is applied, $y^\star = y^\lambda/(\lambda-1)$ when $\lambda \neq 0$ and $y^\star = \log(y)$, when $\lambda=0$. The whole growth curve is summarized by three smooth curves $L(t)$, $M(t)$ and $S(t)$, respectively the Box–Cox power for transformation to normality, the

median, and the coefficient of variation at time t. These curves are estimated subsequently through an iterative algorithm using penalized cubic splines, where the degree of smoothing is determined by assigning a number of equivalent degrees of freedom (edf). The whole set of centile curves can be drawn using the relation $C_{100\alpha}(t) = M(t)\,[1 + L(t)S(t)z_\alpha]^{1/L(t)}$ when $L(t) \neq 0$; and $C_{100\alpha}(t) = M(t)\exp[S(t)z_\alpha]$ when $L(t) = 0$, where z_α is the inverse probability transform of the standard normal distribution for the proportion α or the centile $p=100\alpha$. By inverting this equation, each measurement y can be expressed as a SD score (or z-score) using the relation $z=\{[y/M(t)]^{L(t)} - 1\}/S(t)L(t)$ when $L(t) \neq 0$; and $z=\ln[y/M(t)]/S(t)$ when $L(t) = 0$. In the absence of skewness, $L(t)$ is one at all ages, and the model simplifies to $C_{100\alpha}(t) = M(t) + \mathrm{SD}(t)z_\alpha$, and $z=[y - M(t)]/\mathrm{SD}(t)$, where $\mathrm{SD}(t) = M(t)S(t)$, the standard deviation at time t.

The centile curve for length and height was fitted as a discontinuous curve, explicitly including the difference between length and height when measured in the same child. Centiles were estimated after adjusting length for height, by subtracting a small constant c. In the absence of skewness, the adjustment is reversed after curve fitting, by adding the constant to the median curve, $M=M_c+c$; and recalculating the S-curve as $S=S_cM_c/(M_c+c)$, where S_c and M_c are the respective curves for length – c. Thus, no further adjustments are necessary, and the LMS curves can be readily used.

The appropriate number of equivalent degrees of freedom was selected as described by Pan and Cole (2004), based on deviance (Cole and Green 1992), Q-tests (Royston and Wright 2000) and worm plot (van Buuren and Fredriks 2001). Smoothness of the LMS curves and points of acceleration and deceleration of the median curve were checked by inspection of their derivatives. Final models were validated by comparing expected and observed proportions of observations within a set of centile bands per age class, with a chi-square goodness of fit test (grid test; Healy et al. 1988). No significant deviations from expected frequencies were observed ($p>0.1$ for all models), but the test is known to be conservative (Royston and Wright 2000). All curves were fitted and validated with a package of dedicated functions developed in R (R Development Core Team 2005). Standard errors of the L-curves were obtained with LMS Chartmaker version 2.4 (Pan and Cole 2009).

Centiles for size at birth were estimated independent from the postnatal growth curve. The LMS coefficients were obtained as the non-linear solution to $C=M\exp(SZ)$ if $|L| < 0.001$, and $C=M\,(1 + LSZ)^{\star\star[1/L]}$ otherwise, where C is a vector of ordered observations, and Z is the vector of corresponding standard normal quantiles.

Centiles of pubertal development were estimated with the status quo approach. For each stage (Tanner stages 2–5 for PH, B, G; menarche; testicular volumes 1–25 mL) presence or absence (including all previous stages) was modelled with a generalized additive model using a probit link. The amount of smoothing was determined by cross-validation (Wood 2000). As these models have no closed formula, the age percentiles were obtained numerically, with a nonlinear optimization routine in R.

Results

The number of boys and girls included in the analysis is shown in Table I. Until 19 years of age, there are generally about 300 or more subjects per year for each sex, but this number decreases thereafter. The sample is larger between birth and 1 year of age, because shortly after birth, 879 subjects were enrolled in a longitudinal part of the study that is not discussed here. From each of these infants, a single measurement between birth and

Table I. Age distribution of the cross-sectional reference sample.

	Boys	Girls
Age (years)*	n	n
0	999	950
1	340	338
2	320	317
3	328	369
4	427	366
5	366	375
6	390	399
7	390	433
8	433	392
9	418	471
10	493	475
11	471	438
12	408	438
13	381	389
14	359	369
15	293	347
16	298	320
17	244	335
18	273	365
19	119	98
20	33	30
21 +	76	116
Total	7916	8175

*Calendar age (age at last birthday).

1 year of age was selected at random for inclusion in the cross-sectional sample. Subjects up to 25 years of age were included for curve fitting, but only data from birth to 21 years of age are presented here.

Centiles for size at birth were estimated from data reported by the parents of 2772 singletons (1404 boys and 1368 girls) recruited in a Well Baby Clinic or Day Care Centre (age 0–3 years). Only children born between 37 and 42 weeks from a non-smoking mother were included in the analysis. The sample included 47.5% first-born children. Reference data for height, weight and head circumference at birth are reported in the respective tables, but they were estimated independently from the postnatal growth curve. The average gestational age was 39.5 weeks (SD 1.2 weeks), and was not significantly different according to sex (p=0.2) or birth rank (p=0.2), but first-born children had on average a 0.3 cm (0.6%; 95% CI: 0.3–0.9%) lower length (p<0.001), a 0.1 kg (3.1%; 95% CI: 2.2–4.2%) lower weight (p<0.001), and a 0.3 cm (0.8%; 95% CI: 0.5–1.2%) smaller head circumference at birth (p<0.001) compared to children of higher birth rank.

Length and height

Reference data for length and height by age in boys and girls are shown in Table II. The difference between length and height at 2 years of age was estimated in 375 children (1.75–2.25 years) who had their length as well as their height measured at the same time. The average difference was 0.96 cm (SD 1.02 cm) and was unrelated to sex (p=0.5) or age

Table II. Length and height (cm) from birth to 21 years of age in boys and girls (Flanders, 2004).

	Boys					Girls				
Age	*L**	*M*	*S*	−2 SD	+2 SD	*L**	*M*	*S*	−2 SD	+2 SD
Length										
Birth	2.0	50.7	0.040	46.4	54.6	1.9	50.0	0.040	45.9	53.9
1 month	1	55.2	0.036	51.2	59.2	1	53.9	0.034	50.2	57.6
3 months	1	61.0	0.035	56.7	65.3	1	59.6	0.034	55.6	63.6
6 months	1	67.9	0.033	63.4	72.4	1	66.4	0.033	62.0	70.7
9 months	1	72.6	0.033	67.9	77.3	1	70.9	0.033	66.2	75.6
12 months	1	76.3	0.032	71.4	81.3	1	74.7	0.033	69.8	79.7
18 months	1	82.7	0.033	77.2	88.3	1	81.4	0.034	75.9	87.0
24 months	1	88.4	0.035	82.3	94.5	1	87.3	0.035	81.2	93.5
Height										
2 years	1	87.4	0.035	81.3	93.5	1	86.3	0.036	80.2	92.5
3 years	1	96.3	0.037	89.1	103.5	1	95.3	0.038	88.1	102.6
4 years	1	103.5	0.039	95.4	111.6	1	102.4	0.040	94.2	110.5
5 years	1	110.3	0.041	101.4	119.3	1	109.5	0.041	100.4	118.6
6 years	1	117.2	0.042	107.4	127.0	1	116.4	0.042	106.5	126.2
7 years	1	123.8	0.042	113.3	134.2	1	123.0	0.043	112.4	133.6
8 years	1	129.9	0.042	118.9	140.8	1	129.0	0.044	117.8	140.2
9 years	1	135.4	0.043	123.9	146.9	1	134.3	0.044	122.4	146.3
10 years	1	140.5	0.043	128.4	152.7	1	139.9	0.046	127.1	152.7
11 years	1	145.8	0.044	132.9	158.8	1	146.6	0.047	132.8	160.4
12 years	1	152.0	0.046	137.9	166.0	1	153.2	0.046	139.1	167.4
13 years	1	158.8	0.048	143.6	174.0	1	158.7	0.043	144.9	172.4
14 years	1	166.0	0.047	150.3	181.7	1	162.4	0.040	149.5	175.4
15 years	1	171.9	0.044	156.6	187.2	1	164.7	0.038	152.3	177.1
16 years	1	175.8	0.041	161.3	190.4	1	165.8	0.036	153.8	177.9
17 years	1	178.1	0.039	164.1	192.2	1	166.2	0.036	154.2	178.2
18 years	1	179.4	0.038	165.6	193.1	1	166.3	0.036	154.3	178.2
19 years	1	180.2	0.038	166.7	193.8	1	166.4	0.036	154.5	178.3
20 years	1	180.8	0.037	167.3	194.2	1	166.5	0.036	154.7	178.4
21 years	1	181.0	0.037	167.7	194.4	1	166.6	0.036	154.7	178.4

*The standard error of *L* is 0.15 in boys and 0.17 in girls (0.11 in boys and 0.13 in girls at birth).

(p=0.2). The reference curves in boys and in girls were modelled without skewness (edf = 0, and *L* was fixed as 1), 12 edf for the median and 7 edf for the coefficient of variation, model L0(1)M12S7R. The age was rescaled (Cole et al. 1998), indicated by the letter 'R' at the end of the model term. When skewness was allowed in the model, there was limited positive skewness before and limited negative skewness after puberty. Three additional degrees of freedom for skewness decreased the deviance by only 6 (boys) or 12 units (girls) and the impact on the ±2 SD lines was generally less than 2 mm. Therefore, the more parsimonious model was chosen. The continuous growth curve is the same as the discontinuous one from birth to 18 months and again from 2.5 years onwards. At the age of 2 years, the LMS coefficients for the continuous curve correspond to the average of those for length and height. Reference values at 2.5 years of age can be obtained by interpolation.

Weight

Reference data and approximate 95% sample limits (±2 SD) for weight are presented in Table III. The curves were fitted with L5M12S7R edf in boys and L4M11S7R edf in girls,

Table III. Weight (kg) from birth to 21 years of age in boys and girls (Flanders, 2004).

	Boys					Girls				
Age	*L* (SE*)	*M*	*S*	−2 SD	+2 SD	*L* (SE*)	*M*	*S*	−2 SD	+2 SD
Birth	0.7 (0.02)	3.5	0.133	2.6	4.4	0.5 (0.02)	3.3	0.130	2.5	4.3
1 month	0.7 (0.24)	4.5	0.109	3.5	5.5	−0.1 (0.22)	4.2	0.107	3.4	5.2
3 months	0.5 (0.21)	6.1	0.108	4.9	7.5	−0.1 (0.20)	5.6	0.107	4.5	7.0
6 months	0.3 (0.18)	7.9	0.106	6.4	9.7	−0.2 (0.18)	7.3	0.106	5.9	9.0
9 months	0.1 (0.16)	9.1	0.105	7.4	11.2	−0.3 (0.16)	8.4	0.106	6.8	10.5
12 months	0.0 (0.15)	10.0	0.105	8.1	12.3	−0.3 (0.15)	9.3	0.106	7.6	11.6
18 months	−0.2 (0.14)	11.5	0.105	9.3	14.2	−0.4 (0.13)	10.8	0.108	8.8	13.6
2 years	−0.3 (0.13)	12.6	0.106	10.3	15.8	−0.4 (0.13)	12.1	0.109	9.8	15.2
3 years	−0.6 (0.12)	14.8	0.111	12.0	18.7	−0.5 (0.12)	14.3	0.115	11.5	18.2
4 years	−0.7 (0.12)	16.8	0.117	13.5	21.7	−0.6 (0.11)	16.3	0.122	13.0	21.3
5 years	−0.9 (0.12)	18.8	0.125	15.0	25.0	−0.7 (0.10)	18.5	0.132	14.5	24.8
6 years	−1.1 (0.11)	21.1	0.134	16.7	28.9	−0.8 (0.10)	20.7	0.144	16.0	28.8
7 years	−1.2 (0.11)	23.6	0.144	18.4	33.5	−0.9 (0.10)	23.4	0.157	17.8	33.9
8 years	−1.3 (0.11)	26.5	0.154	20.4	39.1	−0.9 (0.09)	26.6	0.169	19.8	39.9
9 years	−1.3 (0.10)	29.6	0.165	22.5	45.5	−0.9 (0.09)	29.8	0.180	21.8	46.2
10 years	−1.3 (0.10)	32.9	0.175	24.6	52.1	−0.9 (0.09)	33.3	0.192	23.8	53.1
11 years	−1.2 (0.10)	36.4	0.183	26.8	58.5	−0.8 (0.09)	37.5	0.203	26.3	61.1
12 years	−1.0 (0.10)	40.8	0.191	29.4	65.6	−0.7 (0.09)	42.3	0.205	29.5	68.9
13 years	−0.7 (0.11)	46.1	0.198	32.5	73.5	−0.7 (0.09)	47.1	0.194	33.5	74.4
14 years	−0.5 (0.12)	52.5	0.198	36.7	81.8	−0.9 (0.09)	51.5	0.175	37.9	78.0
15 years	−0.5 (0.12)	58.5	0.183	41.9	87.8	−1.0 (0.09)	54.9	0.162	41.5	81.2
16 years	−0.7 (0.12)	62.8	0.165	46.6	91.2	−1.2 (0.10)	57.2	0.155	43.9	84.0
17 years	−0.8 (0.13)	65.9	0.153	50.1	93.4	−1.3 (0.11)	58.5	0.153	45.2	86.0
18 years	−0.9 (0.15)	68.0	0.146	52.4	95.4	−1.3 (0.12)	59.2	0.152	45.8	87.1
19 years	−0.9 (0.18)	69.6	0.143	54.0	97.1	−1.3 (0.13)	59.6	0.151	46.2	87.8
20 years	−1.0 (0.21)	70.7	0.142	55.1	98.5	−1.4 (0.13)	59.9	0.151	46.5	88.3
21 years	−1.0 (0.23)	71.5	0.141	55.8	99.6	−1.4 (0.14)	60.0	0.151	46.6	88.5

*SE: Standard error of *L*.

using the same notation as for height. Positive skewness (*L* is less than one) was observed at all ages, so lower centiles were relatively closer to the median, while upper extreme centiles were relatively more distant from the median. The skewness was maximal just before and after puberty, while the coefficient of variation had a maximum during mid-puberty. Due to the combined effect of skewness and variability, the distance between the median and the +2 SD line was generally twice the distance between the median and the –2 SD line from 7 years of age onwards (Table III).

Body mass index

Reference data and approximate 95% sample limits (± 2 SD) for the BMI are shown in Table IV. The tabulated reference values are essentially based on height, but for curve fitting, data based on length from 1 year onwards were included to avoid edge effects. The BMI based on length was not corrected for height. The curves were fitted with model L3M9S6 in boys and model L4M9S6 in girls. Because the BMI curve is not monotonously increasing, it was not possible to use rescaled age, and other transformations of the age axis (log or power) resulted in a poor correspondence of empirical and fitted centiles. Compared to weight, the BMI showed more positive skewness, but a smaller coefficient of variation. Therefore, the impact of skewness on the ±2 SD lines was generally comparable to that of

Table IV. Body mass index (BMI, kg/m^2) from 2 to 21 years of age in boys and girls (Flanders, 2004).

	Boys					Girls				
Age (years)	*L* (SE*)	*M*	*S*	−2 SD	+2 SD	*L* (SE*)	*M*	*S*	−2 SD	+2 SD
	−1.8 (0.26)	16.4	0.073	14.4	19.4	−1.2 (0.28)	16.1	0.077	13.9	19.1
3	−1.9 (0.22)	15.9	0.075	14.0	19.0	−1.4 (0.23)	15.8	0.076	13.7	18.7
4	−2.0 (0.19)	15.6	0.078	13.7	18.8	−1.6 (0.19)	15.6	0.079	13.5	18.7
5	−2.2 (0.16)	15.5	0.082	13.4	18.9	−1.8 (0.16)	15.4	0.086	13.3	18.9
6	−2.3 (0.14)	15.4	0.088	13.3	19.3	−1.9 (0.15)	15.4	0.097	13.0	19.6
7	−2.4 (0.13)	15.5	0.096	13.2	20.0	−1.9 (0.13)	15.5	0.110	12.9	20.7
8	−2.4 (0.12)	15.8	0.104	13.3	21.1	−1.9 (0.12)	15.9	0.122	13.2	22.2
9	−2.4 (0.11)	16.1	0.113	13.5	22.5	−1.9 (0.11)	16.4	0.133	13.2	23.7
10	−2.4 (0.11)	16.6	0.122	13.7	24.0	−1.7 (0.11)	16.9	0.142	13.4	25.0
11	−2.4 (0.11)	17.0	0.129	14.0	25.5	−1.6 (0.11)	17.5	0.147	13.7	26.1
12	−2.4 (0.10)	17.6	0.133	14.3	26.7	−1.5 (0.11)	18.1	0.149	14.1	26.9
13	−2.3 (0.11)	18.2	0.135	14.7	27.8	−1.5 (0.11)	18.8	0.148	14.7	27.6
14	−2.2 (0.11)	18.8	0.135	15.3	28.5	−1.4 (0.11)	19.5	0.143	15.3	28.2
15	−2.2 (0.11)	19.5	0.132	15.8	28.9	−1.4 (0.12)	20.2	0.139	16.0	28.8
16	−2.1 (0.12)	20.1	0.128	16.4	29.1	−1.5 (0.12)	20.8	0.136	16.5	29.4
17	−2.0 (0.14)	20.6	0.124	16.9	29.1	−1.5 (0.13)	21.1	0.135	16.9	30.0
18	−2.0 (0.17)	21.0	0.120	17.3	29.2	−1.6 (0.14)	21.4	0.137	17.0	30.7
19	−1.9 (0.21)	21.4	0.118	17.6	29.2	−1.7 (0.18)	21.5	0.140	17.1	31.6
20	−1.8 (0.25)	21.7	0.117	17.8	29.4	−1.9 (0.22)	21.7	0.144	17.2	32.7
21	−1.7 (0.31)	21.9	0.116	18.0	29.5	−2.0 (0.29)	21.8	0.148	17.2	34.0

*SE: Standard error of *L*.

weight, except in adolescent boys, where the BMI is more skewed. At age 18, the adult overweight and obesity cut-offs of 25 and 30 kg/m^2 correspond to the centiles 88.8 and 98.3 in boys, and the centiles 84.5 and 97.2 in girls. On the charts, these lines were used to mark the region of overweight and obesity. Likewise, boundaries for underweight and severe underweight, were based on the adult BMI cut-offs of 18.5, 17 and 16 kg/m^2 (World Health Organization 1995; Cole et al. 2007). In the present sample, these were defined by the centiles 11.2, 1.39 and 0.13 in boys, and centiles 11.9, 2.19 and 0.36 in girls.

Head circumference

The reference curve for head circumference was fitted with L0(1)M12S3R edf in boys and L0(1)M11S4R edf in girls (Table V). The distribution is normal at each age. Since the head circumference was only measured in Well Baby Clinics, Day Care Centres and School Health Centres, reference values are only available up to 20 years of age. However, growth in head circumference has already come to a full stop at 18 years.

Pubertal development

References for pubertal development were established on the physical assessment of 8690 subjects (4219 boys and 4471 girls) aged 6–22 years (Table VI). Pubertal development was only assessed in School Health Centres. The participation rate was 95.5% in children above 7 years of age, but only 111 participants (1.3%; 59 boys and 52 girls) explicitly refused to have one or more characteristics assessed. For the assessment of testicular volume, 60 boys with unilateral cryptorchidism were additionally excluded. The age centiles were obtained for each stage separately with a non-parametric model, and they are therefore not

Table V. Head circumference (cm) from birth to 21 years of age in boys and girls (Flanders, 2004).

	Boys					Girls				
Age	*L**	*M*	*S*	−2 SD	+2 SD	*L**	*M*	*S*	−2 SD	+2 SD
Birth	−1.6	34.8	0.042	32.2	38.0	0.3	34.3	0.040	31.6	37.1
1 month	1	37.8	0.029	35.6	40.0	1	37.0	0.032	34.6	39.3
3 months	1	40.9	0.029	38.6	43.3	1	39.7	0.031	37.3	42.2
6 months	1	43.9	0.029	41.4	46.4	1	42.7	0.029	40.2	45.2
9 months	1	45.6	0.029	43.0	48.2	1	44.4	0.029	41.9	47.0
12 months	1	46.8	0.029	44.1	49.5	1	45.6	0.028	43.0	48.1
18 months	1	48.2	0.029	45.4	51.1	1	47.0	0.029	44.3	49.7
2 years	1	49.1	0.029	46.2	52.0	1	48.0	0.029	45.2	50.8
3 years	1	50.2	0.030	47.2	53.2	1	49.2	0.030	46.3	52.1
4 years	1	50.6	0.030	47.6	53.6	1	49.7	0.030	46.7	52.6
5 years	1	51.1	0.030	48.1	54.2	1	50.1	0.030	47.1	53.1
6 years	1	51.7	0.030	48.6	54.9	1	50.7	0.030	47.7	53.8
7 years	1	52.1	0.030	49.0	55.3	1	51.3	0.031	48.2	54.5
8 years	1	52.5	0.031	49.3	55.7	1	51.9	0.031	48.7	55.1
9 years	1	52.8	0.031	49.6	56.1	1	52.2	0.031	49.0	55.5
10 years	1	53.2	0.031	49.9	56.5	1	52.6	0.031	49.3	55.9
11 years	1	53.6	0.031	50.3	56.9	1	53.1	0.031	49.8	56.4
12 years	1	54.0	0.031	50.7	57.4	1	53.7	0.032	50.3	57.0
13 years	1	54.5	0.031	51.1	57.9	1	54.2	0.032	50.7	57.6
14 years	1	55.1	0.031	51.6	58.5	1	54.5	0.032	51.1	58.0
15 years	1	55.7	0.032	52.1	59.2	1	54.7	0.032	51.3	58.2
16 years	1	56.1	0.032	52.5	59.7	1	54.8	0.032	51.3	58.3
17 years	1	56.4	0.032	52.8	60.0	1	54.8	0.032	51.4	58.3
18 years	1	56.5	0.032	52.9	60.1	1	54.9	0.032	51.4	58.4
19 years	1	56.5	0.032	52.9	60.1	1	54.9	0.032	51.5	58.4
20 years	1	56.5	0.032	52.9	60.1	1	54.9	0.032	51.5	58.4

*The standard error of *L* is 0.29 in boys and 0.29 in girls (0.16 in boys and 0.15 in girls at birth).

necessarily symmetrically distributed around the median. The second and third stage of pubic hair in boys and girls, genital development in boys and breast development in girls, and are generally negatively skewed (upper centiles are relatively closer to the median), while stages 4 and 5 of these characteristics are positively skewed (upper centiles more distant from the median). With probit analysis, the mean±SD age at menarche was estimated at 13.13±1.3 years, which is slightly different from the non-parametric values shown in Table VI.

Discussion

In this paper we present contemporary cross-sectional reference charts for length/height, weight, BMI, head circumference and pubertal development for the Belgian population, based on a representative sample from the region of Flanders (the Dutch speaking northern part of Belgium). These charts are an update of older charts (Graffar and Corbier 1965; Joossens et al. 1977; Wachholder and Hauspie 1986), and include reference material previously unavailable (BMI, pubertal development). The sample is representative for children of Flemish origin, as a correct social and geographical representation of the sample was enforced by applying a rigorous sampling scheme. Moreover, possible seasonal bias was avoided through recruitment and measuring of different age groups spread over the year.

Table VI. Pubertal development from 6 to 21 years of age in boys and girls (Flanders, 2004).

	Age at percentile (years)						
Parameter (stage)*	P3	P10	P25	P50	P75	P90	P97
Boys							
Pubic hair	8.1	9.9	11.0	11.9	11.9	13.5	14.3
P2	10.9	11.6	12.3	13.0	13.0	14.5	15.2
P3	11.9	12.5	13.2	13.9	13.9	15.4	16.1
P4	12.8	13.5	14.3	15.2	15.2	17.6	(21.7)†
P5							
Genital development							
G2	(6.5)†	8.9	10.4	11.4	12.3	13.0	13.7
G3	10.7	11.5	12.2	12.9	13.7	14.4	15.1
G4	11.8	12.5	13.1	13.9	14.6	15.4	16.2
G5	12.7	13.5	14.3	15.3	16.7	18.7	(21.8)†
Testicular volume (mL)							
T4	7.6	9.2	10.4	11.4	12.2	13.0	13.7
T6	10.2	11.0	11.6	12.3	13.1	13.8	14.6
T8	10.8	11.4	12.0	12.8	13.5	14.2	14.9
T10	11.2	11.8	12.5	13.2	13.9	14.7	15.6
T12	11.4	12.1	12.7	13.5	14.3	15.3	17.1
T15	11.8	12.4	13.1	13.9	14.9	16.5	(20.6)†
T20	12.3	13.0	13.8	14.9	16.9	NA	NA
Girls							
Pubic hair	8.3	9.2	10.1	11.0	11.8	12.4	13.1
P2	9.7	10.5	11.2	11.9	12.5	13.1	13.7
P3	10.7	11.3	12	12.7	13.6	14.4	15.4
P4	11.6	12.4	13.2	14.5	16.7	(20.9)†	NA
P5							
Breast development							
B2	7.7	8.8	9.8	10.7	11.5	12.2	12.9
B3	9.5	10.3	11	11.8	12.6	13.3	14.0
B4	10.7	11.3	12	12.8	13.8	14.8	16.0
B5	11.5	12.4	13.4	14.8	17.1	(20.4)†	NA
Menarche							
M	10.9	11.6	12.3	13.0	13.9	14.7	15.7

*Tanner stages for pubic hair (P), genital (G) and breast development (B); discrete testicular volumes in mL (T), and presence of menarche (M).
†Estimates below 7 and above 19 years of age should be treated with caution as data on pubertal development are scarce in these age groups; NA, not available, stage not reached by the corresponding proportion of subjects by 22 years of age.

There is little variation in the number of measured children with age, except during the first year of life, where children were oversampled by design, and above 19 years of age, where the number of observations decreases rapidly. However, the growth curve of monotonously increasing traits like height, weight, and head circumference, was fitted with the age axis compressed to the reciprocal of the growth velocity (rescaled age; Cole et al. 1998), such that the relative density of subjects increases above 18 years of age, where the growth velocity is almost zero.

The growth curves were estimated with the LMS method using penalized splines (Cole and Green 1992). This is an elegant way to summarize a growth curve, and several criteria to check the validity of the estimates exist (Royston and Wright 2000; van Buuren and

Fredriks 2001; Pan and Cole 2004). This approach is also more objective than methods used in the past, and holds a number of important advantages. Each observation can be converted to its standard deviation score (SDS or *z*-score), which can be used to test the normality assumption, and, if the data are normally distributed after transformation, more extreme centiles can be estimated with reasonable accuracy.

Centiles for size at birth were based on reported rather than measured data, and were therefore estimated independently from the post-natal growth curve. By limiting these reported values to the children recruited in Well Baby Clinics and Day Care Centres, we are however confident that reported values are accurate as they can be transcribed from the health booklet or from the maternity discharge form. When split by week of gestational age, there is a good correspondence with reference values for birth weight based on the birth register in Flanders (Devlieger et al. 2000).

Height

Adult median stature (± SD) in the current sample, arbitrarily taken at 21 years of age, is 181.0±6.7 cm in boys and 166.6±6.0 cm in girls. There is no indication that growth comes to a full stop by that age, but the increase in height after the age of 21 years is in the present sample less than 1 mm/year. Compared to the three most recent Belgian growth studies (Joossens et al. 1977; Beckers and Pleysier 1983; Wachholder and Hauspie 1986), the secular increase in height at 18 years of age was about 1.2 cm/decade in boys and 0.8 cm/decade in girls, and the difference between young adult boys and girls has increased from 13 cm to more than 14 cm. Hauspie et al. (1996) estimated a comparable secular trend in adult height between 1960 and 1980 of 1.2 cm/decade in boys, and a slightly higher trend of 1.1 cm/decade in girls. The secular increase is present in all age groups for both sexes, but smaller before the age of 5, and, in boys, larger around the age of peak height velocity. This could indicate a secular trend towards an earlier growth spurt in boys. The same trends are observed in the 3rd and 97th centiles. Comparable trends have been observed in other recent Western European growth studies (Hauspie et al. 1997; Cole et al. 1998; Fredriks et al. 2000b). For this reason, population differences that were seen some 30 years ago, still exist. Compared to the original Tanner–Whitehouse reference (Tanner et al. 1966), Flemish boys are on average 2–5 cm taller, and girls 2–4 cm taller than their British counterparts of the 1950s and 1960s. When comparing height of Flemish children with the more recent UK 1990 growth charts (Cole et al. 1998), we notice that British children are still 1–4 cm smaller than Flemish boys and girls. On the contrary, Dutch children measured in 1996/97 are 2–4 cm taller than their Flemish counterparts. However, because of a north to south gradient observed in the Netherlands, Dutch children living closer to Flanders are on average below their national growth curve (Fredriks et al. 2000b). Children living in the northern provinces of Flanders (Antwerp, Limburg) are also taller than average, but this difference is less than 0.1 SD (data not shown).

Weight

The actual median weight of 71.1 kg in boys and 60.1 kg in girls at the age of 21 years, shows that there has also been a considerable increase in body mass over the past 30–40 years. Given the large skewness that is observed in the present weight charts, the standard deviation is not an appropriate measure of the population variation, but approximate 95% confidence limits of the sample can be obtained from the ±2 SD columns in Table III. By

comparing the present weight charts with those published before (Joossens et al. 1977; Beckers and Pleysier 1983; Wachholder and Hauspie 1986), we estimate a secular trend of about 0.9 kg per decade in boys and 1.0 kg per decade in girls. Besides an increase in median weight over time, we also noticed that the distribution of body weight became increasingly skewed. To document this increasing skewness, we compared the ratio of the distance of the 97th percentile to the median, and third percentile to the median. This ratio is about one at all ages in the weight curve of Wachholder and Hauspie (1986) and about 1.5 in the weight curve of Joossens et al. (1977). In the current sample, this ratio is about 2 in adolescents and young adults, but smaller in preadolescent children. Teenagers on the 3rd percentile or the median are nowadays about 2–4 kg heavier than 30 years ago, but for children for the 97th percentile this difference is as much as 10–15 kg.

Body mass index

Weight-for-age is no longer considered to be a suitable indicator for body mass in older children, and has been gradually replaced by growth charts for the BMI (Cole et al. 2000). The BMI values presented here are the first reference curves for the Belgian population. Unlike charts from the UK (Cole et al. 1995) or the Netherlands (Fredriks et al. 2000a), the curves were fitted on the original age scale for reasons explained in the Results section. The 3rd centile and the median are within ±0.5 BMI units of the corresponding centiles from the UK, the Netherlands, and the WHO (de Onis et al. 2007), but from age 7 onwards, the 97th centile is up to 3 units higher. As the BMI charts are a snapshot of the current distribution of BMIs in Flemish children, this probably reflects the sampling period (2002–2004) rather than population differences. Using the criteria set by the International Obesity Task Force (Cole et al. 2000), the overall prevalence of overweight and obesity is 14.2% and 3.0% in girls, and 11.8% and 2.6% in boys 2–18 years of age. In the same age group, the prevalence of a BMI below the threshold for underweight or thinness grade I, II and III (Cole et al. 2007) is 11.0%, 2.0% and 0.5% in girls and 11.2%, 1.5% and 0.2% in boys. These numbers are broadly comparable or even below recent estimates from other West European countries. Since 'normal limits' for the BMI cannot be arbitrarily set by a statistical criterion like ±2 SD or ±2.5 SD, we adopted the same approach as used by the IOTF (Cole et al. 2000) and Cole et al. (2007) to set objective cut-offs for overweight and underweight or thinness.

Pubertal development

The references for pubertal development presented here are the first for the Belgian population, except for menarche, which has been recorded on a regular basis since the 1920s (Vercauteren and Susanne 1985; Hauspie et al. 1997). The average age at menarche has gradually decreased during the first half of the 20th century, but has remained virtually unchanged since the 1950s. The lack of historical data on other traits of pubertal development does not allow for a comparable analysis in boys. Median ages at B2 and at menarche in girls, and at G2 or T4 in boys are similar to those reported for other North-West European countries (Parent et al. 2003). Age centiles for the pubertal development were estimated with a generalized additive model rather than classical probit analysis. These non-parametric models are more faithful to the data, but have important implications for the estimated age percentiles, since the age distribution is often skewed and possibly leptokurtic. In other studies, skewness has been attributed to large differences in living

conditions, or the presence of immigrant children (for review, see Parent et al. 2003), but the present sample is relatively homogeneous, and only includes children of Belgian or mixed Belgian/West European origin. The age limits for precocious puberty correspond to approximately the third percentile in girls (B2 before 8 years of age; Marshall and Tanner 1969), and to approximately the 10th percentile in boys (G2 or T4 before 9 years of age; Marshall and Tanner 1970). A similar observation was made in the Netherlands with a comparable modus operandi (Mul et al. 2001).

The validity of the Flemish growth charts for the southern part of Belgium (Walloon region) is unknown. However, in the past, both regions were always reasonable close together, although differences in culture and lifestyle exist. In the most recent survey that included both regions (Beckers and Pleysier 1983), Flemish schoolchildren were about 1 cm taller than their Walloon counterparts.

Conclusions

The ongoing secular trend in height and weight makes growth charts that were previously used in Belgium obsolete. The current charts are based on a representative sample of Flemish children without known disorders that affect growth, and should be considered as a growth reference and not a growth standard, since they show current growth rather than optimal growth. For weight monitoring, it is advised that the now-available BMI growth charts are used. The BMI accounts for height and age, and allows objective criteria for normal weight, overweight and underweight to be set. The references for pubertal development will help the interpretation of Tanner stages assessed in children, and provide a baseline for bio-monitoring. Because of the cross-sectional nature of the data, these references give, however, no indication about the normal duration of puberty.

The reference charts are available in Dutch, English and French. They can be downloaded free of charge, along with other material from the project from the web site www.vub.ac.be/groeicurven. See the web site disclaimer section for details. The numerical data in Tables II, III, IV and V allow calculation of centiles and SD scores with reasonable accuracy. More detailed data can be obtained from the authors.

Acknowledgements

We want to express our gratitude to the more than 18 000 children and their parents who participated in the survey; the Well Baby Clinics, Day Care Centres and School Health Services for their sound work on recruitment and measurement of the children; Prof. Dr Guido Moens of the Katholieke Universiteit Leuven and IDEWE Leuven for the additional data on young employees; Inneke Demaeseleer for her work on the university freshmen; Sara Daman for her logistic and administrative assistance, calibration of the equipment, and many visits to participating centres; Prof. Jean De Schepper and Dr André Moreels of the Vrije Universiteit Brussel, Dr Johan Pauwels and Dr Martine Debyser of Kind en Gezin, and Dr Lut Dooms for their assistance and advice. Professor Dr Em. Hans van Wieringen provided photographs illustrating the various stages of pubertal development, and Dr Huiqi Pan gave access to a developmental version of LMS Chartmaker 2.4. Pfizer Belgium kindly provided logistical support. This project was supported by a grant from the Minister of Health of the Flemish Government.

***Declaration of interest*:** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Beckers R, Pleysier R. 1983. Anthropometric analysis of the 1980–1981 Belgian school population. I. Description of the survey project; II. Initial results. Arch Belg 41:89–123.

Cole T, Bellizzi M, Flegal K, Dietz W. 2000. Establishing a standard definition for child overweight and obesity worldwide: International survey. BMJ 320:1240–1243.

Cole T, Flegal K, Nicholls D, Jackson A. 2007. Body mass index cut offs to define thinness in children and adolescents: International survey. BMJ 335:194–197.

Cole T, Freeman J, Preece M. 1998. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. Stat Med 17:407–429.

Cole T, Green P. 1992. Smoothing reference centile curves: The LMS method and penalized likelihood. Stat Med 11:1305–1319.

Cole TJ, Freeman JV, Preece MA. 1995. Body mass index reference curves for the UK, 1990. Arch Dis Child 73:25–29.

de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. 2007. Development of a WHO growth reference for school-aged children and adolescents. Bull World Health Org 85:660–667.

Devlieger H, Martens G, Bekaert A, Eeckels R. 2000. Standaarden van geboortegewicht-voor-zwangerschapsduur voor de Vlaamse boreling. Tijdschr Gen 56:1–14.

Fredriks A, van Buuren S, Wit JM, Verloove-Vanhorick P. 2000a. Body index measurements in 1996–7 compared with 1980. Arch Dis Child 82:107–112.

Fredriks AM, van Buuren S, Burgmeijer R, Meulmeester J, Beuker R, Brugman E, Roede M, Verloove-Vanhorick S, Wit JM. 2000b. Continuing positive secular growth change in The Netherlands 1955–1997. Pediatr Res 47:316–323.

Galet O. 1931. Essai de détermination de quelques moyennes staturales et pondérales correspondant à diverses étappes de développement infantile. Analyse sommaire de l'influence de certaines facteurs sociaux sur l'évolution de la croissance. Bull Soc Royale Belge Anth 46:139–316.

Graffar M, Corbier J. 1965. Comparaison d'enfants belges normaux à des critères adéquates de poids et de la taille. Acta Paediatr Belg 19:235–274.

Hauspie RC., Wachholder A. 1986. Clinical standards for growth velocity in height of Belgian boys and girls, aged 2 to 18 years. Int J Anthropol 1:339–348.

Hauspie R., Vercauteren M., Susanne C. 1996. Secular changes in growth. Horm Res 45:Suppl 28–17.

Hauspie RC, Vercauteren M, Susanne C. 1997. Secular changes in growth and maturation: an update. Acta Paediatr 423:Suppl20–27.

Healy M, Rasbash J, Yang M. 1988. Distribution-free estimation of age-related centiles. Ann Hum Biol 15:17–22.

Joossens JV, Brems-Heyns E, Bande-Knops J, Eggermont E. 1977. Longitudinale–transversale kurven opgesteld volgens 97000 metingen bij meer dan 25000 kinderen gevolgd te Leuven van 1963 tot 1977, gecorrigeerd voor tijdtrend op 1977. Leuven. .

Marshall WA, Tanner JM. 1969. Variations in pattern of pubertal changes in girls. Arch Dis Child 44:291–303.

Marshall WA, Tanner JM. 1970. Variations in the pattern of pubertal changes in boys. Arch Dis Child 45:13–23.

Mul D, Fredriks AM, van Buuren S, Oostdijk W, Verloove-Vanhorick SP, Wit JM. 2001. Pubertal development in The Netherlands 1965–1997. Pediatr Res 50:479–486.

Pan H, Cole TJ. 2004. A comparison of goodness of fit tests for age-related reference ranges. Stat Med 23:1749–1765.

Pan H, Cole TJ. 2009. LMS Chartmaker, A program for calculating age-related reference centiles using the LMS method, Version 2.4. London: Medical Research Council.

Parent AS, Teilmann G, Juul A, Skakkebaek NE, Toppari J, Bourguignon JP. 2003. The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. Endocr Rev 24:668–693.

Quetelet A. 1831. Recherches sur les lois de la croissance de l'homme. Mémoires de l. Academie de Bruxelles 7:1–27.

R Development Core Team 2005. R: A Language and Environment for Statistical Computing. Vienna : R Foundation for Statistical Computing. .

Roede M, van Wieringen H. 1985. Growth diagrams 1980: Netherlands third nation-wide survey. Tijdschr Soc Gezondheidsz 63:Suppl1–34.

Royston P, Wright EM. 2000. Goodness-of-fit statistics for age-specific reference intervals. Stat Med 19:2943–2962.

Tanner J, Whitehouse R, Takaishi M. 1966. Standards from birth to maturity for height, weight, height velocity and weight velocity: British children, part II. Arch Dis Child 41:613–635.

Tanner J. 1962. Growth at adolescence2nd edn (pp. –). Oxford: Blackwell Scientific.

van Buuren S, Fredriks M. 2001. Worm plot: A simple diagnostic device for modelling growth reference curves. Stat Med 20:1259–1277.

Vercauteren M., Susanne C. 1985. The secular trend of height and menarche in Belgium: Are there any signs of a future stop?. Eur J Pediatr 144:306–309.

Vlaams Ministerie van Onderwijs 2000. Statistisch Jaarboek van het Vlaams Onderwijs 1999–2000. Brussels : Ministerie van de Vlaamse Gemeenschap. .

Wachholder A., Hauspie RC. 1986. Clinical standards for growth in height of Belgian boys and girls, aged 2 to 18 years. Int J Anthropol 1:327–338.

Wood SN. 2000. Modelling and smoothing parameter estimation with multiple quadratic penalties. J R Stat Soc B 62:413–428.

World Health Organization. 1995. Physical status: the use and interpretation of anthropometry. Report of a WHO expert committee. World Health Organ Tech Rep Ser 854:1–452.

This paper was first published online on Early Online on 21 October 2009

NOTE:

The footnote below table II should read: "*The standard error of L is 0.11 in boys and 0.13 in girls at birth; for postnatal growth L was fixed as 1."

The footnote below table V should read: "*The standard errors of L is 0.16 in boys and 0.15 in girls at birth; for postnatal growth L was fixed as 1."

4.2

Breastfeeding, growth and growth standards: Performance of the WHO growth standards for monitoring growth of Belgian children

Roelants M, Hauspie R, Hoppenbrouwers K.

Annals of Human Biology, 37(1): 2–9, 2010

Annals of Human Biology, January–February 2010; 37(1): 2–9

ORIGINAL ARTICLE

Breastfeeding, growth and growth standards: Performance of the WHO growth standards for monitoring growth of Belgian children

MATHIEU ROELANTS[1,2], ROLAND HAUSPIE[1] & KAREL HOPPENBROUWERS[2]

[1]*Laboratory for Anthropogenetics, Vrije Universiteit Brussel, Belgium,* [2]*Centre for Youth Health Care, Department of Public Health, Katholieke Universiteit Leuven, Leuven, Belgium*

(Received 12 November 2008; accepted 1 June 2009)

Abstract

Background: In 2006, the World Health Organization published universal growth standards for all children from birth to 5 years of age, based on a sample of breastfed children.

Aims: The present study documented breastfeeding prevalence in the Flemish Growth Survey 2004, and compared growth of exclusively breastfed children in Flanders with local reference charts and WHO growth standards.

Subjects and methods: A subset of 3287 children 0–3 years of age from the Flemish reference population was studied. Prevalence and duration of breastfeeding were estimated with the status quo method. SD scores (SDS) of length/height, weight, BMI, and head circumference were plotted by age.

Results: Breastfeeding is initiated for 68.2% of children, and approximately 25% were exclusively breastfed until at least 6 months of age. Breastfed children grow according to a typical pattern that deviates from the local reference curve. The average length of breastfed children is reasonably close to the WHO growth standard, but this does not hold for weight, BMI, or head circumference. In Flanders, breastfed children are more comparable to the local reference than to the WHO growth standards.

Conclusions: Growth of breastfed children is similar to the WHO standards for length, but not for other traits. In Flanders, the use of the recent local growth reference is advised for both breastfed and formula-fed children.

Keywords: *Auxology, breastfeeding, infant growth, growth standards*

Introduction

Formula-fed children have long been known to grow according to a different pattern compared to breastfed children (Dewey et al. 1995; Cole et al. 2002). Growth reference

Correspondence: Mathieu Roelants, Laboratory of Anthropogenetics, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium. E-mail: mathieu.roelants@vub.ac.be

ISSN 0301-4460 print/ISSN 1464-5033 online © 2009 Informa UK Ltd.
DOI: 10.3109/03014460903089500

charts are typically based on a representative sample of children from the target population, and therefore include both formula and breast fed infants. This is a major shortcoming, since unaware parents could be encouraged to give up breastfeeding because their child apparently performs badly on the growth chart. In 2006 the World Health Organization (WHO) published universal growth standards for children from birth to 5 years of age all over the world, regardless their ethnicity, socio-economic background, or type of feeding. These charts are the result of the Multicentre Growth Reference Study (MGRS) conducted by the WHO, and are based on a sample of children that were exclusively or predominantly breastfed for at least 4 months (WHO Multicentre Growth Reference Study Group 2006b). The idea behind this initiative was that a number of reports showed that young children grow alike under optimal environmental conditions (Garza and de Onis 2004). This is controversial, since population differences in growth are well described and attributed not only to environmental conditions, but also among others to climate and genetic background (Eveleth and Tanner 1990).

In 2004, new growth references were introduced in Flanders, Belgium. As usual, these references are based on a representative mixture of breast- and formula-fed infants, but the prevalence of breastfeeding is relatively low (Roelants et al. in press). In this paper we compare growth of Belgian children exclusively breastfed for at least 6 months, to the local growth reference, and to the WHO standards.

Materials and methods

The Flanders Growth Survey 2004 is a growth study based on 18 025 subjects aged 0–25 years, representative for the region of Flanders, Belgium. The present analysis is limited to 3287 infants and toddlers (1668 boys and 1619 girls) between 0 and 3.15 years of age who were recruited through Well Baby Clinics or day care centres. The sample includes only healthy children of Belgian origin, born at term (37–42 weeks of gestational age). Low birth weight children born at term were not excluded. Length, weight and head circumference were measured by trained staff members, using a standardized technique and equipment. Weight, length, and head circumference at birth were obtained by questionnaire. Growth reference curves from birth to young adulthood were estimated with the LMS method (Cole and Green 1992) as described elsewhere (Roelants et al. in press).

The socio-demographic profile and feeding habits were recorded through a parental questionnaire. For the present analysis, social stratification is based on the educational level of the mother (the highest attained degree). A low educational level includes mothers who did not complete secondary education (no education, primary, or only lower secondary); the medium level includes mothers who completed secondary education, typically up to 18 years of age; and a high level includes all mothers who achieved a degree in higher education. To document feeding practice during infancy, parents were asked until what age their child received breast feeding, from what age their child was on formula feeding, and from what age their child received solid foods. Children were considered as exclusively breastfed when they were breastfed for at least 6 months, and no formula milk was introduced before that age. This includes all children younger then 6 months who did not yet receive any formula milk. Solids were not accounted for since they are usually introduced between 4 and 6 months of age, and the number of exclusively breastfed children would have become too low. The reference group consists of all children on which the Flemish growth charts are based, including the group of exclusively breastfed children by definition.

Statistical analysis

Description of the prevalence and average duration of breastfeeding in the reference group is based on the current feeding practice, with the status-quo approach. The average duration of breastfeeding was estimated with log-binomial regression, and the result is expressed as the half-life, or the age where the prevalence is half of the prevalence at birth. The relation between breastfeeding and the socio-economic status (SES) was assessed by logistic regression, and results are expressed as odds ratios (OR). Half-lives and OR are reported with their 95% confidence intervals.

Length (≤2 years of age), height (>2 years of age), weight, BMI and head circumference were converted to SD scores (SDS) according to the Flemish growth reference, and according the WHO growth standards (WHO Multicentre Growth Reference Study Group 2006c, 2007). Subjects were grouped by age in 3-month intervals from 0 to 12 months, and half-year intervals thereafter (3.0–3.15 years of age for the oldest age group). Size at birth was based on recall data from all subjects in the sample, and constitutes a separate age class. Observations from both sexes were combined to obtain sufficiently large numbers. Mean SDS values are presented with their standard errors (±1 SEM) in graphs. Differences among groups were tested with ANOVA when appropriate.

For proportions, 95% confidence intervals were obtained through the Wilson Score method (Newcombe 1998). All analysis was done in R version 2.7.2 (R Development Core Team 2008).

This analysis is part of the Flemish Growth Survey 2004 for which the study protocol, procedures and questionnaires were approved by the ethics committee of the Vrije Universiteit Brussel.

Results

Current infant feeding practice in the reference group is documented in Table I. In the complete reference group, the prevalence of any breastfeeding at birth is 68.2% (66.6–69.8%), but shortly after birth this drops to approximately 58%. Once initiated, the estimated half-life of any breastfeeding is 5.3 months (4.8–5.9 months). The prevalence of exclusive breastfeeding at birth is 66.0% (64.3–67.6%), and the half-life is about 3.1 months (2.5–3.7 months). By the end of the first month, the prevalence has dropped by 20%, and from that point onwards it is generally 10–15% lower compared to any breastfeeding. Compared to the medium educated mothers, the odds for any breastfeeding were significantly lower for children born to a mother with a lower education (OR 0.50; 0.37–0.66), and higher for children born to a higher educated mother (OR 2.79; 2.38–3.29). For exclusive breastfeeding, the odds were not significantly different when the mother had a lower education (OR 0.88; 0.60–1.28), but still higher (OR 1.87; 1.56–2.26) in case of a higher education.

Before the age of 6 months, children were assigned to the exclusive breastfed group based on current feeding practice as shown in Table I. In this age group, current exclusive breastfeeding was recorded for 386 out of 1125 children (34.3%; 31.6–37.2%). From 6 months of age onwards, exclusive breastfeeding was determined by means of recall data. In the age group 6–12 months, exclusive breastfeeding was recorded for 113 out of 529 children (21.4%; 18.0–25.1%); in the age group 1–2 years, for 101 out of 669 children (15.1%; 12.5–18.1%), and above 2 years of age for 93 out of 648 children (14.4%; 11.8–17.3%).

Table I. Prevalence of current breastfeeding according to age.

Age group (calendar month)	Number of children	Any breastfeeding (%)	Exclusive breastfeeding (%)
0	178	58.2	47.6
1	230	50.4	37.4
2	159	43.7	35.3
3	176	42.0	30.8
4	232	42.7	29.2
5	212	39.5	24.0
6	189	37.4	26.5
7	128	22.0	12.9
≥8	1783	5.3	4.3

In Figure 1, average SDS for length/height, weight, BMI, and head circumference, based on the Flemish reference are plotted by age group. Birth length is 0.10 SD higher for exclusively breastfed children ($p = 0.02$), after correction for a high (+0.12 SD; $p = 0.04$) or low ($p = 0.5$) education of the mother (compared to a medium education), and smoking during pregnancy (±0.16 SD; $p = 0.04$). Shortly after birth, there is an increase in average length SDS, but between 4 and 6 months of age, there is a slowdown, and between 6 months and about 2.5 years, breastfed children are on average smaller than the reference population. A similar pattern can be observed for weight, except that average weights of breastfed children are already comparable to the reference population from 1.5 year onwards. Corrected for smoking during pregnancy, which has a strong negative effect on birth weight (±0.40 SD; $p < 0.001$), children in the low SES group have on average a 0.14 SD lower birth weight compared to a medium SES ($p = 0.07$), but this increases again with 0.44 SD if they are exclusively breastfed ($p = 0.02$). Children in the high SES group have on average a 0.08 SD higher birth weight ($p = 0.08$), and this increases with 0.26 SD if they are exclusively breastfed ($p = 0.01$). The BMI of exclusively breastfed children is on average 0.1–0.2 SD below the median before the age of 1.5 years, and comparable thereafter. Head circumference shows a similar trend as length during the first 6 months of age, and fluctuates around the median thereafter.

Figure 2 shows WHO-based SDS of the same traits for exclusively breastfed children and for the total Flemish reference sample. Birth length is about 0.5 SD above the WHO median in both groups. From 6 months of age onwards, the average length SDS of the reference group is systematically about 0.25 SD higher then the WHO median. The average length SDS of exclusively breastfed children gradually drops between birth and 6 months of age. Thereafter, the pattern is similar to that of the SDS based on the Flemish reference in Figure 1, but closer to the median. Weight SDS is about 0.25 SD at birth and drops to ±0.25 SD by 4 months of age in both groups. After that age, exclusively breastfed children follow the same trend as the reference group, but their average SDS is 0.1–0.2 SD lower, except between 1 and 1.5 years of age, where the difference is 0.5 SD. BMI SDS trends are also comparable in both groups, with slightly lower values for exclusively breastfed children. For head circumference, the average SDS of both groups fluctuates around 0.5 SD.

Overall, a comparison of Figure 1 and Figure 2 shows that for length and head circumference, the pattern of SDS by age of exclusively breastfed children is similar when using either the local growth reference or the WHO growth standards as a reference. For weight and BMI this is not conclusive. More striking is that despite differences in growth, the pattern of SDS by age of exclusively breastfed children is for all traits except length, similar to that of the total reference group, when compared to the WHO growth standards.

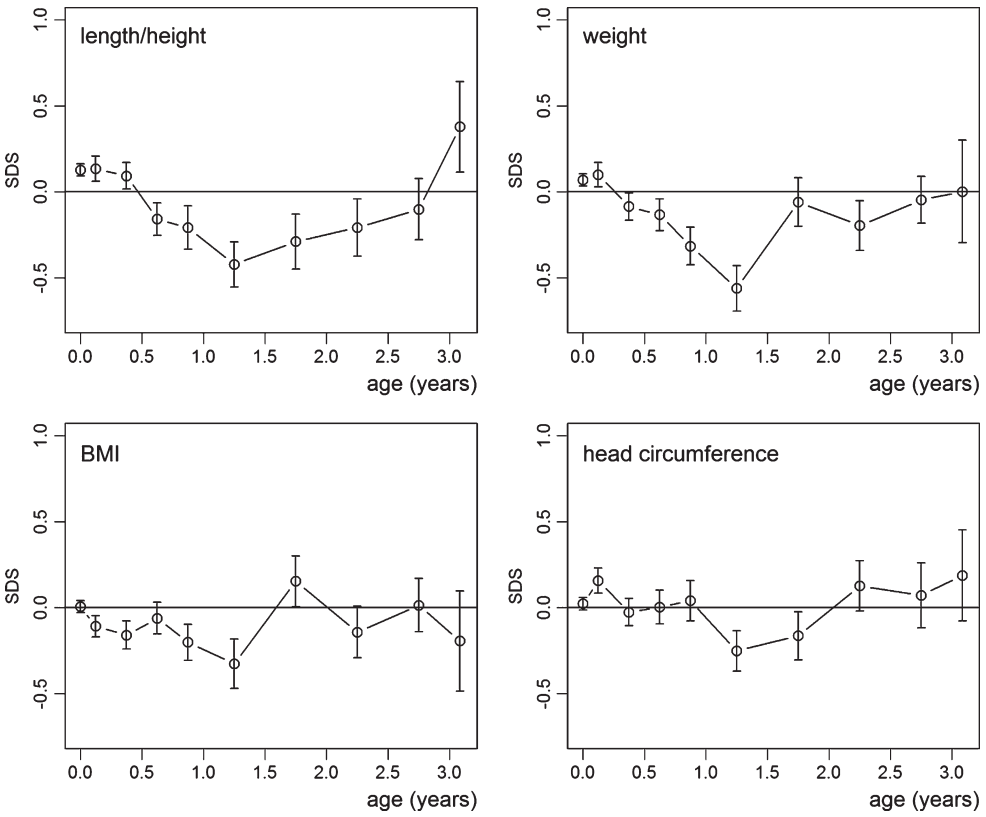


Figure 1. Growth in length/height, weight, BMI, and head circumference of exclusively breastfed children in Belgium: Mean SD score (±SEM) by age group, calculated from the Belgian reference charts, boys and girls combined.

Discussion

The prevalence and duration of breastfeeding in this representative sample of children taken from the Flemish Growth Survey 2004 (Roelants et al. in press), were documented using status quo data, which is highly reliable since it does not depend on parental recall. Only breast milk and formula milk were considered, as solids were usually introduced between 4 and 6 months of age, and the number of children in the exclusively breastfed group would have become too small. The prevalence of breastfeeding reported here is comparable to routinely collected data from the Well Baby Clinics in Flanders (Kind en Gezin 2008). As for the average duration, this is to our knowledge the only recent report based on a representative sample of Flemish infants. Older children were assigned to the exclusively breastfed group using strict criteria (no formula milk before 6 months of age), but the prevalence in these age groups is only half of the expected amount. This is probably due to recall bias, but it might also reflect a change in policy, since the prevalence of breastfeeding in Belgium has almost doubled over the past decade (Yngve and Sjöström 2001). This prevalence remains however among the lowest in Europe (Yngve and Sjöström 2001; Kind en Gezin 2008).

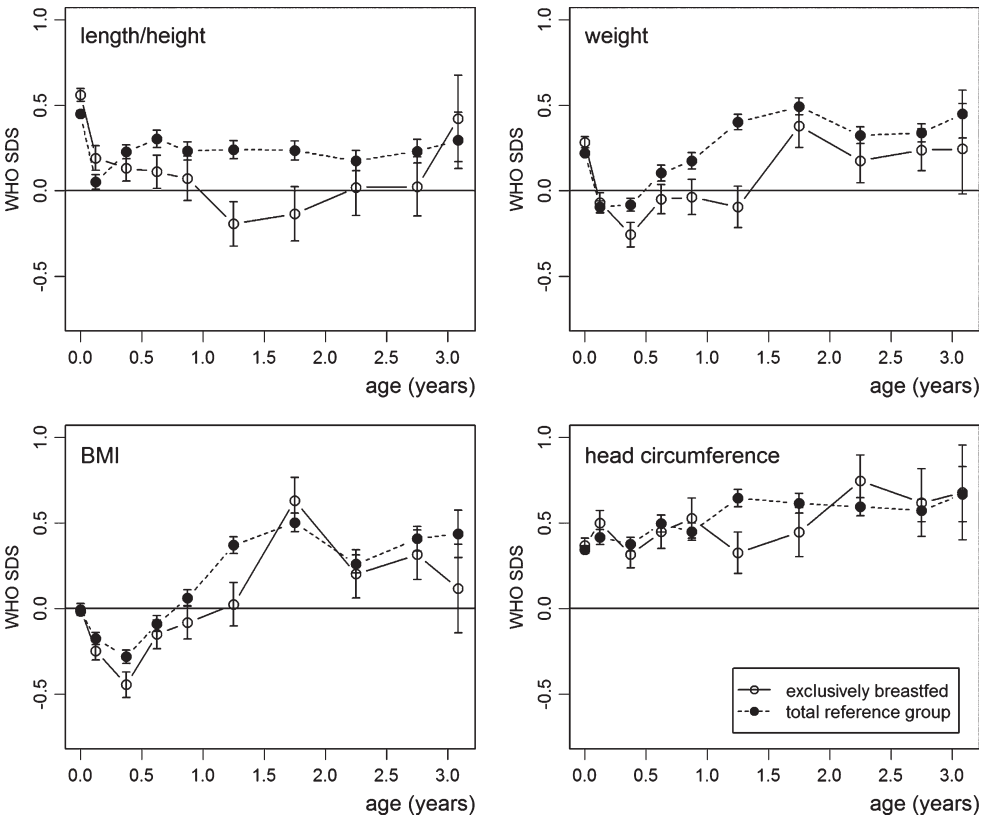


Figure 2. Comparison of average SD scores (±SEM) of length/height, weight, BMI, and head circumference by age group, in all children (dashed line) and exclusively breastfed children (full line) according to the WHO growth standards, boys and girls combined.

Comparison of growth in length and weight of exclusively breastfed children to the local reference confirms the typical pattern that has been observed before (Dewey et al. 1995; Cole et al. 2002). Growth in BMI is comparable to weight, and growth in head circumference comparable to length, but the effect of breastfeeding on these traits is smaller. Compared to the WHO standards, the overall picture is that despite these differences, growth of exclusively breastfed children resembles more that of the Flemish reference group, than that of the MGRS. A notable exception is postnatal growth in length, where Flemish breastfed children are remarkably similar to the WHO standard, while the complete reference group is consistently taller by 0.25 SD. At birth, length is about 0.5 SD above the WHO standard. This was also observed in the Norwegian MGRS study site (WHO Multicentre Growth Reference Study Group 2006a), and in the UK ALSPAC study (Wright et al. 2008). Wright et al. suggest that this is due to the size of the mother, a hypothesis which is corroborated by data from the Netherlands (van Buuren and van Wouwe, 2008). Birth weight is also higher in Flemish children, consistent with data from the UK ALSPAC study (Wright et al. 2008). MGRS site differences for weight have not been published.

By comparing average trends, we have until now treated the WHO growth charts as a reference and not as standard for optimal growth. The WHO claims that a growth reference

does not provide a sound basis for growth monitoring (de Onis and Garza 2006). However, most standards are also a reference, albeit in case of the MGRS, one for exclusively or predominantly breastfed children of high SES. From a statistical point of view, references have important advantages. Their validity can be readily tested in the target population, and the expiry date can be estimated from the analysis of secular trends. From an auxological point of view, an appropriate reference accounts for both the genetic background, however small, and for the shared environment. However, we cannot ignore the influence of breastfeeding on growth, and share the WHO point of view that this should be considered as the norm (Garza and de Onis 2004). In this analysis, we have however shown that population differences are dominant over feeding practice, and therefore tend to conclude that the use of an appropriate local reference, if available, is preferred for monitoring growth of individual children. Considering the WHO charts as a standard for growth of Flemish children also introduces a conceptual problem of over- and under-achievement, and is probably erroneous. Specifically for weight and BMI, the significance of an average SDS that is below (at 2–6 months of age) or above (from 18 months of age onwards) the median, even if criteria for optimal growth are met, is hard to interpret. The striking similarity of growth in weight of Flemish children, breastfed or not, with growth of children in the UK, compared to the WHO standard (Wright et al. 2008), illustrates that the observed trends are probably not random, but reflect true population differences. The pattern of growth in head circumference is even more problematic, as it is consistently 0.5 SD above the WHO standard.

In conclusion, we have shown that the use of a recent local growth reference is more appropriate for monitoring growth of children in Flanders, including those that are exclusively breastfed. This result can probably be generalized to other countries in the European region. The WHO growth standards are however a valuable alternative when such reference is not available. These standards also provide an independent benchmark to analyse and compare growth in different populations. Finally we support the basic principle of the MGRS that growth of breastfed children should be seen as the norm. We also endorse the principle that all children are equal, but not that they grow alike, since diversity is a typical characteristic of the human race.

Acknowledgements

This study was supported by a grant from the Minister of Public Health of the Flemish Government.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Cole TJ, Green PJ. 1992. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med 11:1305–1319.

Cole TJ, Paul AA, Whitehead RG. 2002. Weight reference charts for British long-term breastfed infants. Acta Paediatr 91:1296–1300.

de Onis M, Garza C. 2006. Foreword. Acta Pædiatr Suppl 450:5–6.

Dewey KG, Peerson JM, Brown KH, Krebs NF, Michaelsen KF, Persson LA, Salmenpera L, Whitehead RG, Yeung DL. 1995. Growth of breast-fed infants deviates from current reference data: A pooled analysis of US, Canadian, and European data sets. Pediatrics 96:495–503.

Eveleth PB, Tanner JM. 1990. Worldwide variation in human growth, (2nd edn). Cambridge: Cambridge University Press.

Garza C, de Onis M, for the WHO Multicentre Growth Reference Study Group. 2004. Rationale for developing a new international growth reference. Food Nutr Bull 25:S5–S14.

Kind en Gezin. 2008. Het kind in Vlaanderen 2007. Brussels: Kind en Gezin. http://www.kindengezin.be/Images/KIV%2007_tcm149-55931.pdf. 14 October 2008.

Newcombe RG. 1998. Two-sided confidence intervals for the single proportion: Comparison of seven methods. Stat Med 17:857–872.

R Development Core Team. 2008. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Roelants M, Hauspie R, Hoppenbrouwers K. References for growth and pubertal development from birth to 21 years in Flanders (Belgium). Ann Hum Biol. In press.

van Buuren S, van Wouwe JP. 2008. WHO Child Growth Standards in action. Arch Dis Child 93:549–555.

WHO Multicentre Growth Reference Study Group. 2006a. Assessment of differences in linear growth among populations in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 450:56–65.

WHO Multicentre Growth Reference Study Group. 2006b. Enrolment and baseline characteristics in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 450:7–15.

WHO Multicentre Growth Reference Study Group. 2006c. WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development. Geneva: World Health Organization.

WHO Multicentre Growth Reference Study Group. 2007. WHO Child Growth Standards: Head circumference-for-age, arm circumference-for-age, triceps skinfold-for-age and subscapular skinfold-for-age: Methods and development. Geneva: World Health Organization.

Wright C, Lakshman R, Emmett P, Ong KK. 2008. Implications of adopting the WHO 2006 Child Growth Standard in the UK: Two prospective cohort studies. Arch Dis Child 93:566–569.

Yngve A, Sjöström M. Breastfeeding in countries of the European Union and EFTA: Current and proposed recommendations, rationale, prevalence, duration and trends. Public Health Nutrition 2001 4:631–645.

4.3

Growth of Belgian and Norwegian children compared to the WHO growth standards: prevalence below -2 and above +2 SD and the effect of breastfeeding

Júlíusson PB, Roelants M, Hoppenbrouwers K, Hauspie R, Bjerknes R.

Archives of Disease in Childhood, 96(10): 916-21, 2011

Original article

Growth of Belgian and Norwegian children compared to the WHO growth standards: prevalence below −2 SD and above +2 SD and the effect of breastfeeding

P B Júlíusson,[1] M Roelants,[2,3] K Hoppenbrouwers,[3] R Hauspie,[2] R Bjerknes[1,4]

[1]Section for Pediatrics, Department of Clinical Medicine, University of Bergen, Bergen, Norway
[2]Laboratory for Anthropogenetics, Vrije Universiteit Brussel (VUB), Brussels, Belgium
[3]Department of Youth Health Care, Katholieke Universiteit Leuven, Leuven, Belgium
[4]Department of Paediatrics, Haukeland University Hospital, Bergen, Norway

Correspondence to
Pétur B Júlíusson, Section for Pediatrics, Department of Clinical Medicine, University of Bergen, 5021 Bergen, Norway; petur.juliusson@med.uib.no

PBJ AND MR contributed equally to this work.

Accepted 13 November 2009
Published Online First 30 November 2009

ABSTRACT

Background New national growth references have been published in Belgium and Norway. The WHO recommends universal use of their 2006 Child Growth Standards based on data from breastfed children.

Objective To compare the growth of Belgian and Norwegian children with the WHO standards.

Participants 6985 children 0–5 years of age from Belgium and Norway.

Design Proportion of children below −2 SD and above +2 SD of the WHO standards was calculated for length/height, weight, body mass index and head circumference. Average SD scores of exclusively breastfed children of non-smoking mothers were compared with national reference data and with the WHO standards.

Results Generally, the number of Belgian and Norwegian children below −2 SD lines of the WHO standards was lower and above +2 SD higher than expected. The largest differences were for head circumference (0.97% Belgian and 0.18% Norwegian children below −2 SD, 6.55% Belgian and 6.40% Norwegian children above +2 SD) and the smallest for length/height (1.25% Belgian and 1.43% Norwegian children below −2 SD, 3.47% Belgian and 2.81% Norwegian children above +2 SD). The growth pattern of breastfed children of non-smoking mothers was in both countries more alike the local national growth references than the WHO standards.

Conclusions There are significant deviations in the proportion of children outside normal limits (±2 SD) of the WHO standards. This was true for all children, including those who were exclusively breastfed. Hence, adoption of the WHO growth charts could have consequences for clinical decision-making. These findings advocate the use of national references in Belgium and Norway, also for breastfed children.

What is already known on this topic

- In 2006, the WHO issued growth standards showing children's growth under optimal conditions.
- The WHO advises universal use of these standards.

What this study adds

- The growth of Belgian and Norwegian children deviates significantly from the WHO growth standards, even when they are breastfed according to recommendations.
- In both countries head circumference is systematically 0.5 SD or more above the WHO growth standard.

Because of the variations in growth between populations, there has been a long tradition of constructing growth references aimed at specific target populations. Recently, new national growth references from birth to adulthood were published for Belgian[1] and Norwegian children.[2] In 2006, the WHO also issued universal growth standards that intend to show how children grow under optimal conditions (hence, "standards"), irrespective from which world region or population they originate (hence, "universal").[3] To achieve this goal, the WHO set up a multi-centre growth reference study (MGRS) in six countries (Brazil, Ghana, India, Norway, Oman, USA) and only included children of high social class (except in Norway and the USA where social class was not an inclusion criterion) of non-smoking mothers that were breastfed according to recommendations.[4] The results for length demonstrated that differences between the six countries were relatively small (although the medians differed up to 0.5 SD at certain ages), and universal use of these charts for all children worldwide was, therefore, recommended.[5] To this date, site differences for weight, body mass index (BMI) or head circumference have not been reported.

The WHO growth standards represent an improvement over the previous US National Centre for Health Statistics (NCHS)-based growth reference charts, but the implementation is nevertheless not straightforward in developed countries where updated references are available. Children in Hong Kong have shown to be shorter when compared with the WHO standard[6] and children in The Netherlands longer.[7] In the UK, birth weight and birth length have been found to be higher,[8] and consequently, the WHO standards were only adopted from 2 weeks of age and onwards. When comparing SD scores for length/height, weight, BMI and head circumference, three of us (MR, KH and RH) have recently shown that Belgian breastfed children grow similar in length when compared with the WHO standards, but not for other traits like weight, BMI and head circumference.[9] Collectively, these observations challenge the validity of the WHO standards as a universal tool for growth monitoring that should replace local references.[7] This is an important

issue because the choice of a local reference or the WHO international standards may affect thresholds used in clinical guidelines and can, therefore, have considerable consequences for both patients, parents and the healthcare system.

The aim of the present study was to compare the proportion of Belgian and Norwegian children outside the normal limits (defined by ±2 SD) of the WHO growth standards for length/height, weight, BMI and head circumference, with the expected number of 2.3%. A separate analysis was performed for children of a non-smoking mother breastfed according to recommendations.

PARTICIPANTS AND METHODS

The Flanders Growth Study (FGS) is a representative cross-sectional survey of 18 051 participants 0–21 years of age, recruited in Flanders (northern part of Belgium), in 2001–2004.[1] The Bergen Growth Study (BGS) is a cross-sectional survey of 8299 participants 0–19 years of age, recruited in the city of Bergen, Norway in 2003–2006.[2] The present analysis was limited to children of Belgian (n=4754) and Norwegian (n=2231) origin between 0 and 5 years of age. All children were healthy and born at term (37–42 weeks of gestational age), but low birth weight children born at term were not excluded (table 1).

The FGS and BGS are highly comparable in design and modus operandi, and both surveys included biometry and an assessment of social status and infant feeding practice. Length (before 2 years of age), height (from 2 years of age), weight and head circumference were measured by trained staff members, using a standardised technique and equipment. Weight, length and head circumference at birth, feeding practice and smoking in pregnancy were recorded through a parental questionnaire in both studies. Children were considered as exclusively breastfed when they were breastfed for at least 6 months, and no formula milk was introduced before that age. This includes also children <6 months who did not yet receive any formula milk at the time of measurement. Solids were not accounted for, but these are typically introduced between 4 and 6 months of age. A total of 683 Belgian and 527 Norwegian children were assigned to the exclusively breastfed subgroup (excluding 69 Belgian and 43 Norwegian children because of smoking during pregnancy). In addition, 304 Norwegian children were exclusively breastfed for at least 6 months and received continuous breastfeeding (exclusive or not) up to at least 12 months of age (excluding 16 children because of smoking during pregnancy). Only 26 children in the Belgian sample met this criterion.

STATISTICAL ANALYSIS

Participants were grouped into 4 months interval between birth and 1 year of age and yearly intervals thereafter. The prevalence of children below and above certain thresholds for length, height, weight, BMI and head circumference is reported by age group or adjusted for age (per year) otherwise. For the analysis of breastfed children of non-smoking mothers, all traits were converted to SD scores according to the WHO growth standards.[3] The reference group consists of all children on which the Belgian and Norwegian growth charts are based, including the group of exclusively breastfed children by definition.

Table 1 Number of children in the Belgian and Norwegian sample

	Belgium			Norway		
Age	Boys	Girls	Total	Boys	Girls	Total
0–1	999	951	1949	601	558	1159
1–2	340	338	678	132	141	273
2–3	320	317	637	151	120	271
3–4	328	369	697	125	145	270
4–5	427	366	793	131	127	258
Total	2414	2340	4754	1140	1091	2231

Means and proportions are always presented with 95% CIs. CIs for simple proportions are the exact Clopper–Pearson 95% confidence limits.[10] Age-adjusted rates and confidence limits are based on the γ distribution,[11] assuming an equal number of children in each age group. Size at birth was based on recall data from all participants in the sample and constitutes a separate age class. Global prevalence estimates are only based on postnatal growth measurements.

The Flanders Growth Study was approved by the ethics committee of the Vrije Universiteit Brussel and the Bergen Growth Study by the Regional Committee for Medical Research Ethics and the Norwegian Data Inspectorate.

RESULTS

Percentage of all children outside −2 SD and 2 SD

Compared to the WHO growth standards, we generally saw fewer children below −2 SD and more children above +2 SD than expected (fig 1). Remarkably, similar trends were observed in Belgian and Norwegian children.

At birth, the percentage of Belgian children below −2 SD was as expected for length, weight and head circumference, but too high for the BMI. The percentage of children above +2 SD was highly elevated for length (8%) and head circumference and comparable for weight and BMI. The percentage of Norwegian children below −2 SD at birth was as expected for length and BMI, but too low for weight and head circumference. The percentage of children above +2 SD was highly elevated for all variables, and up to 13% for length and head circumference.

The age-adjusted prevalence of all children 0–5 years of age with a *length* below −2 SD was 1.25% (0.92–1.65) in Belgium and 1.43% (0.90–2.17) in Norway. The corresponding percentage above +2 SD was 3.47% (2.90–4.14) and 2.81% (2.11–3.69), respectively. In both countries, there was a tendency to converge with the WHO standard for length towards the older age groups, and the proportion of children outside normal limits became reasonably close to the expected amount of 2.3% by the age of three (fig 1). For *weight*, the overall percentage below −2 SD was 0.65% (0.45–0.93) for Belgian and 0.56% (0.29–1.04) for Norwegian children. The percentage above +2 SD was, respectively, 3.08% (2.54–3.71) and 3.39% (2.56–4.44). Unlike for length, there was no apparent trend with age. A similar observation was made for the BMI, where the overall percentage below −2 SD was 0.84% (0.59–1.18) for Belgian and 0.54% (0.29–0.99) for Norwegian children. The corresponding percentage above +2 SD was 3.01% (2.45–3.67) and 4.29% (3.30–5.50). For *head circumference*, the overall percentage below −2 SD was 0.97% (0.70–1.33) for Belgian and only 0.18% (0.05–0.53) for Norwegian children. The prevalence of children above +2 SD was relatively high with, respectively, 6.55% (5.76–7.42) and 6.40% (5.19–7.83). Apart from birth, the largest discrepancy was observed between 1 and 3 years of age.

The effect of breastfeeding

To document the effect of breastfeeding, we compared the growth of a subgroup of 1210 breastfed children born to non-smoking mothers with the WHO growth standards in both countries. This was done by comparing mean SD scores rather than the prevalence outside the normal range, because the

Original article

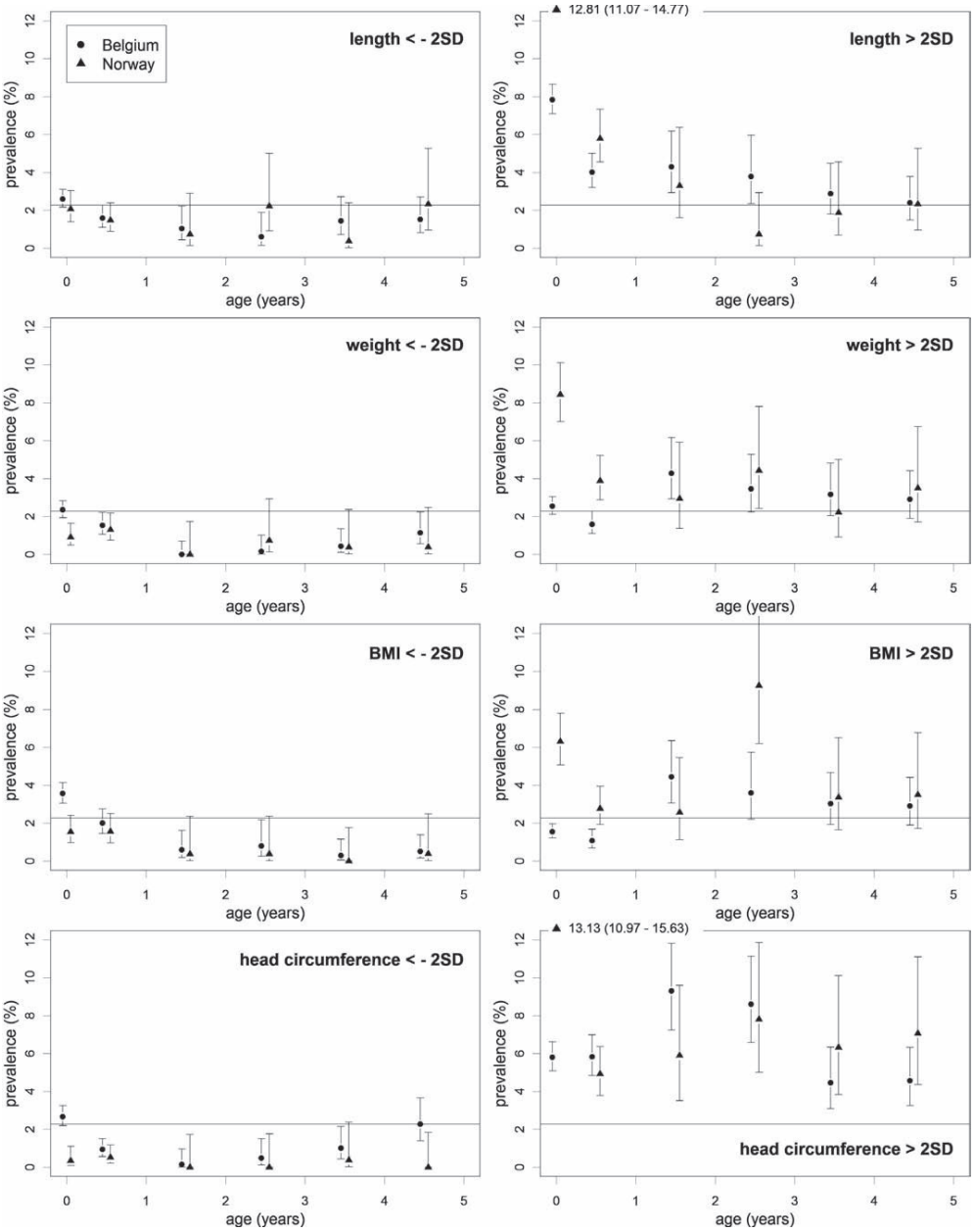


Figure 1 Percentage (95% CI) of observations below −2 SD (left panels) or above +2 SD (right panels) of the WHO growth standards, in Belgium (circles) and Norway (triangles). Data are grouped by calendar age and plotted half way the interval. Birth data are analysed separately and plotted at 0 years. Horizontal lines indicate the expected prevalence of 2.3 percentiles. Prevalence higher than 12% is marked at the top of the figure.

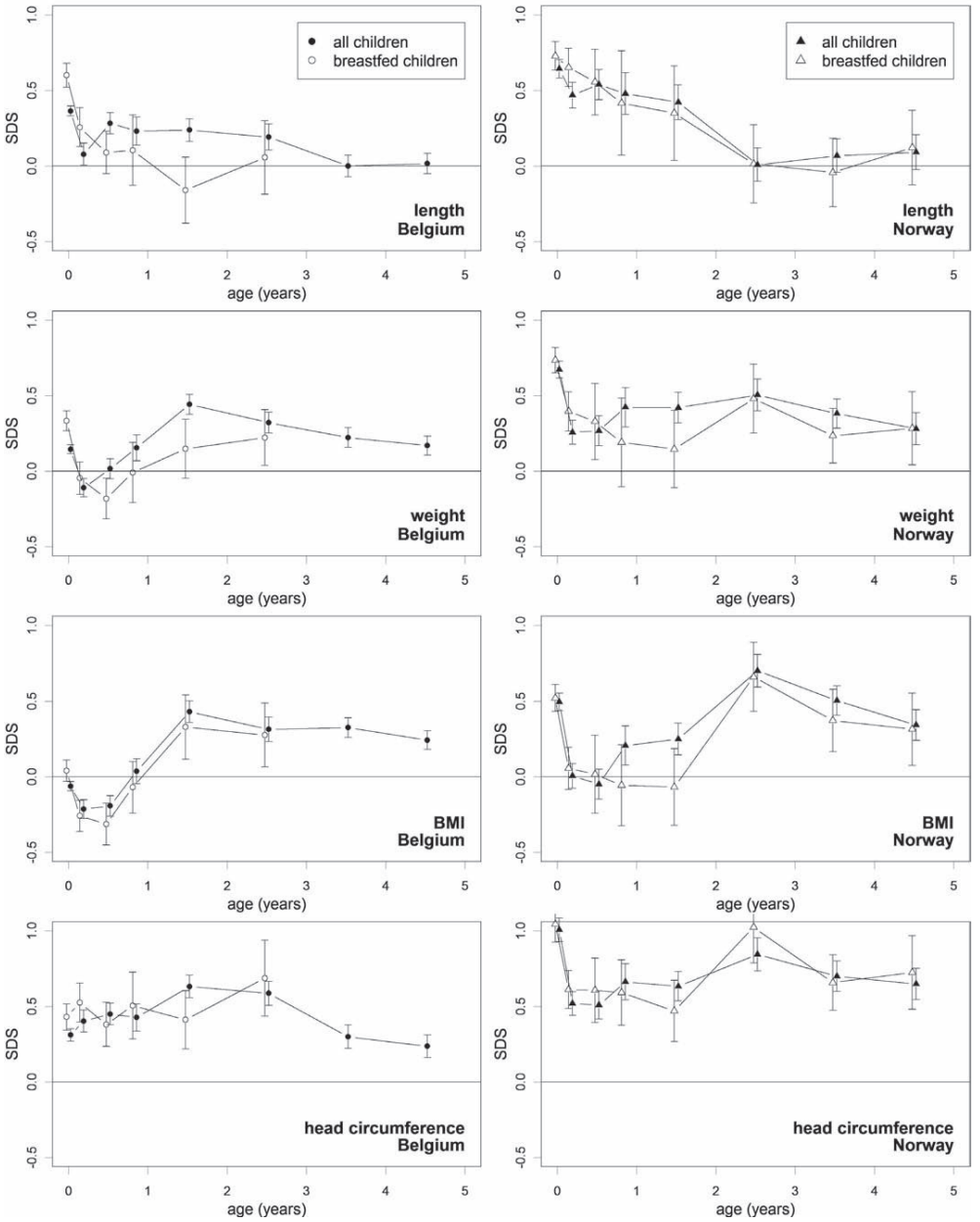


Figure 2 Mean SDS based on WHO growth standards from birth to 5 years of age for all children (open symbols) and children breastfed according to the WHO recommendations (filled symbols) in Belgium (left) and Norway (right). Observations are grouped per 4 months the first year and yearly thereafter. Error bars represent 95% confidence limits of the mean.

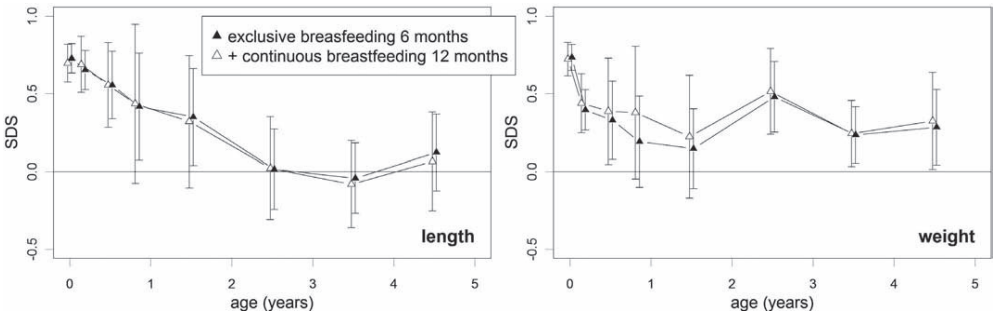


Figure 3 Mean SDS of children exclusively breastfed for at least 6 months (filled triangles) and children that also received continuous partial breastfeeding up to at least 12 months of age (open triangles) in the Norwegian sample. Error bars represent 95% confidence limits of the mean.

number of children below –2 SD or above +2 SD would become too small. The prevalence of exclusive breastfeeding for at least 6 months was 26.5% in the Belgian sample[9] and 40.6% in the Norwegian sample (based on recall data from children 6 months to 5 years of age). This analysis demonstrated that the growth of exclusively breastfed children was more similar to that of all children in the local population (including those that were not breastfed) than to the WHO growth standards (fig 2). Between 6 months and 2 years of age, Belgian breastfed children had generally a smaller length and less weight and are, therefore, somewhat closer to the WHO growth standards, although the overall pattern was still more like that of the reference population. For Norwegian children, this was true for weight and BMI. To document the possible additional effect of continuous breastfeeding up to at least 12 months of age, the growth of a subgroup of 304 Norwegian children that comply with these criteria was compared to those that were exclusively breastfed for 6 months (fig 3). The length and weight were highly similar in the two groups, and no additional effect could be demonstrated in this study. Similar results were obtained for BMI and head circumference (data not shown).

DISCUSSION

In this paper, we have demonstrated significant deviations in the proportion of Belgian and Norwegian children that are located outside the normal limits of the WHO growth standards, generally defined as an SDS below –2 SD (2.3 percentile) or above +2 SD (97.7 percentile). Although these cut-offs do not warrant immediate action or referral to specialised care, they are often included in clinical guidelines, and adoption of the WHO growth charts could, therefore, affect the clinical decision making process.[8] Compared to the WHO standards, the number of Belgian and Norwegian children below –2 SD is generally lower, and the number above +2 SD higher than expected, for length/height, weight, BMI and head circumference, with the largest discrepancy observed for head circumference. As the WHO standards are based on a sample of breastfed children born to a non-smoking mother, one could speculate that these differences mirror different inclusion criteria. However, our observation, that the growth of exclusively breastfed children in both Belgium and Norway was generally closer to that of the local reference population than to that of the WHO standards, refutes this hypothesis and probably reflects that genetic and environmental background factors have stronger influence on growth than feeding practices. This perception is strengthened by the similarity of the growth patterns observed in both countries as compared to the WHO growth standards.

In the present study, criteria for exclusive breastfeeding were chosen to mimic those of the MGRS. These criteria differ slightly from current WHO recommendations, which advise the introduction of solids at the age of 6 months.[12] In the MGRS, the previous recommendation of introducing solids between 4 and 6 months of age was used. Solids were not accounted for in the present study, but it was previously shown that only a minority of breastfed children received solids before the age of 4 months.[13 14] In children >2 years, the MGRS inclusion criteria were less strict (at least 3 months of predominant breastfeeding, with no additional requirements about complimentary foods), while we applied the same definition throughout the whole age range. Due to the cross-sectional design of this study, the feeding practice of children younger then 6 months was only documented up to the time of measurement. This is in contrast with the MGRS, where the feeding practice of all children <2 years of age was documented over the whole trajectory. While this could affect the composition of the target group, any consequences on the growth of breastfed children between birth and 6 months cannot be due to breastfeeding per se.

The average length is larger for Norwegian children but relatively comparable for Belgian children when compared with the WHO growth standards, although these differences tend to disappear by 3 years of age. The proportion of children below –2 SD was nevertheless considerably smaller than expected. A comparable pattern of average growth in length was also observed in the Norwegian sample of the WHO standards,[5 7] which suggests a possible population effect.

In the case of BMI, adoption of the WHO growth standards would lead to a lower prevalence of underweight and an increased prevalence of overweight in Belgium and Norway. A comparable observation was made in the UK by Wright *et al*,[8] who saw this as a positive feature, partly because of the current increase in childhood overweight. This is, however, contradicted by the observation in our study that, compared to the WHO growth standards, weight and BMI in breastfed children is broadly comparable to that of all children. Any comparisons with the WHO standards for weight and BMI are nevertheless limited by the fact that the WHO did not publish site differences for these traits.[15] Comparable results were obtained when plotting SD scores of weight for length/height (data not shown).

The most striking differences were observed for head circumference. When applied in Belgium or Norway, the proportion below –2 SD of the WHO growth standards was less than half the expected amount (only 8% of the expected number in Norway); approximately 70% of all children are above the median (data not shown); and the proportion above +2 SD would increase up to threefold. In Norway, in each annual birth cohort of approximately 60 000 children about 4000 would be above the 97th percentile (expected 1500) and only about 100 would be located below the 3rd percentile (expected 1500). Although clinical guidelines for head circumference are seldom based on a single measurement, clinicians and medical staff of well baby clinics and school health centres would find a very high proportion of their population with large heads, while small heads would be virtually non-existing. This might lead to a large uncertainty concerning referral criteria and unwarranted examinations of many children and introduce unnecessary worries in many parents.

One of the main reasons for constructing the WHO growth standards was the observation that breastfed children grow differently from the previous NCHS/WHO international growth references, with faster growth in length and weight during the first 6 months and slower growth thereafter.[16] A tendency towards this pattern was observed in the present study, although the general trend was more like that of the local reference population. In our opinion, the present observations do not favour the use of the WHO growth standards for monitoring growth of children of Belgian or Norwegian origin, a result that may also be generaliseable to other North/West European populations. The WHO growth standards are nevertheless a valuable instrument, (a) as an independent platform for international comparisons and (b) as a tool for growth monitoring when a suitable local reference is unavailable, although differences in growth should be accounted for. Moreover, the concept of the WHO growth study has put breastfed children forward as the "norm" rather than the other way around, a view that we fully support.

In conclusion, we have demonstrated a significant deviation in the proportion of Belgian and Norwegian children that are located outside the normal limits for length/height, weight, BMI and head circumference of the WHO growth standards. Consequently, the implementation of the WHO growth standards would have unwanted clinical implications that are not supported by this study. This suggests that at least in Belgium and Norway, the use of a local reference based on a representative sample of healthy children is more appropriate for all children, bottle-fed or breastfed.

Acknowledgements The authors are most grateful to the children and their families for taking part in the Flanders Growth Study and Bergen Growth Study.

Funding The Flanders growth study was supported by a grant from the Minister of Public Health of the Flemish Government. The Bergen Growth Study was supported by The Western Norway Regional Health Authority, University of Bergen, Novo Nordisk AS and Merck Serono SA.

Competing interests None.

Ethics approval The Flanders Growth Study was approved by the ethics committee of the Vrije Universiteit Brussel, and the Bergen Growth Study by the Regional Committee for Medical Research Ethics and the Norwegian Data Inspectorate.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Roelants M**, Hauspie R, Hoppenbrouwers K. References for growth and pubertal development from birth to 21 years in Flanders, Belgium. *Ann Hum Biol* 2009;**36**:680–94.
2. **Júlíusson PB**, Roelants M, Eide GE, *et al.* [Growth references for Norwegian children]. *Tidsskr Nor Laegeforen* 2009;**129**:281–6.
3. WHO Multicentre Growth Reference Study Group. WHO child growth standards: methods and development. Geneva: World Health Organization, 2006.
4. WHO Multicentre Growth Reference Study Group. Enrolment and baseline characteristics in the WHO Multicentre Growth Reference Study. *Acta Paediatr Suppl* 2006;**450**:7–15.
5. WHO Multicentre Growth Reference Study Group. Assessment of differences in linear growth among populations in the WHO Multicentre Growth Reference Study. *Acta Paediatr Suppl* 2006;**450**:56–65.
6. **Hui LL**, Schooling CM, Cowling BJ, *et al.* Are universal standards for optimal infant growth appropriate? Evidence from a Hong Kong Chinese birth cohort. *Arch Dis Child* 2008;**93**:561–5.
7. **van Buuren S**, van Wouwe JP. WHO child growth standards in action. *Arch Dis Child* 2008;**93**:549–51.
8. **Wright C**, Lakshman R, Emmett P, *et al.* Implications of adopting the WHO 2006 Child Growth Standard in the UK: two prospective cohort studies. *Arch Dis Child* 2008;**93**:566–9.
9. **Roelants M**, Hauspie R, Hoppenbrouwers K. Breastfeeding, growth and growth standards: performance of the WHO growth standards for monitoring growth of Belgian children. *Ann Hum Biol* 2009;**1**:1–8.
10. **Newcombe RG**. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* 1998;**17**:857–72.
11. **Fay MP**, Feuer EJ. Confidence intervals for directly standardized rates: a method based on the gamma distribution. *Stat Med* 1997;**16**:791–801.
12. World Health Organization. Complementary feeding: report of the global consultation, and summary of guiding principles for complementary feeding of the breastfed child. Geneva: World Health Organization, 2002.
13. **Lenaers S**, Goffin I, Alliët Ph, *et al.* Onderzoek naar de voedingssituatie van jonge kinderen. Sein, Hasselt, Belgium, 2002. http://www.kindengezin.be/Images/Document _ tcm149-12808.pdf (Accessed November 2009).
14. **Lande B**, Andersen LF, Baerug A, *et al.* Infant feeding practices and associated factors in the first six months of life: the Norwegian infant nutrition survey. *Acta Paediatr* 2003;**92**:152–61.
15. WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards based on length/height, weight and age. *Acta Paediatr Suppl* 2006;**450**:76–85.
16. WHO Working Group on Infant Growth. An evaluation of infant growth: the use and interpretation of anthropometry in infants. *Bull World Health Organ* 1995;**73**:165–74.

4.4

Conditional growth and correlations between longitudinal measurements of length, height, weight and head circumference in infancy and childhood

Roelants M, Hoppenbrouwers K., Hauspie R.

Submitted 2013

Conditional growth and correlations between longitudinal measurements of length, height, weight and head circumference in infancy and childhood

Mathieu Roelants[1,2], Karel Hoppenbrouwers[2], Roland Hauspie[1]

ABSTRACT

Background: Growth distance charts are not suitable to assess changes in growth over time. Conditional change in the standard deviation score (SDS), relative to a growth chart, has been suggested as a better alternative for growth velocity, but requires an estimate of the correlation between measurements. **Aims**: To estimate and validate a reference for conditional growth based on the correlation between measurements. **Methods**: The correlation structure of length, weight and head circumference from birth to 3 years of age was estimated in 862 newborn infants, measured with monthly (up to 1 year) or quarterly (1 – 3 years) intervals. Annual correlations in height were estimated in 2222 schoolchildren who were measured twice with an interval of one year. **Results**: Correlations of weight by age and interval correspond well with published data from the UK, and those of height in schoolchildren with data from France. An algorithm to obtain the correlation between length, height, weight, or head circumference at different ages is provided and validated. **Conclusions**: Conditional growth, based on a simple descriptive algorithm for the correlation structure is shown to be a flexible and valid method for the assessment of growth between two time points.

[1]Laboratory of Anthropogenetics, Department of Biology, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium;
[2] Centre of Youth Health Care, Department of Public Health and Primary Care, KU Leuven – University of Leuven, Kapucijnenvoer 35, 3000 Leuven
Corresponding author: Mathieu Roelants, Department of Public Health and Primary Care, KU Leuven, Kapucijnenvoer 35, 3000 Leuven. mathieu.roelants@med.kuleuven.be

INTRODUCTION

When comparing a measurement of length, weight or head circumference to a growth reference chart, clinicians often assess the change in centile position from a previous measurement, using informal criteria to judge which change is "acceptable" or "normal". However, cross-sectional growth distance charts give no guidance regarding changes in growth because the centile lines do not represent individual growth curves. Such criteria have been shown to be biased (Cole 1995), and may lead to large numbers of children referred to specialized care (van Buuren *et al.*, 2004).

Appropriate instruments to monitor changes in growth are references for growth velocity or growth increments. They are based on longitudinal data, and show the distribution of changes in growth over time at a particular age and a particular interval, just like the distance chart shows the size at a given age. However, growth velocity is a complex parameter that depends on age and has a variance that depends both on age and the time interval between measurements (van 't Hof *et al.*, 2000). Therefore velocity references are usually limited to a specific interval between measurements (van 't Hof *et al.*, 2000, WHO, 2009). The most complete velocity reference is probably that by Chrzastek-Spruch *et al.* (1994) who estimated mean increments and percentiles 10 and 90 over all possible intervals between monthly (0 – 12 months) and quarterly (1-3 years) measurements. Using velocity references is however cumbersome, and often involves either an approximation (when the nearest age or interval is used as a reference), or a complex interpolation of the mean and variance between different charts.

Regression standards for conditional growth are an alternative approach that has been suggested by Healy (1974). Based on earlier work by Cameron (1980) and Berkey *et al.* (1983), Cole (1995) and Wright *et al.* (1994) used correlations to estimate the change in standard deviation score (SDS) of a measurement, conditional on the SDS of a previous measurement, i.e. the change in position on a growth distance chart. Working on the SDS scale has several advantages over the measurement scale (Cole, 1995). If an appropriate growth reference is used, the SDS follows a standard Gaussian distribution, even for skew distributed measurements like weight. The expected position of a measurement on the centile

chart is given by: $SDS^*_2 = r_{1,2}$ x SDS_1 , i.e. conditional on SDS_1, where $r_{1,2}$ is the correlation between the measurements at ages t_1 and t_2, and '*'denotes the predicted value. The variance of SDS^*_2 is $1 - r_{1,2}^2$, so the conditional "normal range" (say 95%) of a measurement is $SDS^*_2 \pm 1.96\sqrt{(1 - r_{1,2}^2)}$, where 1.96 is the normal equivalent deviate that corresponds to a coverage of 95%. Alternatively, the result can be expressed as a conditional gain (Cole, 1995) or "thrive index" (Wright *et al.*, 1994), by dividing the difference between the measured and predicted SDS by its corresponding standard deviation: Conditional gain = $(SDS_2 - r_{1,2}$ x $SDS_1) / \sqrt{(1 - r_{1,2}^2)}$, which is an SDS in itself, and can be expressed as a percentile (Cole, 1995). This approach is simple yet attractive because it has a direct link with cross-sectional distance charts, and both the expected outcome and variance depend on the correlation only. Along with a suitable growth distance chart, an algorithm for the correlation structure is the only requirement to assess conditional growth. Moreover, if correlations between successive measurements are comparable across populations, it would be straightforward to implement this approach in other populations (Cole, 1995).

Correlations of weight SDS in infancy have been published by Cole (1995), based on 223 complete cases from the Cambridge Growth Study, and by and Wright *et al.* (1994), using data extracted from health records in Newcastle. To our knowledge, these are the only suitable data published to date. The correlation structure of the Cambridge series was summarized by Cole (1995) with a six-term fractional polynomial to estimate fisher transformed correlation coefficients. Wade and Ades (1998) explored several correlation structures for growth from birth to 8 years of age, including a 4 parameter exponential autoregression model of age and interval. The most simple model is without doubt given by Argyle *et al.* (2008), who estimated the correlation coefficient from a two-parameter exponentiated ratio of ages, but earlier work also included different functions of age or interval, and models with more parameters (Argyle, 2002).

In this paper, we present correlations between SDS in length, weight and head circumference over variable intervals from birth to 3 years, and explore candidate algorithms to summarize the correlation structure. Correlations are compared with existing data, and the validity of conditional growth SDS is assessed by comparing actual and predicted growth.

MATERIALS AND METHODS

Study population

The correlation structure in infancy was estimated from a sample of 862 term born (37-42 weeks) singletons (440 boys and 422 girls), recruited from September 2002 to June 2003 in 50 Well Baby Clinics and 25 Day Care Centres for a three year follow-up (longitudinal infant sample). Up to 20 measurements were scheduled with monthly (0 – 1 year) or quarterly (1 – 3 years) intervals. Including birth records, a total of 12595 contacts were registered, but the individual number of visits ranged from 3 to 21. A large majority (n = 713, 82.7%) of children were measured on 10 or more occasions, including 73 who completed the whole planned schedule. The number of children who were lost to follow up was limited since the majority (n = 525; 60.9%) were measured on one or more occasions during the third and final year of the study. In principle, measurements were taken on the precise age, with a small allowance of 7 (1 – 12 months) or 14 (15 – 36 months) days, excluding 717 out of 9175 visits (7.8%) during the first year, and 398 out of 3420 visits (11.6%) during the second and third year. The total number of valid contacts eligible for analysis of the correlation structure was therefore 11480, except for missing measurements of height (n = 89), weight (n = 26) or head circumference (n = 417). Model validation was based on all observations.

Correlations in height during childhood were estimated from a sample of 2222 schoolchildren (1117 boys and 1105 girls), 3 – 9 years of age, recruited in 59 School Health Centres. Data from 2920 children 9 – 18 years of age were not included in this analysis because the pubertal spurt may have started. All children were measured on two occasions, with an interval of approximately one (0.85 – 1.15) year (single increment sample).

All children included in this analysis were of Belgian origin, and free from known diseases or conditions that could affect growth for the duration of the study. Social stratification showed no differences with national census data. Both samples are part of a representative sample of Flemish children that was used to estimate Belgian growth reference charts (Roelants *et al.*, 2009), to which the infants

contributed a single, randomly chosen, measurement between birth and 12 months, and the schoolchildren contributed only the first of two measurements. More details on study design and sample selection are available in Roelants *et al.* (2009).

This study was approved by the Ethics committee of the Vrije Universiteit Brussel. Written informed consent was obtained from a parent or legal guardian of each participating child.

Biometry

Recumbent length (up to 2 years), standing height (from 2 years onwards), weight and head circumference were measured by trained staff members of the participating centres using a standard measuring technique. Size at birth was obtained by parental questionnaire. Length/height, weight and head circumference were converted to SD scores using the national growth reference (Roelants *et al.*, 2009).

Correlation structure

The Pearson product-moment correlation coefficients of length, weight and head circumference SDS in infancy were calculated for each combination of the 21 target ages (210 in total) using all available pairs of measurements. Data from boys and girls were combined as they did not systematically differ from each other. No age correction was applied since all observations were confined to a narrow age range, and both the mean and variance of the SD scores are constant. No correction for gestational age was applied for the SDS at birth, but the sample included only term born singletons.

For the correlations between yearly measurements of height from 3 – 9 years of age, children were grouped by age at last birthday (calendar age). No correction for the grouping was necessary as both the mean and variance are constant within age groups when working on the SDS scale.

The distribution of SD scores at each age was tested with the Kolmogorov-Smirnov test for a standard Gaussian distribution, and with an omnibus test for skewness and kurtosis (D'Agostino *et al.*, 1990). Departures from the standard Gaussian distribution were further examined with a t-test for zero mean and chi-square test for unit variance.

Several models from the literature were explored for their suitability to describe the correlation structure between ages t_1 and t_2 within the longitudinal infant sample. Fractional polynomials of fisher transformed correlation coefficients (Cole, 1995) did not result in a decreasing correlation with time interval, presumably because of noise in the data. The extended autoregression model used by Wade and Ades (1998), $r_{1,2} = \sigma^2 \left(\rho_1 + \rho_2 \log(t)\right)^{(t_2 - t_1)^\gamma}$, fitted nicely to correlations in length up to 12 months, but not for older ages, nor for weight or head circumference. Argyle *et al.* (2008) used a two-parameter function of t_1 and t_2 for correlations in weight, but different variants of the form $r = \kappa \left(t_1 + \tau_1\right)^{\lambda_1} / \left(t_2 + \tau_2\right)^{\lambda_2}$ were explored earlier by Argyle (2002). The two parameter model has $\kappa = 1$, $\tau_1 = \tau_2$ and $\lambda_1 = \lambda_2$, but the use of a multiplier κ is discussed in Argyle *et al.* (2008). Variants of this model were estimated from the 210 correlations of all combinations of t_1 and t_2 using weighted nonlinear regression with the variance stabilizing Fisher transformation, $z = \log[(1+r)/(1-r)]/2$, applied to both sides of the equation. Weights were set to $(n - 3)$, where n is the number of pairs of observations for the corresponding observed r. The adequacy of all models was assessed visually by analysis of weighted model residuals, and by comparing fitted and observed correlations along t_1 and t_2. Final models were selected by the Akaike (AIC) and Bayesian Information Criteria (BIC). The standard deviation of the unweighted differences between observed and predicted correlations is given as a measure of prediction error. We also tried to estimate the correlation structure directly from the observed SDS using the relation: $\text{SDS}_2 = f(r|t_1,t_2) * \text{SDS}_1$, where $f(r|t_1,t_2)$ is the function to be estimated, using iterated weighted nonlinear regression with weights equal to 1 for the first iteration, and $1/(1- f_{i-1}(r|t_1,t_2)^2)^2$ thereafter, where $f_{i-1}(r|t_1,t_2)$ is the correlation function estimated during the previous iteration. However, these models either did not converge or produced unrealistic biased estimates.

Validation

Conditional growth assessment based on correlation models was verified by converting the second measurement of all available pairs in the data set to a conditional growth $\text{SDS} = (\text{SDS}_2 - r_{1,2} \times \text{SDS}_1) / \sqrt{(1 - r_{1,2}^2)}$. The distribution of this variable was analysed for each combination of ages. Validation statistics are based on all 12595 contacts, but excluding measurements taken less than 3 weeks (0-12 months) or less than 2.5 months (12 – 36 months) apart, as they imply

extrapolation. Conditional annual growth SDS in children 3 – 8 years of age was assessed analogously in half year cohorts.

All statistical analysis was done in R, version 2.15.2 (The R foundation for statistical computing, Vienna, Austria, 2012).

RESULTS

Correlation structure

Correlations of length/height, weight, and head circumference SDS for all target ages in infancy and early childhood are listed in tables 1 – 3. Analysis of the marginal distribution of SDS, showed evidence for departure from the standard Gaussian distribution in 4 age groups for length, and in one age group for weight, all attributable to a non-zero mean; and in 6 age groups for head circumference, attributable to a non-zero mean and smaller variance. There was no evidence for either skewness or kurtosis in the data. Compared with the correlation structure based on internal SDS, which showed little evidence of deviation from the standard Gaussian distribution, correlations differed at most 0.01, and were 9 out of 10 times identical up to the second decimal.

[Tables 1 – 3 are shown at the end of this chapter]

The correlation structure of length, weight and head circumference shows a typical pattern of decreasing correlation as the time between measurements increases, and increasing correlation as the child grows older (figure 1). Although similar patterns are observed in all three traits, the magnitude of the correlation coefficients, and rate of change with age or time interval are different. Correlations of head circumference are usually higher and appear to level out from the second year onwards, while those of length and weight continue to decrease with increasing time interval between measurements. Correlations between lengths are mostly higher than those between weights during the first months, and lower thereafter.

One-year correlations of height SDS continue to increase in childhood from about 0.973 in 3–year old children to 0.986 by the age of six years (table 4). After 9 years

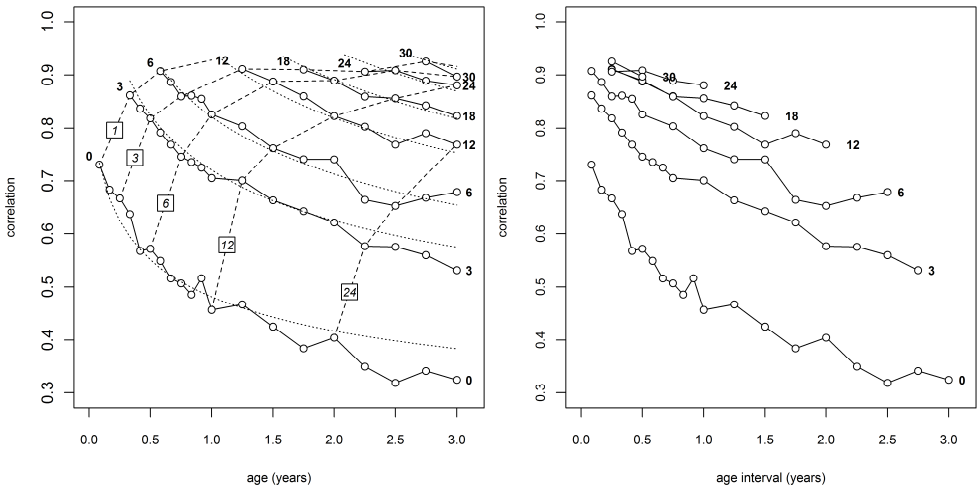


Figure 1: Evolution of the correlation between measurements of length/height with age_1 and age_2 (left panel) and with the measurement interval by age at the first measurement (right panel). Bold numbers indicate the age at the first measurement. Full lines connect points with the same age at the first measurement, and dashed lines (left panel) connect points with the same measurement interval (numbers in squares indicate the interval in months). Dotted lines show the predicted correlation by age and interval.

of age, the correlations decrease temporarily (to 0.96 in girls and 0.95 in boys), coinciding with the pubertal growth spurt (data not shown). Contrary to the previous tables these correlations apply to a whole year cohort. When grouped in quarterly intervals, the correlation increases from 0.967 at 3 years of age to 0.986 at 5 years of age, and remains constant thereafter (data not shown). The increase from the 12 month correlation between 24 and 36 months in table 1 to the correlation at 3 years of age is relatively large (from 0.88 to 0.967), but no data are available to study trends in this age range.

Table 4: Correlation of height SDS and height SDS one year (0.85 – 1.15 year) later, per age group.

age first measurement[1]:	3	4	5	6	7	8
correlation:	.973	.981	.986	.986	.985	.988
n:	273	353	478	389	375	354

[1] Calendar age, 3 years includes all children aged 3.0 to 3.9 etc.

Correlation models in infancy

Given the similarities in the shape of the correlation structure, models that were initially proposed for weight (Cole, 1995; Argyle *et al.*, 2008) or height (Wade and Ades, 1998) were fitted to all traits, but only those by Argyle et al. (2008) produced

satisfactory results. The two parameter model proposed by Argyle et al. (2008) was not flexible enough for the whole correlation structure and produced biased results, but less so for weight (τ = 0.04 and λ = 0.27). Starting from $r = \kappa(t_1 + \tau_1)^{\lambda_1} / (t_2 + \tau_2)^{\lambda_2}$ and setting in turn $\kappa = 1$, $\tau_1 = \tau_2$ and $\lambda_1 = \lambda_2$, the best fitting models with the smallest BIC were in each case of the form $\lambda_1 = \lambda_2$. With ages t_1 and t_2 expressed in years, simplified model expressions are for length/height: $r = 0.95 * [(t_1 + 0.05) / (t_2 + 0.08)]^{0.22}$; and for weight: $r = 0.99 * [(t_1 + 0.03) / (t_2 + 0.02)]^{0.26}$. The corresponding residual standard errors (RSE) are 0.023, 0.036, respectively, but the RSE of weight reduced to 0.018 for measurements taken after one month of age. For head circumference, however, none of the tested models gave a reasonable approximation above 2 years of age. Based on measurements up to 2 years the best fitting model was $r = 0.99 * [(t_1 + 0.01) / (t_2 + 0.13)]^{0.15}$, with a residual standard error of 0.018. A model based on all data up to 3 years of age was highly comparable, but validation statistics showed significant bias for measurement pairs with one or both measurements taken above 2 years of age.

Validation of conditional growth

The distribution of the conditional gain z-scores within age groups in infancy (defined by combinations of t_1 and t_2) was close to theoretical values of zero mean and unit variance (table 5). The percentage outside the 95% prediction intervals was on average 5%, but in about 20 – 30 % of age groups more children were above than below this interval (2.7 – 2.9% vs. 2.0 – 2.2%). This means that the correlation was more frequently under- than overestimated. The difference cannot be large, because a low correlation has the effect of a low conditional mean and a more narrow distribution of SD scores at the same time. The group standard deviations are on average close to one, or slightly below, which confirms a reasonable match of the correlation coefficients. Validation statistics of the yearly increment sample are comparable in children 3.5 to 8 years of age, with averages for the group means and standard deviations of -0.02 (0.21), and 0.99 (0.08) respectively, and 5.8% (2.0%) of children outside the 95% prediction interval. In the age group 3 to 3.5 correlations were on average underestimated, and the mean conditional SDS was larger than 0.5. The number of children in this age range is however small (n = 105).

Table 5: Summary of the conditional change in SDS in infancy[1].

	Length/height	Weight	Head circumference[2]
Mean	-0.01 (0.10)	0.01 (0.07)	-0.01 (0.10)
SD	0.98 (0.04)	0.98 (0.05)	0.97 (0.05)
Percentage outside 95% prediction interval	5.0 (1.0)	4.8 (1.1)	4.9 (1.1)

[1] Mean (and standard deviation) of 210 age group means, standard deviations, and percentages outside the 95% prediction interval. [2] Includes only measurements from birth to 24 months (136 age groups).

DISCUSSION

Conditional SDS is a simple and flexible method to assess changes in growth between two time points, but requires a suitable growth distance reference to convert measurements to SD scores, and a model for the correlation structure between time points. In this paper we have presented correlations and provided a descriptive correlation model based on longitudinal data of length/height, weight, and head circumference in infancy and height in childhood.

The correlation structures given in tables 1 – 3 are based on the longitudinal follow-up of a large number of infants. Although less than 10% completed the whole measurement schedule, the compliance was reasonably high considering the effort that was required from the parents, and the majority of children were still actively involved during the final study year. Our data of weight from birth to 2 years are quite comparable with the Cambridge (Cole, 1995) and Newcastle (Wright *et al.*, 1994) series, to our knowledge the only published correlations based on SDS. The mean (SD) difference with all 105 correlations of the Cambridge series is 0.005 (0.022), and less than 0.03 in 9 out of 10 age groups (correlations were bilinear interpolated because the Cambridge infants were weighed every 4 weeks). The mean difference with the Newcastle series is 0.016 (0.024), and less than 0.03 in 7 out of 10 age groups. No trend with age or interval was observed. We were not able to trace comparable data on length or head circumference. Although both traits are Gaussian distributed and do not require conversion to SD scores (the correlation does not depend on the scale), they should be measured in narrow age groups, or be properly scaled when calculated directly from measurements (Goldstein, 1981). Correlations based on data that are generated or interpolated with a mathematical growth model, tend to smooth out variability and will overestimate the correlation (Berkey *et al.*, 1983).

Correlations in height are listed up to 8 years of age (calendar age), which allows assessment of follow-up measurements up to 10 years of age (from 8.9 to 9.9 years). Above this age, the use of conditional growth based on a single follow-up measurement alone is questionable as the group becomes heterogeneous with respect to the timing of their pubertal spurt. The yearly correlations in height differ less than 0.01 units from those reported by Cole (1997) for French children when interpolated to the mid-age interval. This confirms the proposition by Cole that correlations between measurements are comparable across populations, even when their growth distances differ.

A flexible method for conditional growth over variable ages relies on a suitable algorithm for the correlation structure. Cole (1995) used fractional polynomials to devise an interpolating algorithm for the Cambridge series, but this approach was not suitable for our data, presumably because of noise in the data. The model chosen by Wade and Ades (1998) guaranteed an increasing correlation with age and decreasing correlation with time interval, but the best approximation was obtained with the models used by Argyle (2002) and Argyle *et al.* (2008). The two-parameter model of Argyle *et al.* (2008) was biased in older children and over longer intervals, but a variant with a multiplier and different constants added to ages t_1 and t_2 provided satisfactory results. Using the same model, parameter estimates (κ, τ_1, τ_2, λ) for the Cambridge series (0.99, 0.02, -0.02, 0.26), and Newcastle series (0.99, 0.02, -0.02, 0.24) are highly comparable with our model for weight.

The validation analysis showed a fair agreement between predicted and observed growth, although the lower 95% prediction limits are somewhat conservative in infants. The sample size per age combination did not allow examination of more stringent limits like -2.5 or -3 SDS, but such criteria could be verified with data from the field, if the measurement practice is sufficiently standardized. Our study included measurements taken over relatively short intervals, which should be considered an absolute minimum when the measurement error is not too large (Himes, 1999). In practice, conditional standards should not be used over shorter intervals as this implies unwarranted extrapolation.

Normal ranges for conditional growth are relatively wide, especially when the correlation is low, which in turn explains the large referral rates when using deflection on a distance chart as an ad hoc criterion (van Buuren *et al.*, 2004). When the correlation is 0.95, a conditional 95% range covers approximately two centile band widths, but this increases to 4 centile band widths if the correlation is only 0.7. When more than two measurements are available, Cole (1995) suggests that the pair with the longest interval is the most reliable, but the actual choice may depend on the situation. Argyle *et al.* (2008) proved that their two parameter model has the Markov property, which implies that only the most recent measurement is pertinent. This property does however not hold for the model variant we selected. Healy *et al.* (1988) proposed a two-step procedure involving the probability (or percentile) of two velocities. Although growth measurements are highly correlated, this is not true for subsequent velocities (Healy *et al.*, 1988; Bailey, 1994; Borislavsky, 1988). Because they are uncorrelated, the probabilities of falling below a threshold may be multiplied to approximate the total probability to fall below the threshold on both occasions. For instance, the probability of falling below the 3th percentile on two subsequent occasions is 0.03 x 0.03 or less than 0.1%. The use of similar cut-off criteria may reduce the number of unwarranted referrals to specialized care, without depending on extreme values of the distribution like ±3 SD.

Drop-out and missing data are a major concern in most longitudinal studies. Likewise, a large number of contacts in our study are missing. However, we believe that the risk for selective drop-out and selection bias is small because the social and demographic profile of the longitudinal subsample compares well with the reference sample, and the majority of children were still involved after the age of 2 years. A major advantage of our study is the large number of individual children involved, and the high density of contacts, up to the age of 3 years. Unfortunately, there was little overlap between the longitudinal and increment samples. This is a design flaw related to different recruitment channels for children below and above 3 years of age. The study setting also did not allow for repeated measurements during the same contact, although such measurement pairs would be a welcome addition to the data set. Ideally, our data should be validated externally, preferably in the field. However, given the good agreement with the Cambridge

and Newcastle correlations, our study can also be considered as a confirmation of the analysis published by Cole (1995).

In summary, conditional growth SDS is a valid, flexible, and possibly population independent approach to assess the dynamics of growth. It depends on the availability of the correlation between lengths, weights or head circumferences at both time points. We estimated the correlation structure in a longitudinal study with a large number of participants and a high density of scheduled contacts, and summarized the correlation structure with a simple algorithm.

REFERENCES

Argyle J (2002) Statistical analysis of child growth data. Doctoral thesis, Durham University (http://etheses.dur.ac.uk/4113/)

Argyle J, Seheult AH, Wooff DA. (2008). Correlation models for monitoring child growth. Statistics in medicine, 27(6), 888–904

Bailey BJ. (1994). Monitoring the heights of prepubertal children. Annals of human biology, 21(1), 1–11

Berkey CS, Reed RB, Valadian I. (1983). Longitudinal growth standards for preschool children. Annals of human biology, 10(1), 57–67

Borysławski K. (1988). Structure of monthly increments of length, weight and head circumference in the first year: a pure longitudinal study of 200 Wrocław infants. Annals of human biology, 15(3), 205–12

Cameron N. (1980). Conditional standards for growth in height of British children from 5.0 to 15.99 years of age. Annals of human biology, 7(4), 331–7

Chrzastek-Spruch H, Hauspie R, Kozlowska M. (1994) Assessment of child physical development up to the 3rd year of age on the basis of data from Lublin longitudinal growth study. Humanbiologia Budapestinensis 25; 307–314

Cole TJ. (1995) Conditional reference charts to assess weight gain in British infants, Archives of Disease in Childhood, 73,8–16

Cole TJ. (1997) Growth monitoring with the British 1990 growth reference, Archives of Disease in Childhood, 76,1–3

D'Agostino RB, Belanger A, D'Agostino Jr RB. (1990) A suggestion for using powerful and informative tests of normality. The American Statistician Vol. 44, No. 4 (Nov., 1990), pp. 316–321

Goldstein H. (1981) The effects of age grouping on the estimate of a correlation coefficient. Annals of Human Biology, 8(2), 181–183

Healy MJR, Yang M, Tanner JM, Zumrawi FY (1988). The Use of Short-Term Increments in Length to Monitor Growth in Infancy. In J. C. Waterlow (Ed.), Linear Growth Retardation in Less Developed Countries (Vol. 14, pp. 41–55). Nestle Nutrition Workshop Series, Vol. 14. Nestec Ltd., Vevey/Raven Press, Ltd., New York.

Healy MJR. (1974). Notes on the statistics of growth standards. Annals of human biology, 1(1), 41–6

Himes JH. (1999) Minimum time intervals for serial measurements of growth in recumbent length or stature of individual children. Acta Paediatrica, 88(2):120–5

Roelants M, Hauspie R, Hoppenbrouwers K. (2009). References for growth and pubertal development from birth to 21 years in Flanders, Belgium. Annals of Human Biology, 36(6): 680–694

Van Buuren S, Bonnemaijer-Kerckhoffs DJA, Grote FK, Wit J M, Verkerk PH. (2004) Many referrals under Dutch short stature guidelines. Archives of Disease in Childhood, 89(4), 351–352

van't Hof MA., Haschke F, Darvay S., Euro-Growth Study Group. (2000). Euro-Growth References on Increments in Length , Weight , and Head and Arm Circumferences During the First 3 Years of Life. Journal of Pediatric Gastroenterology and Nutrition, 31(Supplement 1), S39–S47

Wade AM, Ades AE. (1998). Incorporating correlations between measurements into the estimation of age-related reference ranges. Statistics in medicine, 17(17), 1989–2002

WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards: Growth velocity based on weight, length and head circumference: Methods and development. Geneva: World Health Organization, 2009 (242 pages).

Wright CM, Matthews JN, Waterston A, Aynsley-Green A. (1994). What is a normal rate of weight gain in infancy? Acta paediatrica, 83(4), 351–6

Table 1: Correlation matrix of length (≤ 2 years of age) and height (> 2 years of age) SDS over variable intervals from birth to 36 months (below diagonal). Correlations are based on all available pairs[1] (corresponding numbers above diagonal).

age	birth	1	2	3	4	5	6	7	8	9	10	11	12	15	18	21	24	27	30	33	36
birth		709	724	672	685	651	651	647	576	551	566	511	538	505	461	426	458	311	323	246	250
1	.73		631	582	587	560	542	541	494	466	482	440	461	437	400	368	392	277	281	221	224
2	.68	.85		616	621	570	565	559	501	482	491	442	460	438	401	373	393	276	285	218	231
3	.67	.82	.87		602	546	549	537	479	459	467	423	447	413	383	352	375	272	277	215	222
4	.64	.78	.83	.86		564	552	546	481	464	475	439	457	420	384	361	376	270	276	214	219
5	.57	.71	.79	.84	.87		557	546	494	469	472	443	460	427	394	365	388	280	294	224	222
6	.57	.71	.76	.82	.85	.90		552	497	468	480	449	465	431	399	367	387	277	296	218	222
7	.55	.65	.73	.79	.82	.88	.91		501	474	482	436	460	424	390	365	381	263	278	210	216
8	.52	.64	.71	.77	.82	.86	.89	.91		431	430	403	424	394	361	342	353	251	258	200	199
9	.51	.63	.71	.75	.78	.84	.86	.89	.92		450	413	437	381	354	334	340	237	243	194	193
10	.49	.62	.67	.74	.77	.82	.86	.87	.90	.93		434	438	398	361	335	357	242	254	194	203
11	.52	.61	.69	.72	.76	.82	.86	.86	.90	.92	.93		412	364	339	321	333	243	252	203	201
12	.44	.58	.65	.69	.74	.79	.82	.84	.88	.90	.91	.93		384	356	325	346	243	251	184	191
15	.47	.58	.66	.70	.71	.77	.80	.81	.85	.88	.89	.89	.91		408	364	389	268	272	213	221
18	.43	.54	.61	.66	.67	.74	.76	.77	.81	.85	.83	.87	.89	.90		369	369	256	262	209	209
21	.38	.54	.58	.64	.66	.73	.74	.74	.81	.82	.81	.85	.85	.87	.91		358	241	242	196	197
24	.40	.54	.55	.62	.62	.70	.74	.71	.80	.80	.81	.82	.82	.85	.89	.91		266	275	206	202
27	.35	.49	.50	.58	.58	.64	.66	.66	.77	.75	.77	.80	.80	.84	.86	.88	.91		257	196	197
30	.32	.48	.47	.57	.58	.6	.65	.66	.74	.73	.74	.75	.76	.83	.86	.86	.91	.92		205	203
33	.34	.52	.51	.56	.57	.62	.67	.64	.75	.74	.76	.79	.78	.82	.84	.87	.89	.91	.93		190
36	.32	.42	.49	.53	.56	.60	.68	.62	.71	.71	.70	.76	.76	.79	.82	.86	.88	.89	.90	.92	
n[2]	856	715	728	678	691	655	654	652	581	554	570	514	542	510	464	428	461	313	326	248	251
mean[2]	-0.09	0.08	0.08	0.00	0.05	0.01	0.00	0.02	-0.02	-0.04	0.0	-0.02	0.00	-0.03	-0.03	0.13	0.00	-0.11	-0.03	0.04	0.04
sd[2]	1.03	0.97	0.98	0.99	0.97	0.95	0.99	0.98	0.99	0.99	0.98	1.02	1.01	1.01	1.00	1.00	0.99	0.94	0.97	1.01	0.96

[1] The total number of pairs is 79734; [2] marginal number of measurements (n), mean and SD.

Table 2: Correlation matrix of weight SDS over variable intervals from birth to 36 months (below diagonal). Correlations are based on all available pairs[1] (corresponding numbers above diagonal).

age	birth	1	2	3	4	5	6	7	8	9	10	11	12	15	18	21	24	27	30	33	36
birth		712	731	681	688	655	657	659	584	554	569	517	545	508	465	428	460	316	326	246	252
1	.77		635	586	586	561	543	550	498	467	483	443	461	437	403	369	393	280	282	220	225
2	.65	.87		625	624	575	571	570	508	488	493	447	465	441	406	375	395	281	286	218	232
3	.55	.76	.92		606	551	554	545	485	461	468	428	452	417	388	356	382	276	282	216	224
4	.52	.69	.84	.94		566	555	554	485	465	476	443	460	422	387	364	377	272	276	214	221
5	.45	.62	.77	.88	.94		560	556	497	472	477	446	465	429	398	367	390	282	296	224	223
6	.48	.60	.75	.85	.91	.95		562	505	470	484	455	473	435	404	369	389	280	298	218	223
7	.45	.55	.72	.84	.89	.93	.97		513	482	491	447	472	431	400	371	388	268	283	214	217
8	.45	.53	.70	.80	.85	.89	.93	.97		437	435	408	431	398	366	347	356	254	260	201	201
9	.44	.50	.66	.78	.82	.86	.91	.94	.96		454	415	442	382	356	334	339	240	245	193	193
10	.40	.47	.63	.73	.80	.83	.88	.91	.94	.97		438	443	402	367	338	360	247	257	196	205
11	.45	.47	.63	.72	.77	.81	.86	.90	.92	.96	.97		421	369	344	324	334	248	256	204	202
12	.41	.46	.61	.68	.75	.78	.84	.88	.90	.94	.96	.97		387	359	329	348	249	255	185	192
15	.45	.44	.58	.67	.72	.72	.79	.83	.85	.90	.91	.92	.94		412	365	391	271	275	214	223
18	.46	.45	.56	.62	.68	.69	.75	.78	.81	.86	.86	.88	.90	.95		372	371	263	267	211	212
21	.48	.48	.57	.62	.67	.68	.74	.75	.78	.81	.82	.83	.86	.91	.95		360	246	245	196	197
24	.43	.46	.57	.63	.67	.66	.72	.76	.79	.80	.80	.83	.85	.90	.94	.96		268	277	205	202
27	.44	.40	.50	.58	.60	.64	.70	.72	.73	.78	.77	.78	.82	.87	.92	.94	.96		262	200	201
30	.40	.35	.46	.54	.58	.60	.65	.67	.71	.75	.74	.77	.79	.86	.91	.93	.95	.96		206	203
33	.45	.43	.55	.59	.62	.63	.65	.64	.67	.71	.72	.73	.77	.84	.88	.92	.94	.94	.96		190
36	.39	.32	.47	.54	.56	.59	.64	.63	.65	.72	.70	.75	.77	.83	.86	.90	.93	.92	.94	.96	
n[2]	859	715	734	684	691	657	658	662	587	556	572	518	547	511	467	430	462	317	328	247	252
mean[2]	-0.05	0.03	0.11	0.03	0.03	0.00	-0.04	-0.01	-0.06	-0.03	0.02	-0.03	0.01	0.00	-0.03	0.06	0.00	-0.06	-0.03	-0.04	0.00
sd[2]	1.04	0.96	0.92	0.95	0.97	0.96	0.98	1.02	0.98	1.02	0.99	1.00	1.00	1.01	1.00	1.02	1.01	0.98	1.00	0.99	0.98

[1] The total number of pairs is 80478; [2] marginal number of measurements (n), mean and SD.

Table 3: Correlation matrix of head circumference SDS over variable intervals from birth to 36 months (below diagonal). Correlations are based on all available pairs[1] (corresponding numbers above diagonal).

age	birth	1	2	3	4	5	6	7	8	9	10	11	12	15	18	21	24	27	30	33	36
birth		588	613	575	574	547	543	548	487	461	472	426	449	413	381	357	375	270	279	212	223
1	.67		627	575	573	544	520	532	484	448	466	430	446	412	389	354	375	270	278	214	219
2	.63	.84		615	611	561	550	555	491	472	479	440	448	417	390	359	378	271	281	210	227
3	.56	.78	.87		593	536	533	526	469	446	456	416	435	395	369	339	361	266	277	210	218
4	.57	.78	.85	.92		547	532	535	469	448	457	428	443	397	368	346	357	261	271	207	213
5	.55	.71	.82	.88	.92		536	535	479	455	452	425	444	407	376	348	366	270	289	219	217
6	.54	.70	.81	.87	.90	.93		534	477	445	461	431	444	405	382	344	364	266	284	207	214
7	.52	.67	.78	.83	.86	.89	.92		494	460	467	430	451	408	383	354	369	258	275	207	211
8	.51	.68	.80	.84	.87	.91	.92	.93		414	414	392	409	373	348	325	335	246	252	192	196
9	.53	.68	.78	.83	.85	.89	.92	.93	.96		430	398	420	360	336	315	323	227	237	184	186
10	.52	.65	.77	.81	.84	.88	.91	.92	.94	.95		417	422	375	345	322	340	235	250	191	201
11	.50	.66	.77	.81	.85	.86	.89	.90	.93	.94	.96		401	337	324	301	310	234	240	190	193
12	.49	.63	.76	.80	.83	.86	.89	.90	.92	.94	.95	.96		359	337	309	327	234	244	175	184
15	.45	.64	.74	.75	.76	.81	.83	.84	.87	.89	.90	.92	.93		381	339	360	249	259	198	210
18	.43	.64	.73	.77	.78	.82	.84	.86	.88	.89	.88	.90	.92	.95		348	346	250	256	199	200
21	.47	.66	.74	.76	.77	.80	.82	.83	.86	.87	.87	.89	.89	.93	.95		337	231	236	186	190
24	.42	.60	.70	.73	.73	.80	.80	.81	.84	.85	.83	.86	.86	.90	.93	.95		251	266	191	190
27	.42	.60	.71	.73	.76	.80	.82	.80	.84	.87	.85	.88	.88	.90	.92	.94	.93		253	190	193
30	.42	.56	.66	.69	.70	.76	.77	.80	.83	.84	.85	.86	.87	.88	.90	.91	.92	.96		200	198
33	.47	.63	.70	.74	.76	.80	.79	.78	.82	.84	.82	.85	.87	.88	.90	.91	.90	.94	.95		183
36	.46	.60	.71	.74	.77	.82	.82	.82	.83	.85	.85	.87	.89	.87	.91	.92	.91	.94	.94	.96	
n^2	719	708	729	677	681	643	637	648	573	541	556	506	533	487	450	413	443	308	324	240	247
mean[2]	-0.09	0.23	0.07	-0.01	0.00	0.01	0.00	0.02	0.00	0.01	0.06	-0.02	0.04	0.03	-0.06	-0.03	-0.02	0.00	0.11	0.20	0.29
sd^2	1.02	0.90	0.87	0.88	0.90	0.89	0.92	0.97	0.96	0.98	0.96	0.95	1.00	0.98	0.98	0.99	0.95	0.96	0.94	0.96	0.96

[1] The total number of pairs is 75670; [2] marginal number of measurements (n), mean and SD.

4.5

Referral rates for serial measurements of growth

Roelants M, Hauspie R, Hoppenbrouwers K.

Submitted 2013

Referral rates for serial measurements of growth

Mathieu Roelants[1,2], Roland Hauspie[1], Karel Hoppenbrouwers[2]

ABSTRACT

The proportion of children with a measurement of length or weight below a given percentile on a growth chart increases when growth is assessed repeatedly. We documented how this increase evolves with the number of measurements in longitudinal data of length and weight from 862 infants. Referral rates of serial measurements can be predicted from the joint distribution of measurements at different ages. This method is shown to be valid and allows prediction of the expected rate of referral in preventive health programs that include repeated testing.

INTRODUCTION

High referral rates are a common problem of preventive health programs that include repeated testing of children. Regular screening of length and weight in infancy and childhood are typical examples where the total proportion of children that would be referred to specialized care largely exceeds the nominal value given by the percentile cut-off (van Buuren *et al*., 2004; Grote *et al*., 2007). The probability that a child gets referred on the basis of a single measurement of length or weight, or alternatively, the proportion of children that will be referred, is given by the percentile cut-off used (e.g. 0.4% for the 0.4th percentile), but may increase up to 10 times this amount by completion of the recommended health care program

[1]Laboratory of Anthropogenetics, Department of Biology, Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050 Brussels, Belgium.
[2]Centre of Youth Health Care, Department of Public Health and Primary Care, KU Leuven - University of Leuven, Kapucijnenvoer 35, 3000 Leuven.
Corresponding author: Mathieu Roelants, Department of Public Health and Primary Care, KU Leuven, Kapucijnenvoer 35, 3000 Leuven. mathieu.roelants@med.kuleuven.be

(van Buuren *et al*., 2004; Grote *et al*., 2007). If we consider two measurements of length or weight at age_1 and age_2, converted to their corresponding standard deviation scores SDS_1 and SDS_2, we can denote the probability that these measurements are below a cut-off for referral *c* as $P(E_1) = P(SDS_1 < c)$, and $P(E_2) = P(SDS_2 < c)$. The probability that a child is referred on both occasions is given by the intersection of both events, $P(E_1 \cap E_1)$. In practice, children are usually referred only once, and the probability that this occurs for two measurements is given by the union of E_1 and E_2: $P(E_1 \cup E_1) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$, otherwise said, the total probability of referral at age_1 and age_2 minus the probability to be referred twice. This expression also corresponds to the total proportion of children that will be referred at least once when they are all measured at age_1 and age_2. This can be extended to include three or more measurements, but the algebra involved becomes cumbersome, for instance $P(E_1 \cup E_1 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_3 \cap E_3)$. However, the opposite condition, that is, the probability that a child will not be referred at all, can be readily obtained from the joint distribution of SD scores at different ages. Because these probabilities add up to 1, we can therefore convert a difficult problem (the probability that one or more measurements are below a cut-off), to a more easy one as $P(SDS_1 < c$ *and*/*or* $SDS_2 < c) = 1 - P(SDS_1 > c$ *and* $SDS_2 > c)$. Given a vector of *k* measurements converted to SD Scores, $\mathbf{SDS} = (SDS_1, SDS_2, \ldots SDS_k)$, where each $SDS_{i,i = 1..k}$ follows a Gaussian distribution with zero mean and unit variance (as they should if a well-matched growth reference is used), then **SDS** is multivariate Gaussian with a mean vector $\boldsymbol{\mu} = \mathbf{0}$, and a covariance matrix $\boldsymbol{\Sigma}$ that is equal to the correlation matrix. From the correlation matrix, we can therefore estimate the total proportion of children that will fall below a given percentile for any combination of ages.

Correlations between serial measurements of weight, length, and head circumference, have been published for infants and young children in the UK and in Belgium (Wright *et al*., 1994; Cole, 1995; Roelants *et al*. 2013). We used these correlations to estimate the total referral rate when growth in children is assessed at consecutive ages. This result is validated by comparing the expected referral rates with the corresponding proportions in longitudinal data of infants and young children. Total referral rates are also estimated for existing preventive health programs.

MATERIAL AND METHODS

Serial measurements of length and weight were obtained in a sample of 862 term born (37-42 weeks) singletons (440 boys and 422 girls), recruited between September 2002 and June 2003 in 50 Well Baby Clinics and 25 Day Care Centres for a three year follow-up with monthly (1 – 12 months) and quarterly (15 – 36 months) contacts. The individual number of measurements ranged from 2 to 20, but most children (78.9%) were measured on 10 or more occasions, and 60.9% were measured at least once after the age of 2 years. The total number of contacts was 11736, and about 90% was within a narrow range of 7 (1 – 12 months) or 14 (15 – 36 months) days from the target age. Size at birth was not used in this analysis. All children were of Belgian origin, and free from known diseases or conditions that could affect growth for the duration of the study. This sample was previously used to estimate correlations in length and weight (Roelants *et al*., 2013). This study was approved by the Ethics committee of the Vrije Universiteit Brussel (VUB), and written informed consent was obtained from a parent or legal guardian of each participating child.

Measurements were taken by trained staff members using a standard measuring technique. Length and weight were converted to SD scores using local growth reference charts (Roelants *et al.*, 2009). In each age group, we calculated the proportion of children below the 5th percentile, as well as the proportion of children below the 5th percentile at any occasion, up to the corresponding age. This is a theoretical referral rate as children are not actually referred. The 5th percentile was chosen for statistical reasons as explained in the text.

The expected proportion of children below a given percentile at any age was predicted from the multivariate cumulative Gaussian distribution. The mean was a zero vector, and the covariance matrix was populated with correlations obtained from published correlation models (Cole, 1995; Argyle *et al*., 2008; Roelants *et al*., 2013), and one on the diagonal. As explained in the previous section we calculated the proportion that was above a threshold at all ages considered. Probabilities of the multivariate Gaussian distribution were calculated with the mvtnorm package in R (Genz and Bretz, 2009).

RESULTS

The proportion of children below the 5th percentile at each visit fluctuates around the expected value of 5%, but the total proportion of children who were below the threshold in the past increased significantly with the number of previous measurements (figure 1). Similar trends are observed for weight (table1). The total proportion is a cumulative measure, and thus expected to increase monotonically. In our data, this rate was variable because of small differences in the composition of each age group. The pattern of increase with the number of measurements, more pronounced for the first measurements, was predicted by the model. A small amount of over-prediction was anticipated because the model assumes full compliance, but some contacts were missing in the data, although this number was relatively low (table 1). The predicted referral rates in table 1 were estimated from the Belgian correlation model (Roelants *et al*., 2013), but estimates based on correlation models from the UK (Cole, 1995, Argyle *et al*., 2008) were comparable (data not shown).

Table 1: Observed and predicted proportion of children below the 5th percentile of length and weight at each visit and in the past in selected age groups.

		Number of visits		Children < P5 at current visit		Children < P5 at any visit ≤ age		
age	n	scheduled	Actual[1]	Percentage	95% CI	%	95%CI	Predicted
Length								
1	755	1	1.0	4.0	(2.7–5.7)	4.0	(2.7–5.7)	5.0
3	758	3	2.9	5.5	(4.0–7.4)	8.5	(6.7–10.7)	8.9
6	698	6	5.5	5.0	(3.5–6.9)	11.1	(8.9–13.7)	11.9
9	602	9	8.1	4.1	(2.7–6.1)	12.4	(9.9–15.3)	13.9
12	622	12	10.7	5.1	(3.6–7.2)	13.2	(10.7–16.2)	15.3
18	510	14	12.5	5.1	(3.4–7.4)	16.0	(13.0–19.5)	16.5
24	513	16	14.0	3.9	(2.4–6.0)	14.7	(11.8–18.2)	17.5
36	272	20	17.9	5.1	(2.9–8.6)	16.4	(12.3–21.5)	19.2
Weight								
1	756	1	1.0	4.4	(3.1–6.1)	4.4	(3.1–6.1)	5.0
3	764	3	2.9	3.5	(2.4–5.2)	6.5	(4.9–8.6)	8.4
6	704	6	5.6	4.5	(3.2–6.4)	9.5	(7.5–12.0)	11.1
9	604	9	8.2	4.8	(3.3–6.9)	11.9	(9.5–14.8)	12.8
12	627	12	10.8	4.9	(3.4–7.0)	13.8	(11.3–16.8)	14.0
18	513	14	12.5	5.7	(3.9–8.1)	16.2	(13.2–19.7)	15.2
24	515	16	14.1	5.2	(3.5–7.6)	16.3	(13.3–19.8)	16.3
36	273	20	17.9	5.5	(3.2–9.1)	18.2	(14.0–23.4)	17.7

[1] mean number of contacts

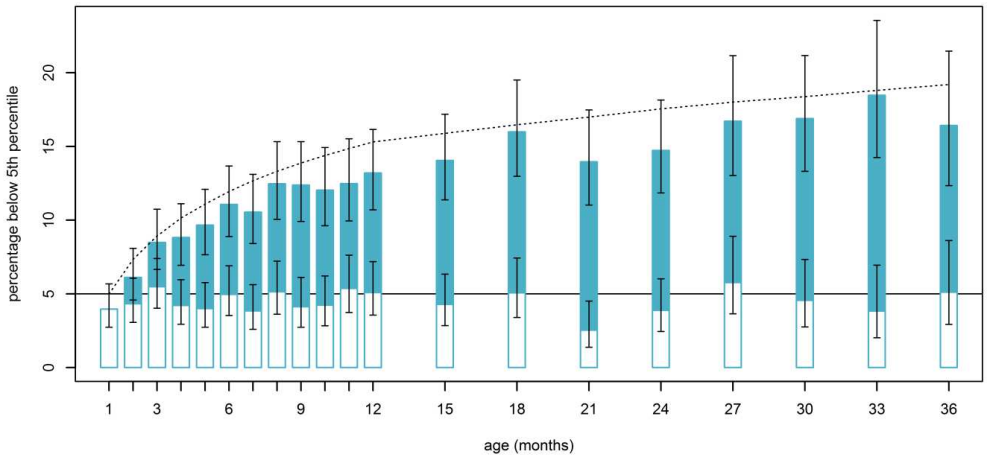


Figure 1: Observed and predicted proportion of children below the 5th percentile of length at each visit (lower stack and horizontal full line), and at any visit (total bar height and dashed curve). Error bars indicate the 95% confidence interval of the observed rate.

DISCUSSION

We have validated a simple method to estimate the expected rate of referral when growth of children is assessed on multiple occasions. Our method assumes that a well-matched growth reference is used, but otherwise depends only on the correlation between measurements at the ages where the children are evaluated.

The increasing proportion of children below the theoretical referral cut-off in our longitudinal data set confirms findings by van Buuren *et al*. (2004), and Grote *et al*. (2007). These studies, both conducted in the Netherlands, found referral rates of 1.8 and 6.2% in children of 3 and 10 years of age respectively, although the expected proportion was only 0.6% at each individual contact. We have additionally demonstrated how this increase occurs with age and the number of measurements. We used the 5th percentile as a theoretical cut-off in our data. This choice was prompted by the need for a sufficiently large number of children below the cut-off (typically 5%) to obtain reasonably precise estimates, and should not be seen as a recommended or even realistic criterion for the assessment of growth. Similar results are obtained with more common cut-offs like -2.5 SD (0.6%) or -3 SD (0.13%), but the rates become more variable, and the confidence intervals

widen. The exact percentile is however not critical, as it was only used as a model to demonstrate how referral rates increase over time, and more importantly, that our correlation-based prediction method is valid.

Referral rates increase with the number of measurements, and this increase is larger for the first few measurements, and larger when the correlation is low. Both properties are responsible for the more pronounced increase that is observed in the data and predicted by the model in early infancy. Correlations are lower in this age period (typically 0.6-0.8 for measurements not too far apart) (Cole, 1995; Wright *et al*., 1994; Roelants *et al*., 2013), when individual growth profiles are more variable (Mei *et al*., 2004), than in childhood (about 0.98 for measurements one year apart) (Bailey, 1994; Cole, 1997, Roelants *et al*., 2013). The rate of increase also depends on the nominal cut-off, and is more pronounced for more extreme percentiles.

The American Academy of Pediatrics recommends the assessment of length and weight at 11 different ages between 1 month and 3 years of age (9 ages up to 24 months for head circumference) (AAP, 2007). Similar health promotion programs are in effect in other countries (Kuo *et al.*, 2006). Our model predicts referral rates of 7 – 9% when using a fixed cut-off of -2 SD (2.3rd percentile), 2 – 3% for -2.5 SD (0.6th percentile) and 0.6 – 0.8% for -3 SD (0.13th percentile). Expected rates can be refined to accommodate different criteria in different age groups (Grote *et al*. 2008, Oostdijk *et al*., 2009). Choosing even lower percentiles as a criterion for referral appears a logical measure to reduce the total referral rate and increase the specificity, but may hamper the detection of growth faltering at each individual contact. An alternative approach that is less dependent on extreme percentiles, is to assess consecutive measurements concomitantly with appropriate tools like velocity charts or conditional growth (Cole, 1995; Roelants *et al*., 2013).

In summary, we have documented how referral rates increase when growth is assessed on multiple occasions, and validated a method to predict referral rates for serial measurements. The results from this study are generalizable, and will apply to other screening programs that include repetitive testing.

REFERENCES

AAP, American Academy of Pediatrics, Committee on Practice and Ambulatory Medicine and Bright Futures Steering Committee. (2007) Recommendations for Preventive Pediatric Health Care. Pediatrics. 120(6):1376

Argyle J, Seheult AH, Wooff DA. (2008). Correlation models for monitoring child growth. Statistics in medicine, 27(6), 888–904

Bailey BJ. (1994). Monitoring the heights of prepubertal children. Annals of human biology, 21(1), 1–11

Cole TJ. (1995) Conditional reference charts to assess weight gain in British infants, Archives of Disease in Childhood, 73,8–16.

Cole TJ. (1997) Growth monitoring with the British 1990 growth reference, Archives of Disease in Childhood, 76,1–3.

Genz A, Bretz F (2009), Computation of Multivariate Normal and t Probabilities. Lecture Notes in Statistics, Vol. 195., Springer-Verlage, Heidelberg.

Grote FK, Oostdijk W, De Muinck Keizer-Schrama SMPF, Dekker FW, Van Dommelen P, Van Buuren S, Lodder-van der Kooij AM, *et al*. (2007). Referral patterns of children with poor growth in primary health care. BMC public health, 7, 77

Grote FK., Van Dommelen P, Oostdijk W, De Muinck Keizer-Schrama SMPF., Verkerk PH, Wit JM, Van Buuren S. (2008). Developing evidence-based guidelines for referral for short stature. Archives of disease in childhood, 93(3), 212–7

Kuo A, Inkelas M, Lotstein D, Samson K, Schor E, Halfon N. (2006) Rethinking well-child care in the United States: an international comparison. Pediatrics. 118(4):1692-702

Mei Z, Grummer-Strawn LM, Thompson D, Dietz WH. (2004). Shifts in percentiles of growth during early childhood: analysis of longitudinal data from the California Child Health and Development Study. Pediatrics, 113(6), e617–27

Oostdijk W, Grote FK, De Muinck Keizer-Schrama SMPF, Wit JM. (2009). Diagnostic approach in children with short stature. Hormone research, 72(4), 206–17

Roelants M, Hoppenbrouwers K, Hauspie R. (2013). Conditional growth and correlations between longitudinal measurement of length, height, weight and head circumference in infancy and childhood. Submitted

Roelants M, Hauspie R, Hoppenbrouwers K. (2009). References for growth and pubertal development from birth to 21 years in Flanders, Belgium. Annals of Human Biology, 36(6): 680–694

Van Buuren S, Bonnemaijer-Kerckhoffs DJA, Grote FK., Wit JM, Verkerk PH. (2004). Many referrals under Dutch short stature guidelines. Archives of Disease in Childhood, 89(4), 351–352

Wright CM, Matthews JN, Waterston A, Aynsley-Green A. (1994). What is a normal rate of weight gain in infancy? Acta paediatrica (Oslo, Norway : 1992), 83(4), 351–6

5. Discussion

The assessment of growth in children depends on the use of a suitable, matched, growth reference, and appropriate criteria for referral. We have addressed several aspects of the assessment of *normal growth* in a single mixed longitudinal survey of healthy children. This study was conducted according to guidelines of the WHO Expert Committee on Physical Status (WHO, 1995b), and adheres to the recommendation of creating a reference that resembles as far as possible a true standard (WHO, 1986). The sample is sufficiently large, representative of healthy children, well documented and selected using reproducible procedures; the study included all anthropometric indicators that are commonly used for routine growth assessment; measurements were taken with standardized techniques and calibrated equipment; and reference curves were estimated using modern statistical methods that allow conversion of measurements to SD-scores and validation of the curve fit (Cole, 1988; Cole and Green, 1992). Study results were disseminated as growth charts (see appendix) and documented in an elaborate report (Roelants, Hauspie and Hoppenbrouwers, 2008), but an overview of the study is given in chapter 3 and a summary of key findings in section 4.1. In this discussion we will focus on the results that are covered by sections 4.1 – 4.5: cross-sectional references for growth distance and pubertal development; the appropriateness of the WHO universal standards; references for the dynamics of growth; and total referral rates of growth monitoring programs.

5.1 A cross-sectional growth distance reference

Our first objective was to estimate a representative growth reference up to current methodological standards. While not explicitly formulated as a research question or hypothesis, the justification to conduct a complete growth reference study (and the sample sizes involved) implies at least two: ongoing secular trends, and differences in growth from comparable populations.

5.1.1 Secular trends

Positive secular trends in length and height have been observed in most affluent populations (Hauspie *et al.* 1996; 1997), and are generally attributed to improved overall living standards (Tanner, 1987). Data on height and weight in the Belgian population are available from 1830, and subsequent growth studies have shown a consistent positive trend in attained height and advancement in the tempo of growth (Hauspie *et al.*, 1997). During the last decades, changes in tempo became however less prominent (Vercauteren et al., 1998), and the secular increase over the past 30-40 years that was observed in our study is smaller compared to estimates from the first half of the 20th century (Hauspie *et al.*, 1997). A systematic slowdown between 1955 and 1997 is well documented in the Netherlands (Fredriks *et al.*, 2000), where the secular increase has now completely ceased (Schönbeck *et al.*, 2013). Similar trends have been observed in other north European countries, for instance Germany (Schaffrath Rosario *et al.*, 2011), but it is not a general trend (Júlíusson *et al.*, 2009). In the absence of large societal changes a current maximum population average is apparently located between 180 and 185 cm in young adult males and between 165 and 170 cm in females (Schönbeck *et al.*, 2013; Schaffrath Rosario *et al.*, 2011).

Between 1830 and 1980, secular trends in length or height were largely realized by the age of 3 (Vercauteren et al., 1998) but between 1960 and our study in 2004, the secular trend in height of children under five has completely ceased, although it is still present in older children and adolescents. The same phenomenon was observed in The Netherlands (Fredriks *et al.*, 2000) and in Norway (Júlíusson *et al.*, 2009). One could speculate that this is due to reaching a maximum or optimum in growth of infants and young children. However, we also observed that there are still marked social differences in the length of infants and young children, and this with the same magnitude as in older children (data not shown). If the secular trend depends only on the level of affluence, this would indicate that there is still some room for a further increase in the Belgian population, before and after the age of five years.

Interestingly, our data also showed some evidence of positive effects of social mobility. A pilot study in university students showed no difference according to the

level of parental education (Demaeseleer and Hauspie, 2001), but a social gradient was still present in those attending secondary education. Although this remains to be studied in detail, it urged us to expand our sample with data from occupational medicine (see chapter 3).

5.1.2 Population differences

Recent reviews of European growth studies confirmed the existence (and persistence) of a north to south gradient of decreasing height (Bonthuis *et al.*, 2012; van Buuren *et al.*, 2012). Apart from growth at adolescence, when children may be temporarily taller or smaller due to a differential timing of the pubertal spurt, these differences in height are usually the largest in young adults, but may be much smaller in childhood (Haas and Campirano, 2006). To discriminate secular trends from true population differences, section 4.1 only includes a comparison with growth references from The Netherlands (Fredriks *et al.*, 2000), UK (Freeman *et al.*, 1995), and US (Kuczmarski *et al.*, 2002). Compared to the UK and the US, Belgian children are taller (and less overweight), but data collection in both studies preceded our study by more than 10 years, and the US sample includes all ethnicities (Kuczmarski *et al.*, 2002). More recent data on height from Germany (Schaffrath Rosario *et al.*, 2011) and Norway (Júlíusson *et al.*, 2013) are highly comparable. Flemish children are however considerably smaller than the Dutch, who are at present day the tallest population in the world (Fredriks *et al.*, 2000; Schönbeck *et al.*, 2013). This is remarkable given the geographical and cultural proximity (we share a border and common language), and an open border policy for several decades. Although the height of the Dutch has stabilized over the past decade (Schönbeck *et al.*, 2013), it would take 20 – 30 years to catch up if the current secular trend can be maintained in the Flemish population. A north to south gradient in the Netherlands (Fredriks *et al.*, 2000; Schönbeck *et al.*, 2013), could only partially explain this difference. Similar gradients in Flanders are very small (the north to south distance is only 100 km.), but historical data showed that children living in the south of Belgium (Wallonia) were on average 1 – 2 centimeter smaller (Beckers and Pleysier, 1983).

Our reference sample only includes children of Belgian origin, which is a common modus operandi in European growth studies (Freeman *et al.*, 1995; Fredriks *et al.*,

1998; Schönbeck *et al.*, 2013; Júlíusson *et al.*, 2013), but not in the US (Kuczmarski *et al.*, 2002) and in a recent study in Germany (Schaffrath Rosario *et al.*, 2011). This may seem impractical as it invalidates the reference curve for more than 10 percent of the target population, but a reference curve based on all children would, in our opinion, be valid for neither subgroup. Compared to the reference population, the height of immigrant children decreases gradually from the age of five onwards to a difference of -1 SD near adult height (data not shown). Height of Flemish children of Moroccan and Turkish origin is however highly comparable to corresponding reference curves that were estimated in these ethnicities in the Netherlands (Fredriks *et al.*, 2003; Fredriks *et al.*, 2004).

5.1.3 Pubertal development

Our references for menarche and pubertal development are different from those of anthropometric distance because they show the distribution of ages at discrete events, rather than the distribution of stages for a given age. Although pubertal development is an indicator of tempo, these references are cross-sectional, i.e. one can derive the mean duration of transition through subsequent stages, but not the normal variation. There have been concerns about an increased prevalence of precocious puberty in the US (Parent *et al.*, 2003) and more recently in Europe (for review, see Sørensen *et al.*, 2012). Due to the lack of suitable historical data, this could not be examined by us, but our data are in close agreement with those from The Netherlands in 1996/97, who did not observe any changes over the previous decades (Mul *et al.*, 2001). However, the most recent Dutch reference study showed a small but significant advancement of mean age at menarche with approximately one month (Talma *et al.*, 2013). In our population, we could confirm the observation by Vercauteren and Susanne (1985) that mean age at menarche has stabilized, at least up to 2004, and that both menarche and the development of secondary sex characteristics are consistent with those observed in northern Europe (Parent *et al.*, 2003).

The distribution of ages at the various stages of pubertal development was found skewed, an observation that was equally made (distributions were even highly comparable) in the Netherlands (Mul *et al.*, 2001), but is not a general finding (Aksglaede et al., 2009; Sørensen *et al.*, 2010). It is yet unclear if this has a

methodological origin, because the studies in Copenhagen (Aksglaede *et al.*, 2009; Sørensen *et al.*, 2010) used a small team of observers, while both we and Mul *et al.* (2001) depended on a large number of observers. These studies are otherwise comparable, and based on a physical examination, including palpation of the budding stage at B2.

5.1.4 Criteria for overweight

The worldwide increase in overweight is problematic for the choice of BMI percentiles that delimit overweight, obesity and thinness (Cole *et al.*, 2000). Because the prevalence of overweight and obesity in our reference sample is unknown, we selected the corresponding cut-off percentiles according to the principle outlined by the IOTF (Bellizzi and Dietz, 1999; Cole and Roede, 1999, Cole *et al.*, 2000). Using a percentile threshold, objectively chosen or not, implies a constant prevalence across age groups. Compared to the IOTF, the overall prevalence of overweight and obesity in our reference sample is 13.0% and 2.8% respectively, but it is lower in infancy and early childhood, and higher in primary schoolchildren. This is a general phenomenon that was observed in our population (Roelants and Hoppenbrouwers, 2001), and in other European countries, for instance, the Netherlands (Schönbeck et al., 2011) and Norway (Júlíusson et al., 2010). However, it has been shown that from the age of two years onwards, children in the top deciles of the BMI distribution have already a high risk to be in the top deciles at a later age (De Kroon *et al.*, 2010). Therefore our thresholds in younger children probably indicate an increased risk to become overweight or obese rather than actually being overweight or obese.

The IOTF BMI criteria were originally proposed as an international criterion for epidemiological purposes, and not as a universal criterion (Cole *et al.*, 2000), but they were nevertheless added as clinical thresholds to several national reference curves, for instance in The Netherlands (van Buuren, 2004) and in Norway (Júlíusson *et al.*, 2009). We preferred to use an analogue, rather than the precise cut-offs, for the reasons outlined above, and also because the IOTF reference was only defined as a discrete cut-off for overweight and obesity when our reference charts were disseminated in 2004. Criteria for underweight, and a completely defined international BMI curve appeared later (Cole *et al.*, 2007; Cole and

Lobstein, 2012). In The Netherlands they combined IOTF cut-offs with local BMI curves, and estimated cut-offs for thinness from the local BMI reference (van Buuren, 2004). We believe that this creates confusion in a clinical context, because the BMI SDS and weight status are not consistent. However, because the precise meaning of being above our percentile cut-offs may differ according to age, we advise to use the IOTF criteria for epidemiological studies and for the estimation of prevalence and trends of actual overweight.

5.2 References or universal standards

A close look at the design and inclusion criteria of the Multicentre Growth Reference Study (MGRS) (WHO, 2006b) shows that the universal standards have a different meaning in developing countries, where they are elite standards (*viz.* Goldstein and Tanner, 1980), and in the western world, where they can be considered as a reference for breastfed children. This allowed us to validate the "universal" aspect of the standards, the result of which pointed to persistent population differences and a possible methodological flaw.

Validating a standard is in principle as difficult as defining one, because markers of optimal growth are not readily available (Goldstein and Tanner, 1980), and standards do not have to match a particular population. However, the WHO used the very reasonable but circumstantial model that growth in breastfed children reflects true optimal growth (we can safely ignore social differences as only few children from the US or Norway were excluded for this reason; WHO, 2006b). Infant feeding was well documented in our data and in the Bergen Growth Study (Júlíusson *et al.*, 2013), which allowed us to compare growth in infants and young children, breastfed according to MGRS feeding recommendations, with the standards. Our study showed a reasonable match of length by age in older infants (somewhat less in the Norwegian sample), but cross-sectional patterns of weight and BMI in breastfed children were more alike those observed in the total reference population. It also appears that a large size at birth followed by a downward trend sustained for several months, is typical feature of growth in length and weight in western populations, including the Norwegian and US infants in the MGRS (sections 4.2 and 4.1; Wright *et al.*, 2008; Mei and Grummer-Strawn, 2011; WHO,

2006c, de Onis *et al.*, 2007a). A large size at birth in these groups coincided with a larger maternal size (WHO, 2006b), but it has been suggested (Binns and Lee, 2006; Binns *et al.*, 2008), and was confirmed in at least one study (van Buuren, 2010), that the downward trend shortly after birth may have a methodological reason (see section 1.5). Our study was based on cross-sectional data, and only considered feeding up to the age of the measurement. This approach is correct, because we validated a cross-sectional curve, and because growth should reflect the past, not the future. It is however different from the WHO/MGRS data set which only included infants that were compliant for the whole two years.

Our findings for head circumference in Belgian and Norwegian children are entirely different but equally important. Head circumference by age was (i) comparable in breastfed and formula fed children; (ii) parallel to the median of the WHO standards, but (iii) consistently 0.5 SD above. Upon publication, these findings have been confirmed in the UK (Wright *et al.*, 2011), and in the US (Daymont *et al.*, 2010), and point to clear population differences.

Based on our comparison of growth in breastfed children from two European countries, we conclude that the proposition of universality in growth remains unproven. Given that universality is a major focal point of the WHO standards, we find it odd that two elaborate reports (530 pages in total) and a complete supplement of Acta Paediatrica (Volume 95, Supplement 450, April 2006) contain no data on site differences beyond those of length and height. Given that feeding recommendations are another major focal point of the MGRS, we find it equally odd that the effect of sample selection was not discussed in these reports. A majority of infants, including those that are breastfed, will exhibit downward tracking relative to the standards, the avoidance of which was one of the main reasons to initiate the MGRS (De Onis *et al.*, 1997). However, we also observed the typical patterns of growth in breastfed versus formula-fed children, and consider local reference charts based on a representative sample of exclusively breastfed children a more valuable alternative for classical infant references.

5.3 Monitoring changes in growth over time

To realize the third objective of this study, we pursued the concept of conditional growth SDS (Cole, 1994b; 1995), and confirmed that it is a flexible and valid alternative for growth velocity or increment charts (WHO, 2007; van 't Hof *et al.*, 2000; Chrzastek-Spruch *et al.*, 1994). The required correlation structure of weight in infancy and height in childhood corresponded to similar data from the UK (Wright *et al.*, 1994; Cole, 1995) and France (Cole, 1994; 1997), and thus confirms the proposition that the correlation between measurements is comparable across populations, despite differences in size may exist (Cole, 1997). Correlations of length and head circumference in infancy were not previously available. Correlation models were based on the model by Argyle *et al.* (2008), with minor modifications.

Under the assumption of a constant variance, regular (unconditional) growth velocities or increments have been shown to be less efficient (Cameron, 1980), and biased, because they do not correct for regression to the mean (Cole, 1993). As a consequence, velocities expressed as unconditional change on the SD scale (sometimes termed "delta SDS" or "deflection"; van Buuren *et al.*, 2004a) should be avoided, except in age groups where the bias is small because the correlation is high (Cole, 1997). However, on the measurement scale, the variance increases with age, and velocities may actually be unbiased and equally efficient as conditional growth (see Cameron *et al.*, 2005 for a similar analysis of regression to the mean), but this proposition remains to be validated.

References for growth velocity, conditional or not, are subject to the same limitations as those of distance. They are based on healthy children and show normal growth, but give no indication about the sensitivity to detect true growth faltering. These references are also designed to assess a single velocity at a time. However, it has been shown that that the correlation between consecutive velocities is small (Healy *et al.*, 1988b; Bailey, 1994), in which case the probabilities of occurrence (the percentiles) may be multiplied (Healy *et al.*, 1988b; Cole, 1997). For instance, the probability of falling twice below the 5^{th} percentile is only 0.25%. The concomitant assessment of serial velocities could also be a more sensitive indicator for disorders that result in a small but consistent deviation from the

reference curve, as is observed in, for instance, Turner syndrome (Ranke *et al.*, 1983).

The WHO did not create conditional reference charts because they felt these to be impractical in the field (WHO, 2009). The proper use of velocity charts is however subject to comparable limitations, unless all children are measured with the same (exact) interval. However, we agree that in order to fully exploit conditional growth, an electronic growth chart is required. Apart from the advantages outlined above, conditional SDS is also attractive because of its close link with changes in position on the distance chart.

5.4 Referral rates of growth monitoring programs

High referral rates are a common concern, particularly because the "diagnostic" or screening value of growth monitoring is ill-defined (Hall, 2000; Garner *et al.*, 2000). The LMS method or equivalent distribution based methods can generate more extreme percentiles, and thus increase the specificity (Cole, 1988). However, due to the problem of multiplicity, the total rate of referral will still exceed the value that corresponds with the nominal cut-off percentile (van Buuren *et al.*, 2004a; Grote *et al.*, 2007). In section 4.5 we presented a systematic analysis of referral rates, and provided and validated a method to estimate the total rate of referral in growth monitoring programs that include multiple measurements. This method is based on the same correlation models that were developed for conditional growth reference (section 4.4).

The method we have developed allows estimating the referral rate, but does not provide a solution to reduce it. A closer look at this model shows that the increase is nonlinear and behaves like a saturation function. Reducing the number of assessments from, say, 10 to 8 has therefore little effect, as most of the inflation is already realized during the first few contacts. A second property is that the referral rate increases more when the correlation is low. This pleads against the use of a single growth velocity for referral (Hall, 2000), and provides another argument to further investigate the concomitant assessment of velocities discussed in the previous section.

5.5 Some final remarks

Although we consider our growth study fairly complete, there are some limitations. A major weakness is the lack of data on the technical error of measurement (TEM). Measurement error is well documented in the literature (Ulijaszek and Kerr, 1999; WHO, 2006d), but should be part of each study as a measure of data quality. We did not include this for practical reasons. Because of the large number of staff members involved in the study, we invested in calibrating equipment and standardizing measurement procedures, rather than to estimate the TEM from a limited (and possibly more motivated) number of persons. We assume that the TEM will be in between that of strict research setting (about 0.35 cm for length or height) and that of routine practice (> 0.5 cm; Ulijaszek and Kerr, 1999; WHO, 2006d), because staff members were obviously aware that the data were being processed for the study.

Another limitation is the lack of velocity data around the age of 3 years, a period of transition from quickly decreasing and variable velocities to a relative stable period of growth (Molinari and Gasser, 2004). The velocity data also cover a wide variety of ages in infancy, but are limited to yearly increments in childhood. Larger intervals, preferably multiples of a year, would be a welcome addition that reflects more closely the school health practice in Flanders, where children are examined every 2 or 3 years (Van Hoek and Hoppenbrouwers, 2005). Finally, for practical reasons, there was no pubertal assessment when the second measurement of the yearly increment data was taken, which limits the maximum age for growth velocities in childhood to approximately 9 – 10 years.

This brings us to a limitation of growth assessment in general and cross-sectional growth references in particular: the assessment of growth at adolescence. Our references allow a cross-sectional assessment of the timing of pubertal development, and give an overview of the (tempo unconditional) distance and velocity, but these tools are clearly too limited for a full assessment of growth during adolescence (Tanner *et al.*, 1966b).

We approached growth assessment as a form of screening in which growth charts serve as a reference, but growth is not a diagnostic test. The focus of our study was

"normal growth", arbitrarily defined as growth that is observed in the majority of healthy children. In this respect we have only addressed the specificity or false positive rate of growth assessment. Also, we only considered data on growth of the child itself, and did not include extraneous parameters like parental height, which may well increase the sensitivity of a single measurement for a given specificity (van Buuren *et al.*, 2004b; Grote *et al.*, 2008; Hall *et al.*, 2008). Data on the overall sensitivity and outcome of growth monitoring programs is still a knowledge gap. Maintaining high standards of measurement and systematic registration of measurements and interventions in the field could provide these much needed data. A proper registration would also be useful to monitor secular trends in height and (over)weight in the population.

5.6 Future perspectives

Throughout the discussion we have hinted at several possibilities for further research, and knowledge gaps regarding the assessment of growth. More specifically for the assessment of longitudinal measurements, we have proposed conditional growth as an alternative for growth velocity or increments over a single measurement interval. However, in practice, there are usually many more than two measurements, and it is yet unclear over which of these intervals growth is best evaluated. Longer intervals are less prone to measurement errors, but may hamper quick action if needed, while the opposite is true for shorter intervals. Concomitant assessment of several velocities (conditional or not) may provide a robust solution for this problem. Another potentially useful, and, to our knowledge, yet unexplored aspect is the inclusion of parental height in the conditional growth model, to discriminate children that grow towards or away from their target height.

6. Summary and conclusions

The assessment of growth, whether that of a single measurement of attained size or the rate of growth between two measurements, depends on the use of an appropriate, well-matched reference, and suitable criteria for referral. In this work, we have developed and validated instruments to give guidance for the assessment of growth distance and growth velocity in a single mixed longitudinal study.

For the assessment of a single measurement, conditional on age, we developed cross-sectional references of length, height, weight, BMI, head circumference, and pubertal development from a representative sample of Flemish children. Sampling and data collection were conducted according to WHO recommendations. Growth distance curves were fitted with the LMS method, a semi-parametric method that allows to estimate more extreme percentiles with reasonable accuracy, and to convert measurements to SD scores and vice versa. Results show an ongoing secular trend in height of 0.8 – 1.2 cm per decade within the Belgian population. Height of Belgian children is roughly comparable with that in other northern European countries, except for the Netherlands, and is more or less consistent with a European north to south gradient of decreasing height. Our references for BMI and Tanner stages of pubertal development are the first for the Belgian population. The BMI curve, and consequently the prevalence of overweight, is largely comparable with that in neighbouring countries. References for pubertal development were estimated from *status presens* data using generalized additive models. Mean age at menarche has not changed over the past 50 years. Menarche and the development of secondary sex characteristics are highly comparable to data from the Netherlands and within ranges that are observed in other northern European populations.

In exclusively breastfed children, length by age is comparable with that of the WHO/MGRS universal growth standards, but growth in weight, BMI, and especially head circumference is more comparable to the local Belgian reference population. These results were validated in a joint analysis of Belgian and Norwegian children, and confirmed trends that were observed in other countries. It is yet unclear how to what extent this is explained by a – in our opinion – possible methodological flaw of the standards. Our findings for head circumference were

later confirmed in the UK and in the US, and indicate persistent population differences in anthropometric characteristics. We feel that these observations, combined with the marked population differences in older children, justify our recommendation that a recent local reference chart is the best tool for growth assessment, but consideration should be given to estimate charts from breastfed children.

To give guidance for the assessment of short-term growth (i.e. the rate of growth between two measurement occasions) we have estimated correlations between measurements and validated conditional growth assessment in our population. Conditional growth SDS is a flexible alternative for velocity, and has a close link with assessing changes in the position on a growth distance chart. Correlations between serial measurements of weight were shown to be comparable with similar data from the UK. Conditional growth assessment is currently limited to a single velocity at a time, but the concomitant assessment of consecutive velocities is a promising approach that deserves further investigation.

The perspective given by our data is limited to growth in healthy children, and is therefore a proxy for the specificity or false positive rate. These rates have been shown to increase when children are frequently assessed. From a systematic analysis of referral patterns we were able to devise and validate a method to estimate the total referral rates in growth monitoring programs. The sensitivity of growth assessment is still largely unknown. High standards of measurement and registration of data could help to fill this knowledge gap, and allow monitoring secular trends in the population.

Altogether, the tools that were developed in our study will help health care providers to assess growth in individual children, and policy makers to design and evaluate growth monitoring programs.

References

This list covers the literature cited in chapters 1 – 3 and 5

Aksglaede L, Sørensen K, Petersen JH, Skakkebaek NE, Juul A. (2009). Recent decline in age at breast development: the Copenhagen Puberty Study. Pediatrics 123: e932–9.

Argyle J, Seheult AH, Wooff DA. (2008). Correlation models for monitoring child growth. Stat Med, 27: 888–904.

Atwood CL, Taube A. (1976). Estimating mean time to reach a milestone, using retrospective data. Biometrics 32: 159–71.

Aulchenko YS, Struchalin MV, Belonogova NM, Axenovich TI, Weedon MN, Hofman A, Uitterlinden AG, et al. (2009). Predicting human height by Victorian and genomic methods. European journal of human genetics : Eur J Hum Genet 17: 1070–5.

Ayatollahi SMT. (2005). Growth velocity standards from longitudinally measured infants of age 0-2 years born in Shiraz, southern Iran. Am J Hum Biol 17: 302–9.

Bailey BJ. (1994). Monitoring the heights of prepubertal children. Ann Hum Biol 21: 1–11.

Bairagi R. (1986). On Components of Variation of Estimated Weight Velocity of Children. J R Stat Soc Ser C Appl Statist 35: 178–82.

Barlow SE, Dietz WH. (1998). Obesity evaluation and treatment: Expert Committee recommendations. Pediatrics 102: e29.

Baumgartner RN, Roche AF, Himes JH. (1986). Incremental growth tables: supplementary to previously published charts. Am J Clin Nutr 43: 711–22.

Bayley N. (1956). Growth curves of height and weight by age for boys and girls, scaled according to physical maturity. J Pediatr, 48: 187–94.

Beckers R, Pleysier R. (1983). [Anthropometric analysis of the 1980-1981 Belgian school population. I. Description of the survey project. II. Initial results]. Archives belges 41(3-4): 89–123.

Bellizzi MC, Dietz WH. (1999). Workshop on childhood obesity: summary of the discussion. Am J Clin Nutr 70: 173S–5S.

Berkey CS, Dockery DW, Wang X, Wypij D, Ferris B. (1993). Longitudinal height velocity standards for U.S. adolescents. Stat Med 12: 403–14.

Berkey CS, Reed RB, Valadian I. (1983). Longitudinal growth standards for preschool children. Ann Hum Biol 10: 57–67.

Binns C, James J, Lee MK. (2008). Why the new WHO growth charts are dangerous to breastfeeding. Breastfeeding review 16(3), 5–7.

Binns C, Lee M. (2007). New growth standards. Lancet, 370: 1542.

Bonthuis M., Van Stralen KJ, Verrina E, Edefonti A, Molchanova E, Hokken-Koelega ACS, Schaefer F, et al. (2012). Use of national and international growth charts for studying height in European children: development of up-to-date European height-for-age charts. PloS one 7: e42506.

Borghi E, De Onis M, Garza C, Van den Broeck J, Frongillo E, Grummer-Strawn, L, Van Buuren S, et al. (2006). Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. Stat Med 25: 247–65.

Borysławski K. (1988). Structure of monthly increments of length, weight and head circumference in the first year: a pure longitudinal study of 200 Wrocław infants. Ann Hum Biol 15: 205–12.

Bowditch HP . (1891). The growth of children studied by Galton's method of percentile grades. 22d Ann Rep Massachusetts State Board of Health, Boston

Brook C. (1983). Earlier recognition of abnormal stature [Letter]. Arch Dis Child 58: 840.

Butler GE, McKie M, Ratcliffe SG. (1990). The cyclical nature of prepubertal growth. Ann Hum Biol 17: 177–98.

Cameron N, Preece MA, Cole TJ. (2005). Catch-up growth or regression to the mean? Recovery from stunting revisited. Am J Hum Biol 17: 412–7.

Cameron N. (1980). Conditional standards for growth in height of British children from 5.0 to 15.99 years of age. Ann Hum Biol 7: 331–7.

Ceriotti F, Hinzmann R, Panteghini M. (2009). Reference intervals: the way forward. Ann Clin Biochem, 46: 8–17.

Chen K, Müller H. (2012). Conditional quantile analysis when covariates are functions , with application to growth data. J R Stat Soc Ser B Stat Methodol, 74: 67–89.

Chrzastek-Spruch H, Hauspie R, Kozlowska M. (1994). Assessment of child physical development up to the 3rd year of age on the basis of data from Lublin longitudinal growth study. Humanbiol Budapest, 25, 307–314.

Cole TJ. (1986). Weight/height^p compared to weight/height^2 for assessing adiposity in childhood: influence of age and bone age on p during puberty. Ann Hum Biol, 13: 433–51.

Cole TJ. (1988). Fitting Smoothed Centile Curves to Reference Data. J R Stat Soc Ser A Stat Soc, 151: 385–418.

Cole TJ. (1993). The use and construction of anthropometric growth reference standards. Nutr Res Rev, 6: 19–50.

Cole TJ. (1994a). Do growth chart centiles need a face lift? BMJ, 308: 641–2.

Cole TJ. (1994b). Growth charts for both cross-sectional and longitudinal data. Stat Med, 13: 2477–92.

Cole TJ. (1995). Conditional reference charts to assess weight gain in British infants. Arch Dis Child, 73: 8–16.

Cole TJ. (1997). Growth monitoring with the British 1990 growth reference. Arch Dis Child, 76: 47–9.

Cole TJ. (1998). Presenting information on growth distance and conditional velocity in one chart: practical issues of chart design. Stat Med 17: 2697–707.

Cole TJ, Bellizzi MC, Flegal KM., Dietz WH. (2000). Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ 320: 1240–3.

Cole TJ, Cortina-Borja M, Sandhu J, Kelly FP, Pan H. (2008). Nonlinear growth generates age changes in the moments of the frequency distribution: the example of height in puberty. Biostatistics 9: 159–71.

Cole TJ, Flegal KM, Nicholls D, Jackson AA. (2007). Body mass index cut offs to define thinness in children and adolescents: international survey. BMJ 335: 194.

Cole TJ, Freeman JV, Preece MA. (1998). British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. Stat Med 17: 407–29.

Cole TJ, Green PJ. (1992). Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med 11: 1305–19.

Cole TJ, Lobstein T. (2012). Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. Pediatr Obes 7: 284–94.

Cole TJ, Roede MJ. (1999). Centiles of body mass index for Dutch children aged 0-20 years in 1980 – a baseline to assess recent trends in obesity. Ann Hum Biol 26: 303–8.

Daymont C, Hwang WT, Feudtner C, Rubin D. (2010). Head-circumference distribution in a large primary care network differs from CDC and WHO curves. Pediatrics 126: e836–42.

De Kroon ML, Renders CM, Van Wouwe JP, Van Buuren S, Hirasing RA. (2010). The Terneuzen birth cohort: BMI changes between 2 and 6 years correlate strongest with adult overweight. PloS one 5: e9155.

De Onis M, Garza C, Habicht JP. (1997). Time for a new growth reference. Pediatrics 100: e8.

De Onis M, Garza C, Onyango AW, Borghi E. (2007a). Comparison of the WHO child growth standards and the CDC 2000 growth charts. J Nutr 137: 144–8.

De Onis M, Onyango AW, Borghi E, Garza C, Yang H. (2006). Comparison of the World Health Organization (WHO) Child Growth Standards and the National Center for Health Statistics/WHO international growth reference: implications for child health programmes. Public Health Nutr 9: 942–7.

De Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. (2007b). Development of a WHO growth reference for school-aged children and adolescents. Bull World Health Organ 85: 660–7.

De Onis M, Siyam A, Borghi E, Onyango AW, Piwoz E, Garza C. (2011). Comparison of the World Health Organization growth velocity standards with existing US reference data. Pediatrics, 128: e18–26.

De Onis M, Wijnhoven TM, Onyango AW. (2004). Worldwide practices in child growth monitoring. J Pediatr 144: 461–5.

De Onis M, Yip R. (1996). The WHO growth chart: historical considerations and current scientific issues. Bibl Nutr dieta 53:74–89.

Demaeseleer I, Hauspie R (sup). (2001). Biometrische analyse van jonge volwassenen in Vlaanderen. Masterthesis, Vrije Universiteit Brussel, Brussels

Dibley MJ, Goldsby JB, Staehling NW, Trowbridge FL. (1987). Development of normalized curves for the international growth reference: historical and technical considerations. Am J Clin Nutr 46: 736–48.

Dietz WH, Bellizzi MC. (1999). Introduction: the use of body mass index to assess obesity in children. Am J Clin Nutr 70: 123S–5S.

Dietz WH, Robinson TN. (1998). Use of the body mass index (BMI) as a measure of overweight in children and adolescents. J Pediatr 132: 191–3.

Emery JL, Waite AJ, Carpenter RG, Limerick SR, Blake D. (1985). Apnoea monitors compared with weighing scales for siblings after cot death. Arch Dis Child 60: 1055–60.

Eusebio JS, Nube M. (1981). Attainable growth. Lancet, 2: 1223

Falkner F, Buzina R, Chopra J, György P, Jelliffe D, Jelliffe P, McKigney J, et al. (1972). The creation of growth standards: a committee report. Am J Clin Nutr, 25: 218–20.

Fatti LP, Senaoana EM, Thompson ML. (1998). Bayesian updating in reference centile charts. J R Stat Soc Ser A Stat Soc 161: 103–115.

Finney D. (1971). Probit analysis (3rd ed.). Cambridge, UK: Cambridge University Press.

Fredriks AM, van Buuren S, Burgmeijer RJ, Meulmeester JF, Beuker RJ, Brugman E, Roede MJ, et al. (2000). Continuing positive secular growth change in The Netherlands 1955-1997. Pediatr Res 47: 316–23.

Fredriks AM, van Buuren S, Jeurissen SER, Dekker FW, Verloove-Vanhorick SP, Wit JM. (2003). Height, weight, body mass index and pubertal development reference values for children of Turkish origin in the Netherlands. Eur J Pediatr 162: 788–93.

Fredriks AM, van Buuren S, Jeurissen SER, Dekker FW, Verloove-Vanhorick SP, Wit JM. (2004). Height, weight, body mass index and pubertal development references for children of Moroccan origin in The Netherlands. Acta Paediatr 93: 817–24.

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: 17–24.

Garner P, Panpanich R, Logan S. (2000). Is routine growth monitoring effective? A systematic review of trials. Arch Dis Child 82: 197–201.

Garrow JS, Webster J. (1985). Quetelet's index (W/H2) as a measure of fatness. Int J Obes 9: 147–53.

Goldstein H, Tanner JM. (1980). Ecological considerations in the creation and the use of child growth standards. Lancet, 315: 582–5.

Goldstein H. (1981). The effects of age grouping on the estimate of a correlation coefficient. Ann Hum Biol, 8: 181–3.

Graitcer PL, Gentry EM. (1981). Measuring children: one reference for all. Lancet, 318: 297–9.

Grote FK, Oostdijk W, De Muinck Keizer-Schrama SMPF, Dekker FW, Van Dommelen P, van Buuren S, Lodder-van der Kooij AM, et al. (2007). Referral patterns of children with poor growth in primary health care. BMC public health 7: 77.

Grote FK, Van Dommelen P, Oostdijk W, De Muinck Keizer-Schrama SMPF, Verkerk PH, Wit JM, van Buuren S. (2008). Developing evidence-based guidelines for referral for short stature. Arch Dis Child 93: 212–7.

Guo S, Roche AF, Moore WM. (1988). Reference data for head circumference and 1-month increments from 1 to 12 months of age. J Pediatr 113: 490–4.

Guo SM, Roche AF, Fomon SJ, Nelson SE, Chumlea WC, Rogers RR, Baumgartner RN, et al. (1991). Reference data on gains in weight and length during the first two years of life. J Pediatr 119: 355–62.

Gupta N, Goel K, Shah P, Misra A. (2012). Childhood obesity in developing countries: epidemiology, determinants, and prevention. Endocr Rev, 33: 48–70.

Haas JD, Campirano F (2006). Interpopulation variation in height among children 7 to 18 years of age. Food Nutr Bull 27: S212–23.

Habicht JP, Martorell R, Yarbrough C, Malina RM, Klein RE. (1974). Height and weight standards for preschool children. How relevant are ethnic differences in growth potential? Lancet 303: 611–4.

Hall D, Cole T, Elliman D, Gibson P, Logan S, Wales J. (2008). Growth monitoring. Arch Dis Child 93: 717–8.

Hall DM. (2000). Growth monitoring. Arch Dis Child 82: 10–5.

Hauspie RC, Bergman P, Bielicki T, Susanne C. (1994). Genetic variance in the pattern of the growth curve for height: a longitudinal analysis of male twins. Ann Hum Biol, 21: 347–62.

Hauspie RC, Molinari L. (2004). Parametric models for postnatal growth. In: Hauspie R, Cameron N, Molinari L (Eds.). Methods in Human Growth Research (pp. 205–233). Cambridge, UK: Cambridge University Press.

Hauspie RC, Pagezy H. (1989). Longitudinal study of growth of African babies: an analysis of seasonal variations in the average growth rate and the effects of infectious diseases on individual and average growth patterns. Acta Paediatr Suppl 350: 37–43.

Hauspie RC, Vercauteren M, Susanne C. (1996). Secular changes in growth. Horm Res, 45: S8–17.

Hauspie RC, Vercauteren M, Susanne C. (1997). Secular changes in growth and maturation: an update. Acta Paediatr Suppl 423: 20–7.

Hauspie RC, Wachholder A. (1986). Clinical standards for growth velocity in height of Belgian boys and girls, aged 2 to 18 years. Int J Anthropol, 1: 339–48.

He X. (1997). Quantile Curves without Crossing. Am Stat 51: 186–92.

Heagerty PJ, Pepe MS. (1999). Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in US children. J R Stat Soc Ser C Appl Stat 48: 533–51.

Healy MJR. (1962). The effect of age-grouping on the distribution of a measurement affected by growth. Am J Phys Anthropol 20: 49–50.

Healy MJR. (1974). Notes on the statistics of growth standards. Ann Hum Biol 1: 41–6.

Healy MJR. (1995). Statistics from the inside. 14. Diagnostic and screening tests and reference values. Arch Dis Child 72: 358–61.

Healy MJR, Rasbash J, Yang M. (1988a). Distribution-free estimation of age-related centiles. Ann Hum Biol 15: 17–22.

Healy MJR, Yang M, Tanner JM, Zumrawi FY. (1988b). The Use of Short-Term Increments in Length to Monitor Growth in Infancy. In Waterlow JC (Ed.), Linear Growth Retardation in Less Developed Countries (pp. 41–55). Nestle Nutrition Workshop Series, Vol. 14. Nestec Ltd., Vevey/Raven Press, Ltd., New York.

Hermanussen M. (1998). The analysis of short-term growth. Horm Res 49: 53–64.

Hermanussen M, Grasedyck L, Kromeyer-Hauschild K, Prokopec M, Chrzastek-Spruch H. (2002). Growth tracks in pre-pubertal children. Ann Hum Biol 29: 667–76.

Hermanussen M, Largo RH, Molinari L. (2001). Canalization in human growth: a widely accepted concept reconsidered. Eur J Pediatr 160: 163–7.

Himes JH. (1999). Minimum time intervals for serial measurements of growth in recumbent length or stature of individual children. Acta Paediatr 88: 120–5.

Hoey HM, Tanner JM, Cox LA. (1987). Clinical growth standards for Irish children. Acta Paediatr Suppl 338: 1–31.

Joossens J, Brems-Heyns E, Bande-Knops J, Eggermont E. (1977). Longitudinale-transversale kurven opgesteld volgens 97000 metingen bij meer dan 25000 kinderen gevolgd te Leuven van 1963 tot 1977, gecorrigeerd voor tijdtrend op 1977. Leuven.

Júlíusson PB, Eide GE, Roelants M, Waaler PE, Hauspie R, Bjerknes R. (2010). Overweight and obesity in Norwegian children: prevalence and socio-demographic risk factors. Acta Paediatr 99: 900–5.

Júlíusson PB, Roelants M, Eide GE, Moster D, Juul A, Hauspie R, Waaler PE, Bjerknes R. (2009). [Growth references for Norwegian children]. Tidsskr Nor Laegeforen 129: 281–6.

Júlíusson PB, Roelants M, Nordal E, Furevik L, Eide GE, Moster D, Hauspie R, Bjerknes R. (2013). Growth references for 0-19 year-old Norwegian children for length/height, weight, body mass index and head circumference. Ann Hum Biol 40: 220–7.

Karlberg P, Taranger J, Engström I, Karlberg J, Landström T, Lichtenstein H, Lindstrom B, et al. (1976). I. physical growth from birth to 16 years and longitudinal outcome of the study during the same age period. Acta Paediatr Suppl 258: 7–76.

Kerac M, Blencowe H, Grijalva-Eternod C, McGrath M, Shoham J, Cole TJ, Seal A. (2011). Prevalence of wasting among under 6-month-old infants in developing countries and implications of new case definitions using WHO growth standards: a secondary data analysis. Arch Dis Child 96: 1008–13.

Keys A, Fidanza F, Karvonen MJ, Kimura N, Taylor HL. (1972). Indices of relative weight and obesity. J Chron Dis 25: 329–43.

Kuczmarski RJ, Ogden CL, Guo SS, Grummer-Strawn LM, Flegal KM, Mei Z, Wei R., et al. (2002). 2000 CDC Growth Charts for the United States: methods and development. Vital Health Stat Series 11 246: 1–190.

Lampl M, Veldhuis JD, Johnson ML. (1992). Saltation and stasis: a model of human growth. Science 258: 801–3.

Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, et al. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature, 467: 832–8.

Li J, Park WJ, Roche AF. (1998). Decanalization of Weight and Stature During Childhood and Adolescence. Am J Hum Biol 359: 351–9.

Lobstein T, Baur L, Uauy R. (2004). Obesity in children and young people: a crisis in public health. Obes Rev 5 Suppl 1: 4–104.

Maher B. (2008). Personal genomes: The case of the missing heritability. Nature, 456: 18–21.

Marceau K, Ram N, Houts RM, Grimm KJ, Susman EJ. (2011). Individual differences in boys' and girls' timing and tempo of puberty: modeling development with nonlinear growth models. Dev Psychol 47: 1389–409.

Marshall WA. (1971). Evaluation of growth rate in height over periods of less than one year. Arch Dis Child 46: 414–20.

Marshall WA, Tanner JM. (1969). Variations in pattern of pubertal changes in girls. Arch Dis Child 44: 291–303.

Marshall WA, Tanner JM. (1970). Variations in the pattern of pubertal changes in boys. Arch Dis Child 45: 13–23.

Mayr A, Hothorn T, Fenske N. (2012). Prediction intervals for future BMI values of individual children: a non-parametric approach by quantile boosting. BMC Med Res Methodol 12: 6.

Mei Z, Grummer-Strawn LM, Pietrobelli A, Goulding A, Goran MI, Dietz WH. (2002). Validity of body mass index compared with other body-composition screening indexes for the assessment of body fatness in children and adolescents. Am J Clin Nutr 75: 978–85.

Mei Z, Grummer-Strawn LM, Thompson D, Dietz WH. (2004). Shifts in percentiles of growth during early childhood: analysis of longitudinal data from the California Child Health and Development Study. Pediatrics 113: e617–27.

Mei Z, Grummer-Strawn LM. (2011). Comparison of changes in growth percentiles of US children on CDC 2000 growth charts with corresponding changes on WHO 2006 growth charts. Clin Pediatr, 50: 402–7.

Michaelsen KF, Petersen S, Greisen G, Thomsen BL. (1994). Weight, length, head circumference, and growth velocity in a longitudinal study of Danish infants. Dan Med Bull 41: 577–85.

Molinari L, Gasser T. (2004). The human growth curve: distance, velocity and acceleration. In: Hauspie R, Cameron N, Molinari L (Eds.). Methods in Human Growth Research (pp. 27–54). Cambridge, UK: Cambridge University Press.

Mul D, Fredriks AM, van Buuren S, Oostdijk W, Verloove-Vanhorick SP, Wit JM. (2001). Pubertal development in The Netherlands 1965-1997. Pediatr Res 50: 479–86.

Nelder J, Wedderburn RM. (1972). Generalized Linear Models. J R Stat Soc Ser A 135: 370–84.

Neyzi O, Bundak R, Molzan J, Günöz H, Darendeliler F, Saka N. (1993). Estimation of annual height velocity based on short- versus long-term measurements. Acta Paediatr 82: 239–44.

Ong KK, Ahmed ML, Emmett PM, Preece MA, Dunger DB. (2000). Association between postnatal catch-up growth and obesity in childhood: prospective cohort study. BMJ 320: 967–71.

Pan H, Cole TJ. (2004). A comparison of goodness of fit tests for age-related reference ranges. Stat Med 23: 1749–65.

Pan H, Goldstein H. (1997). Multi-level models for longitudinal growth norms. Stat Med 16: 2665–78.

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

Park WJ, Li J, Roche AF. (1997). The decanalization of weight, recumbent length, and head circumference during infancy. Am J Hum Biol 9: 689–698.

PetitClerc C, Solberg H. (1987). International Federation of Clinical Chemistry (IFCC), Scientific Committee, Clinical Section, Expert Panel on Theory of Reference Values. Approved recommendation (1987) on the theory of reference values. Part 2. Selection of individuals for the product. J Clin Chem Clin Biochem 25: 639–44.

Piwoz EG, Peerson JM, Brown KH. (1992). Potential for misclassification of infants' growth increments by using existing reference data. Am J Clin Nutr, 56: 58–64.

Power C, Lake JK, Cole TJ. (1997). Measurement and long-term health risks of child and adolescent fatness. Int J Obes Relat Metab Disord 21: 507–26.

Prader A, Largo RH, Molinari L, Issler C. (1989). Physical growth of Swiss children from birth to 20 years of age. First Zurich longitudinal study of growth and development. Helv Paediatr Acta Suppl 52:1–125.

Quételet A. (1835). Sur l'homme et le développement de ses facultés, ou Essai de physique sociale (p. 327). Bachelier, Paris.

Ranke MB, Pflüger H, Rosendahl W, Stubbe P, Enders H, Bierich JR, Majewski F. (1983). Turner syndrome: spontaneous growth in 150 cases and review of the literature. Eur J Pediatr 141: 81–8.

Rigby RA, Stasinopoulos DM. (2004). Smooth centile curves for skew and kurtotic data modelled using the Box-Cox power exponential distribution. Stat Med 23: 3053–76.

Rigby RA, Stasinopoulos DM. (2005). Generalized additive models for location, scale and shape. J R Stat Soc Ser C Appl Stat 54: 507–54.

Rigby RA, Stasinopoulos DM. (2006). Using the Box–Cox t distribution in GAMLSS to model skewness and kurtosis. Statistical Modelling, 6(3), 209–229.

Roche AF, Guo S, Moore WM. (1989a). Weight and recumbent length from 1 to 12 mo of age: reference data for 1-mo increments. Am J Clin Nutr, 49(4), 599–607.

Roche AF, Guo SS, Yung D. (1989b). Monthy growth increments from a longitudinal study of Canadian infants. Am J Hum Biol 1: 271–9.

Roche AF, Himes JH. (1980). Incremental growth charts. Am J Clin Nutr, 33: 2041–52.

Roelants M, Hauspie R, Hoppenbrouwers K. (2008). Vlaamse Groeicurven 2004. Brussel, Leuven.

Roelants M, Hoppenbrouwers K. (2001). Prevalentie van obesitas bij kinderen en jongeren op basis van de body mass index. TJGZ, 88–94.

Rolland-Cachera MF, Péneau S. (2011). Assessment of growth: variations according to references and growth parameters used. Am J Clin Nutr, 94: 1794S–8.

Royston P. (1995). Calculation of unconditional and conditional reference intervals for foetal size and growth from longitudinal measurements. StatMed 14: 1417–36.

Royston P. (2005). Reference Intervals for Normal Clinical Values. Encyclopedia of Biostatistics.

Royston P, Wright EM. (2000). Goodness-of-fit statistics for age-specific reference intervals. Stat Med 19: 2943–62.

Schaffrath Rosario AS, Schienkiewitz A, Neuhauser, H. (2011). German height references for children aged 0 to under 18 years compared to WHO and CDC growth charts. Ann Hum Biol 38: 121–30.

Schönbeck Y, Talma H, Van Dommelen P, Bakker B, Buitendijk SE, Hirasing RA, van Buuren S. (2011). Increase in prevalence of overweight in Dutch children and adolescents: a comparison of nationwide growth studies in 1980, 1997 and 2009. PloS one 6: e27608.

Schönbeck Y, Talma H, Van Dommelen P, Bakker B, Buitendijk SE, HiraSing RA, Van Buuren S. (2013). The world's tallest nation has stopped growing taller: the height of Dutch children from 1955 to 2009. Pediatr Res 73: 371–7.

Shuttleworth FK. (1934). Standards of development in terms of increments. Child Dev 5: 89–90.

Smith DW, Truog W, Rogers JE, Greitzer LJ, Skinner AL, McCann JJ, Harvey MA (1976). Shifting linear growth during infancy: illustration of genetic factors in growth from fetal life through infancy. J Pediatr 89: 225–30.

Sørensen K, Aksglaede L, Petersen JH, Juul A. (2010). Recent changes in pubertal timing in healthy Danish boys: associations with body mass index. J Clin Endocrinol Metab 95: 263–70.

Sørensen K, Mouritsen A, Aksglaede L, Hagen CP, Mogensen SS, Juul A. (2012). Recent secular trends in pubertal timing: implications for evaluation and diagnosis of precocious puberty. Horm Res Pædiatr 77: 137–45.

Stirnemann JJ, Samson A, Thalabard JC. (2012). Individual predictions based on nonlinear mixed modeling: application to prenatal twin growth. Statistics in medicine 31: 1986–99.

Swan A. (1969). Computing Maximum-Likelihood Estimates for Parameters of the Normal Distribution from Grouped and Censored Data. J R Stat Soc Ser C Appl Stat 18: 65–69.

Talma H, Schönbeck Y, Van Dommelen P, Bakker B, Van Buuren S, Hirasing RA. (2013). Trends in menarcheal age between 1955 and 2009 in the Netherlands. PloS one: 8: e60056.

Tanner JM. (1951). Some notes on the reporting of growth data. Hum Biol, 23: 93–159.

Tanner JM. (1952). The assessment of growth and development in children. Arch Dis Child 27: 10–33.

Tanner JM. (1987). Growth as a mirror of the condition of society: secular trends and class distinctions. Acta Paediatr Jap 29: 96–103.

Tanner JM, Buckler JM. (1997). Revision and update of Tanner-Whitehouse clinical longitudinal charts for height and weight. Eur J Pediatr 156: 248–9.

Tanner JM, Davies PS. (1985). Clinical longitudinal standards for height and height velocity for North American children. J Pediatr 107: 317–29.

Tanner JM, Goldstein H, Whitehouse RH. (1970). Standards for children's height at ages 2-9 years allowing for heights of parents. Arch Dis Child 45: 755–62.

Tanner JM, Whitehouse RH, Takaishi M. (1966a). Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children, 1965 I. Arch Dis Child 41: 454–71.

Tanner JM, Whitehouse RH, Takaishi M. (1966b). Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children, 1965 II. Arch Dis Child 41: 613–35.

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.

Thalange NK, Foster PJ, Gill MS, Price DA, Clayton PE. (1996). Model of normal prepubertal growth. Arch Dis Child 75: 427–31.

Thomis MA, Towne B. (2006). Genetic determinants of prepubertal and pubertal growth and development. Food Nutr Bull 27: S257–78.

Thompson ML, Fatti LP. (1997). Construction of multivariate centile charts for longitudinal measurements. Stat Med 16: 333–45.

Ulijaszek SJ. (2006). The international growth standard for children and adolescents project: environmental influences on preadolescent and adolescent growth in weight and height. Food Nutr Bull 27: S279–94.

Ulijaszek SJ, Kerr DA. (1999). Anthropometric measurement error and the assessment of nutritional status. Brit J Nutr 82: 165–77.

Ulijaszek SJ. (2001). Ethnic differences in patterns of human growth in stature. In Martorell R, Haschke F (Eds.). Nutrition and growth (pp. 1–20). Philadelphia: Nestlé Nutrition Workshop Series. Vol 47. Lippincot Williams and Wilkins.

Van 't Hof MA, Haschke F, Darvay S. (2000). Euro-Growth references on increments in length, weight, and head and arm circumferences during the first 3 years of life. Euro-Growth Study Group. J Pediatr pediatric Gastroenter Nutr 31 Suppl 1: S39–47.

Van 't Hof MA, Wit JM, Roede MJ. (1985). A method to construct age references for skewed skinfold data, using Box-Cox transformations to normality. Hum Biol 57: 131–9.

van Buuren S. (2004). Afkapwaarden van de “body-mass index” (BMI) voor ondergewicht van Nederlandse kinderen. Ned Tijdschr Geneesk 148: 1967–1972.

van Buuren S. (2010). Effects of selective dropout on infant growth standards. Nestlé Nutrition workshop series. Paediatric programme, 65, 167–75; discussion 175–9.

van Buuren S, Bonnemaijer-Kerckhoffs DJA, Grote FK, Wit JM, Verkerk PH. (2004a). Many referrals under Dutch short stature guidelines. Arch Dis Child, 89: 351–2.

van Buuren S, Fredriks M. (2001). Worm plot: a simple diagnostic device for modelling growth reference curves. Stat Med 20: 1259–77.

van Buuren S, Schönbeck Y, Van Dommelen P. (2012). Collection, collation and analysis of data in relation to reference heights and reference weights for female and male children and adolescents (0-18 years) in the EU, as well as in relation to the age at which different stages of puberty are reached in adolescents in the EU. (pp. 1–59). Leiden, The Netherlands.

van Buuren S, Van Dommelen P, Zandwijken GRJ, Grote FK, Wit JM, Verkerk PH. (2004b). Towards evidence based referral criteria for growth monitoring. Arch Dis Child 89: 336–41.

Van Buuren S, Van Wouwe JP. (2008). WHO Child Growth Standards in action. Arch Dis Child 93: 549–51.

Van Hoeck K, Hoppenbrouwers K, 2005. Standaard groei en pubertaire ontwikkeling. Vlaamse Wetenschappelijke Vereniging voor Jeugdgezondheidszorg, Leuven, Belgium.

Vercauteren M, Hauspie RC, Susanne C. (1998). Biometry in Belgian boys and girls: changes since Quételet. In Bodzsár E, Susanne S (Eds.). Secular changes in Europe (pp. 47–63). Budapest, Hungary: Eötvös Univ Press.

Vercauteren M, Susanne C. (1985). The secular trend of height and menarche in Belgium: are there any signs of a future stop? Eur J Pediatr 144: 306–9.

Wachholder A, Hauspie RC. (1986). Clinical standards for growth in height of Belgian boys and girls, aged 2 to 18 years. Int J Anthropol 1: 327–38.

Wade A, Kurmanavicius J. (2009). Creating unbiased cross-sectional covariate-related reference ranges from serial correlated measurements. Biostatistics 10: 147–54.

Wade AM, Ades AE. (1998). Incorporating correlations between measurements into the estimation of age-related reference ranges. Stat Med 17: 1989–2002.

Waterlow JC, Buzina R, Keller W, Lane JM, Nichaman MZ, Tanner JM. (1977). The presentation and use of height and weight data for comparing the nutritional status of groups of children under the age of 10 years. Bull World Health Organ 55: 489–98.

Waterlow JC. (1980). Child growth standards. Lancet 315: 717.

Wei Y, He X. (2006). Conditional growth charts. Ann Stat, 34(5), 2069–97.

WHO Expert Committee on Physical Status: the use and interpretation of anthropometry. (1995b). Physical status: the use and interpretation of anthropometry. WHO technical report series, 845.

WHO Multicentre Growth Reference Study Group. (2006a). WHO Child Growth Standards. Length/height-for-age, weight-forage, weight-for-length, weight-for-height and body mass index-for-age. Methods and development. (p. 336). Geneva.

WHO Multicentre Growth Reference Study Group. (2006b). Enrolment and baseline characteristics in the WHO Multicentre Growth Reference Study. Acta paediatr Suppl 450: 7–15.

WHO Multicentre Growth Reference Study Group. (2006c). Assessment of differences in linear growth among populations in the WHO Multicentre Growth Reference Study. Acta paediatr Suppl 450: 56–65.

WHO Multicentre Growth Reference Study Group. (2006d). Reliability of anthropometric measurements in the WHO Multicentre Growth Reference Study. Acta paediatr Suppl 450: 38–46.

WHO Multicentre Growth Reference Study Group. (2009). WHO Child Growth Standards. Growth velocity based on weight, length and head circumference (p. 262). Geneva.

WHO Working Group on Infant Growth. (1995a). An evaluation of infant growth: the use and interpretation of anthropometry in infants. WHO Working Group on Infant Growth. Bull World Health Organ 73: 165–74.

WHO Working Group on the Growth Reference Protocol. (2000). Growth patterns of breastfed infants in seven countries. Acta Paediatr 89: 215–22.

WHO Working Group on the use and interpretation of anthropometric indicators of nutritional status. (1986). The use and interpretation of anthropometric indicators of nutritional status. Bull World Health Organ 64: 929–41.

Wit JM, Ranke MB, Kelnar CJH. (2007). ESPE Classification of Paediatric Endocrine Diagnoses. Horm Res 68 Suppl 2: 1–120.

Wright C, Avery A, Epstein M, Birks E, Croft D. (1998). New chart to evaluate weight faltering. Arch Dis Child 78: 40–3.

Wright C, Lakshman R, Emmett P, Ong KK. (2008). Implications of adopting the WHO 2006 Child Growth Standard in the UK: two prospective cohort studies. Arch Dis Child 93: 566–9.

Wright CM, Inskip HM, Godfrey K, Williams AF, Ong KK. (2011). Monitoring head size and growth using the new UK-WHO growth standard. Arch Dis Child 96: 386–8.

Wright CM, Matthews JN, Waterston A, Aynsley-Green A. (1994). What is a normal rate of weight gain in infancy? Acta Paediatr 83: 351–6.

Wright EM, Royston P. (1997). A Comparison of Statistical Methods for Age-related Reference Intervals. J R Stat Soc A 160: 47–69.

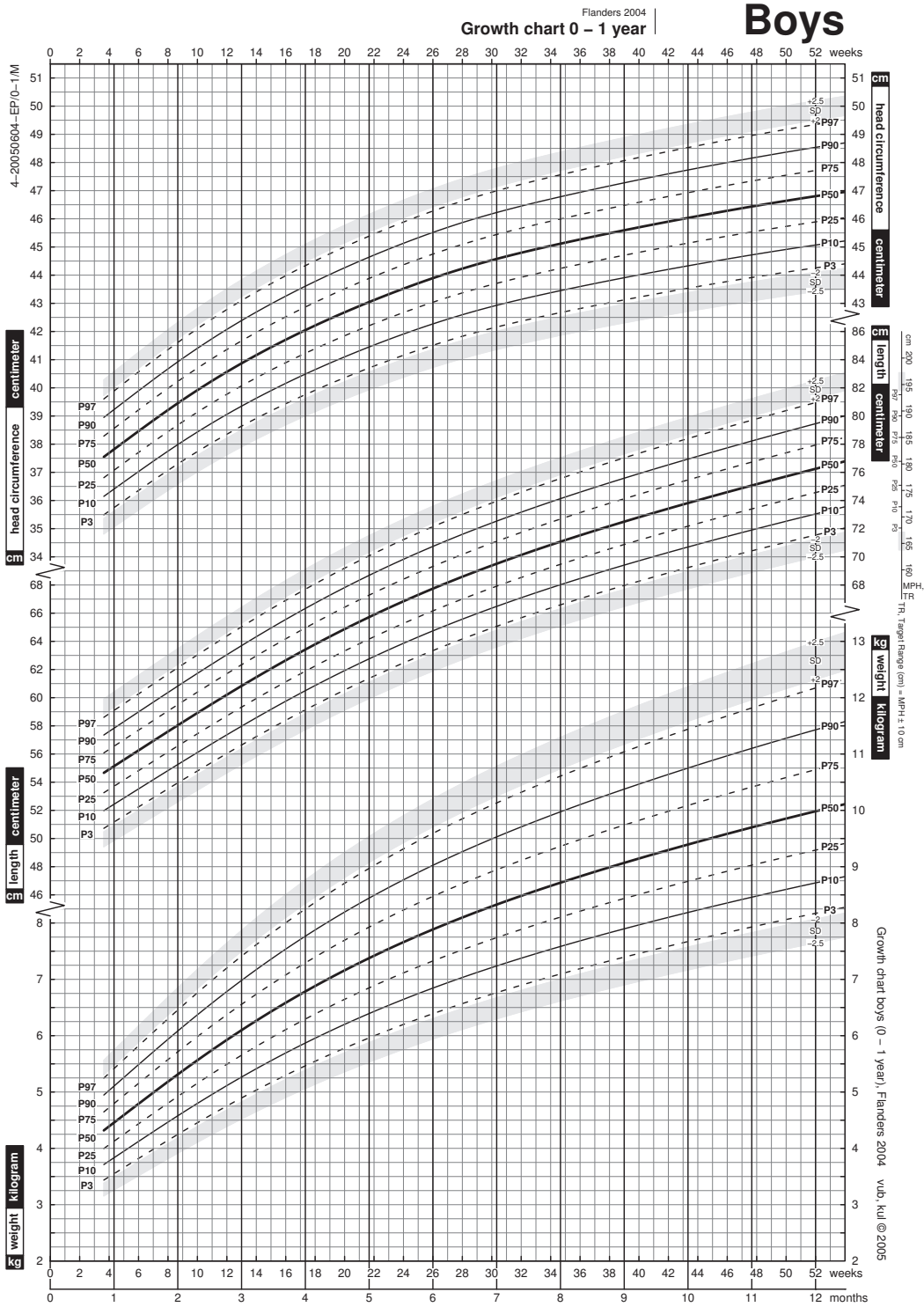
Wright EM, Royston P. (1999). Calculating reference intervals for laboratory measurements. Stat Methods Med Res 8: 93–112.

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden, PA, et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nature genetics 42: 565–9.

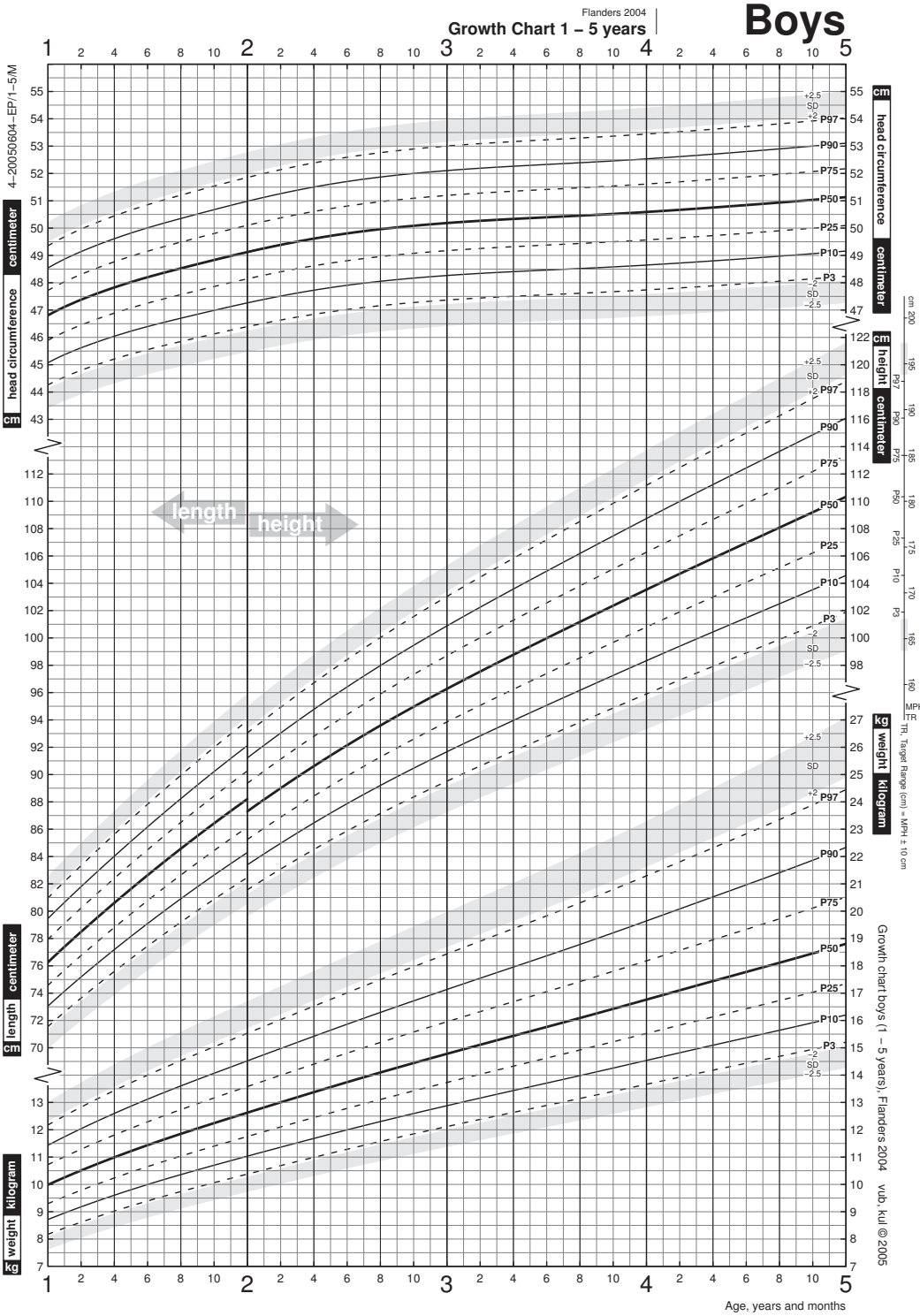
Ziegler EE, Nelson SE. (2012). The WHO growth standards: strengths and limitations. Curr Opin Clin Nutr 15: 298–302.

Appendix: Flemish Growth Charts 2004

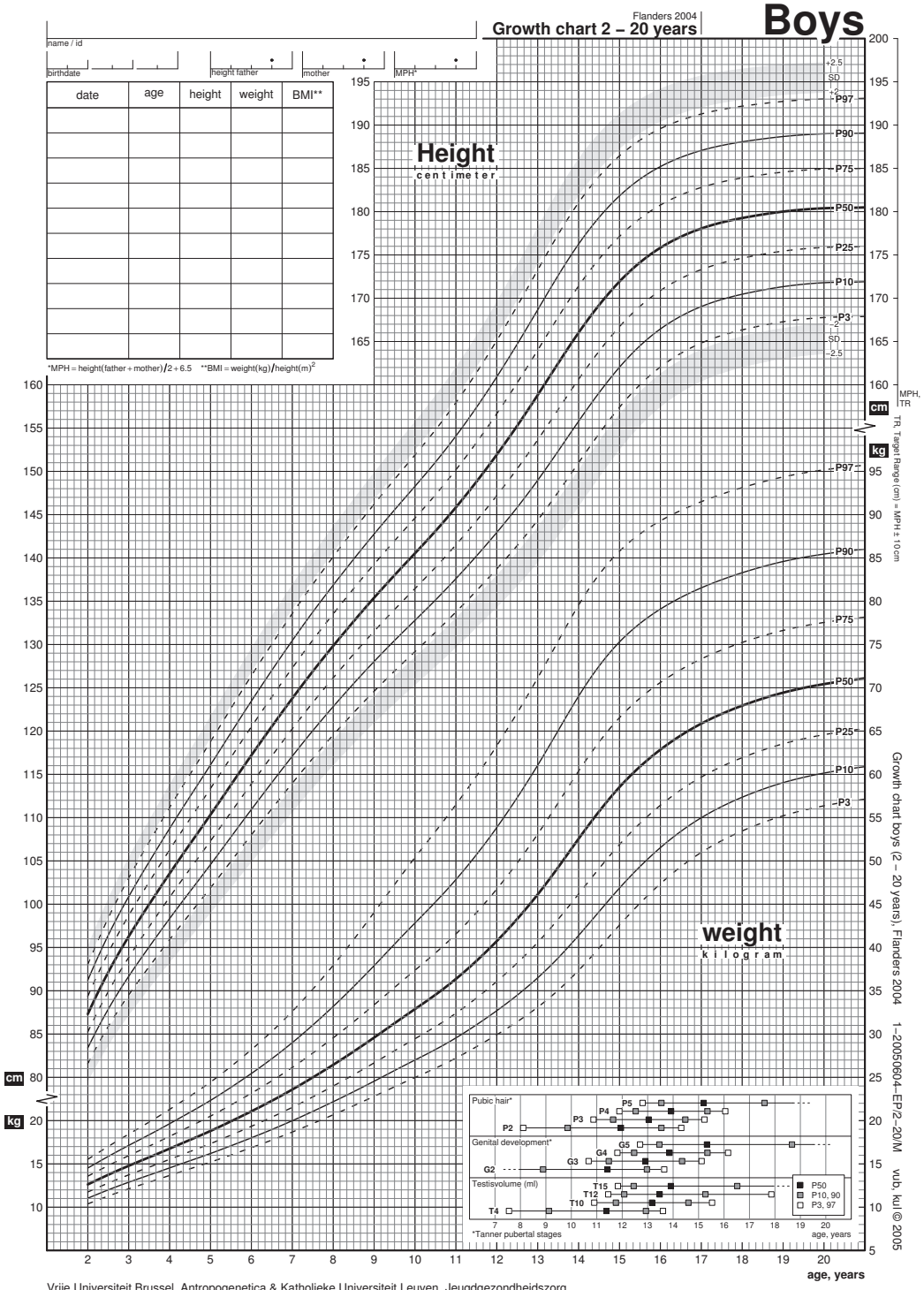
1) Length, weight, head circumference, boys, 0 – 12 months
2) Length/height, weight, head circumference, boys, 1 – 5 years
3) Height, weight, pubertal development, boys, 2 – 21 years
4) BMI, yearly increments in height, boys, 2 – 21 years
5) Length, weight, head circumference, girls, 0 – 12 months
6) Length/height, weight, head circumference, girls, 1 – 5 years
7) Height, weight, pubertal development, girls, 2 – 21 years
8) BMI, yearly increments in height, girls, 2 – 21 years



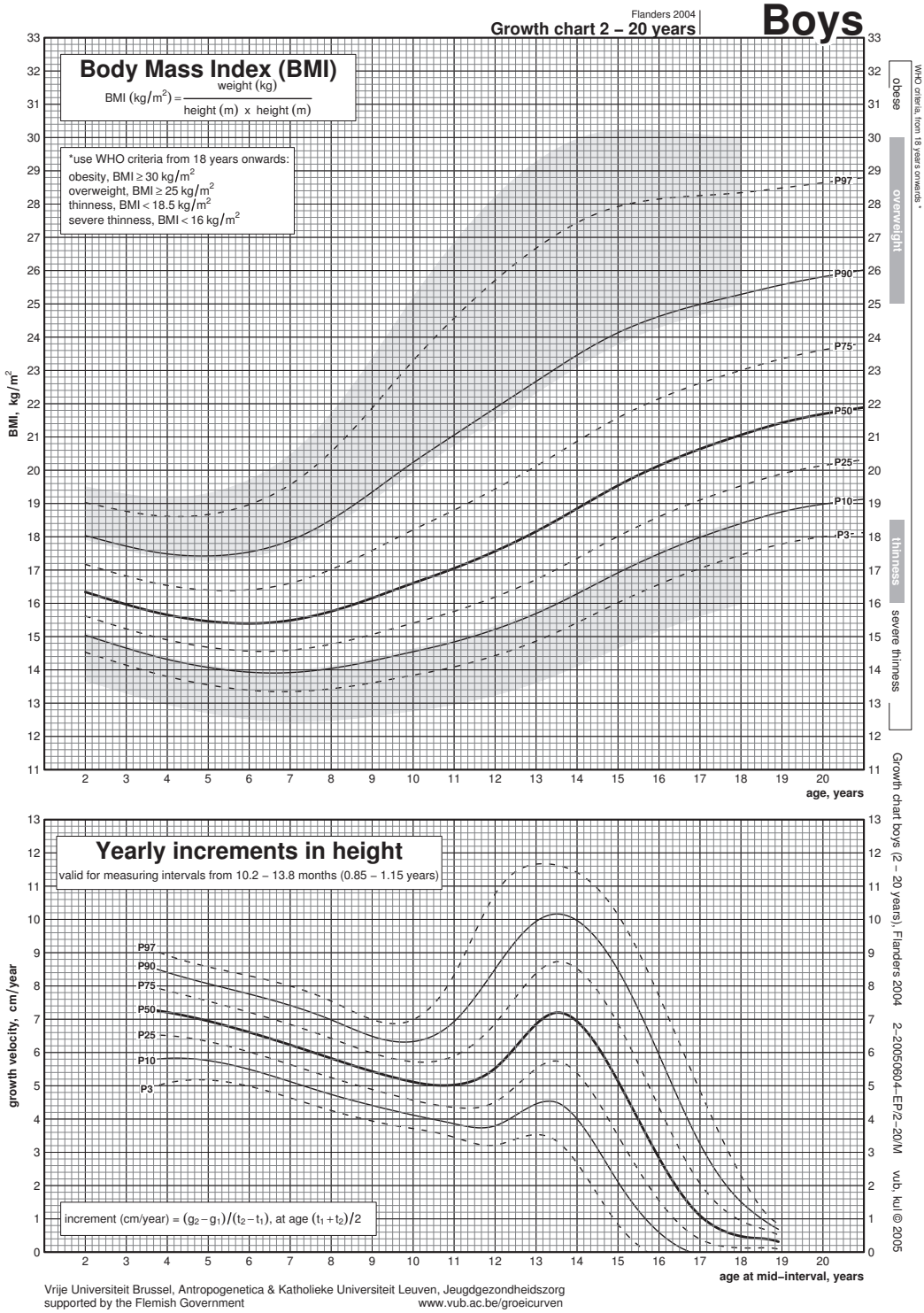
Vrije Universiteit Brussel, Antropogenetica & Katholieke Universiteit Leuven, Jeugdgezondheidszorg
supported by the Flemish government www.vub.ac.be/groeicurven



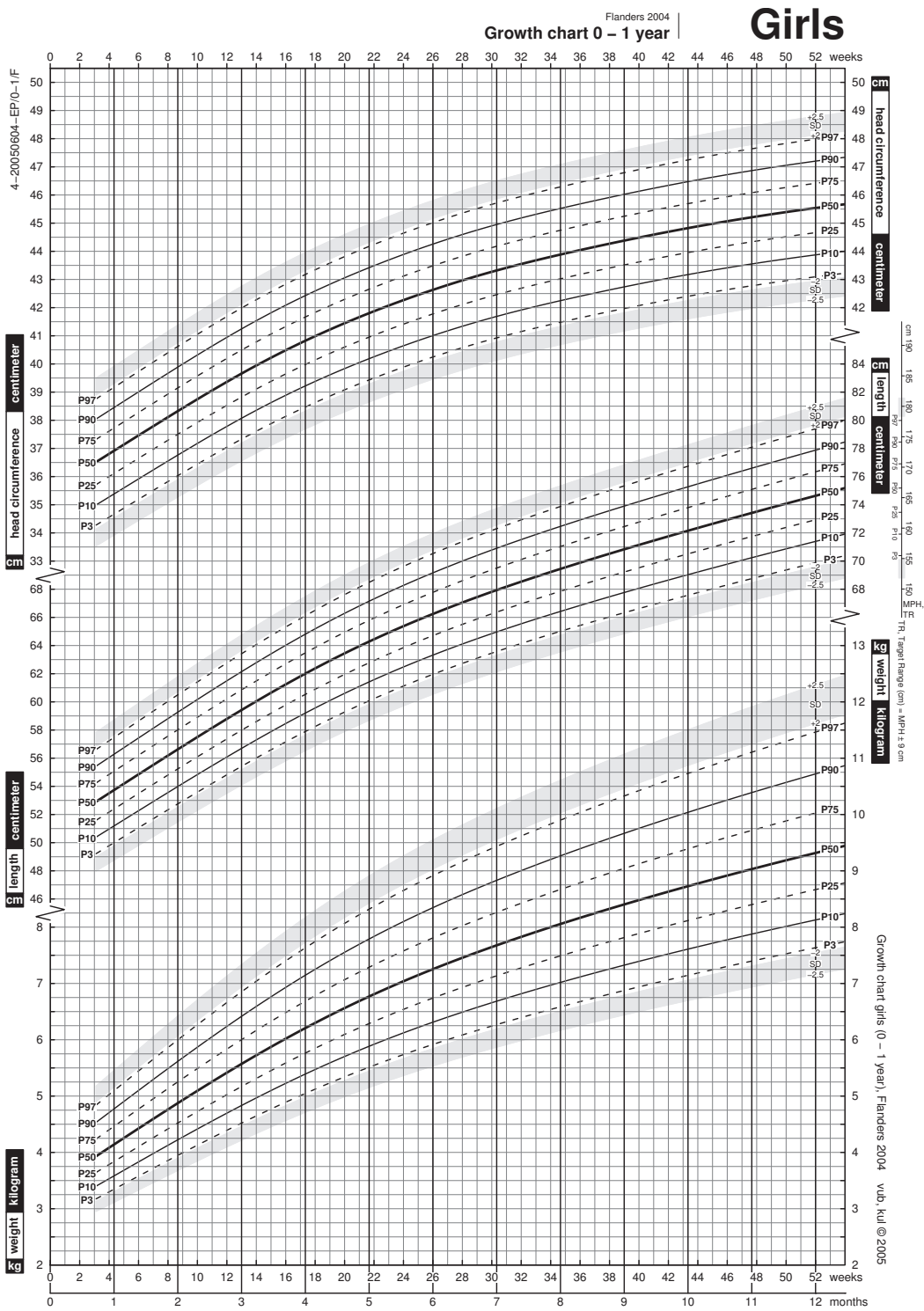
Vrije Universiteit Brussel, Antropogenetica & Katholieke Universiteit Leuven, Jeugdgezondheidszorg
supported by the Flemish government www.vub.ac.be/groeicurven



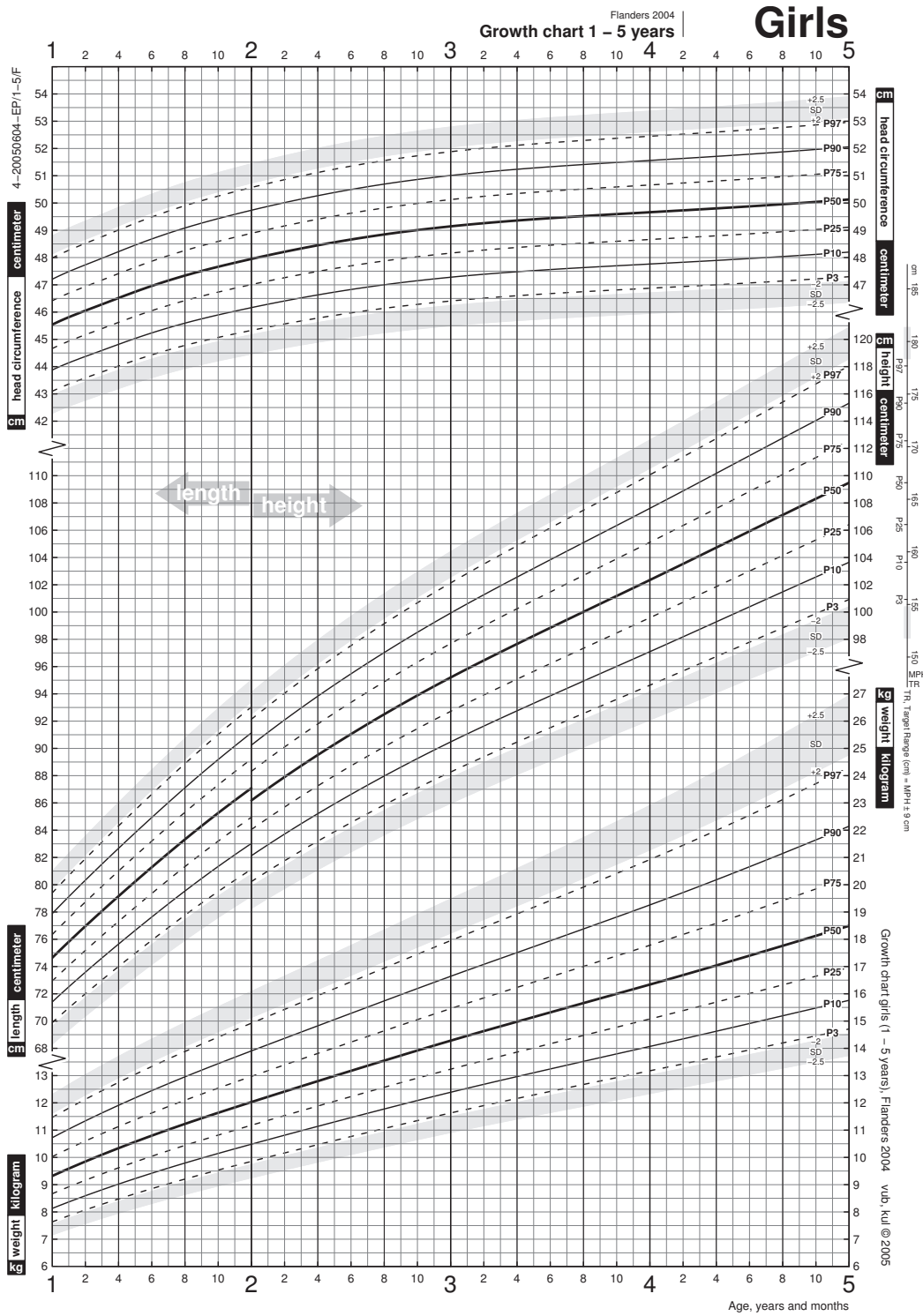
Vrije Universiteit Brussel, Antropogenetica & Katholieke Universiteit Leuven, Jeugdgezondheidszorg
supported by the Flemish Government www.vub.ac.be/groeicurven

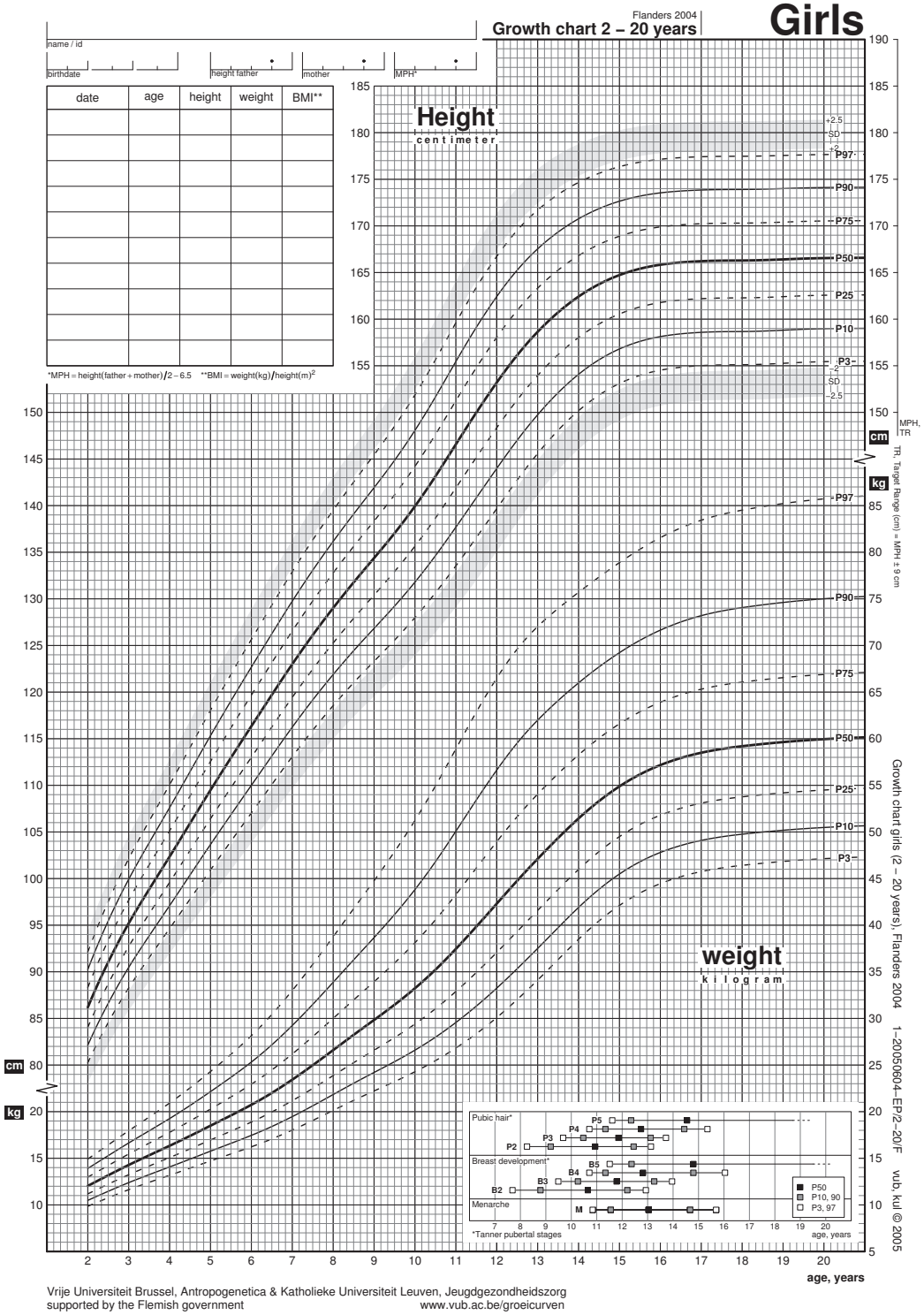


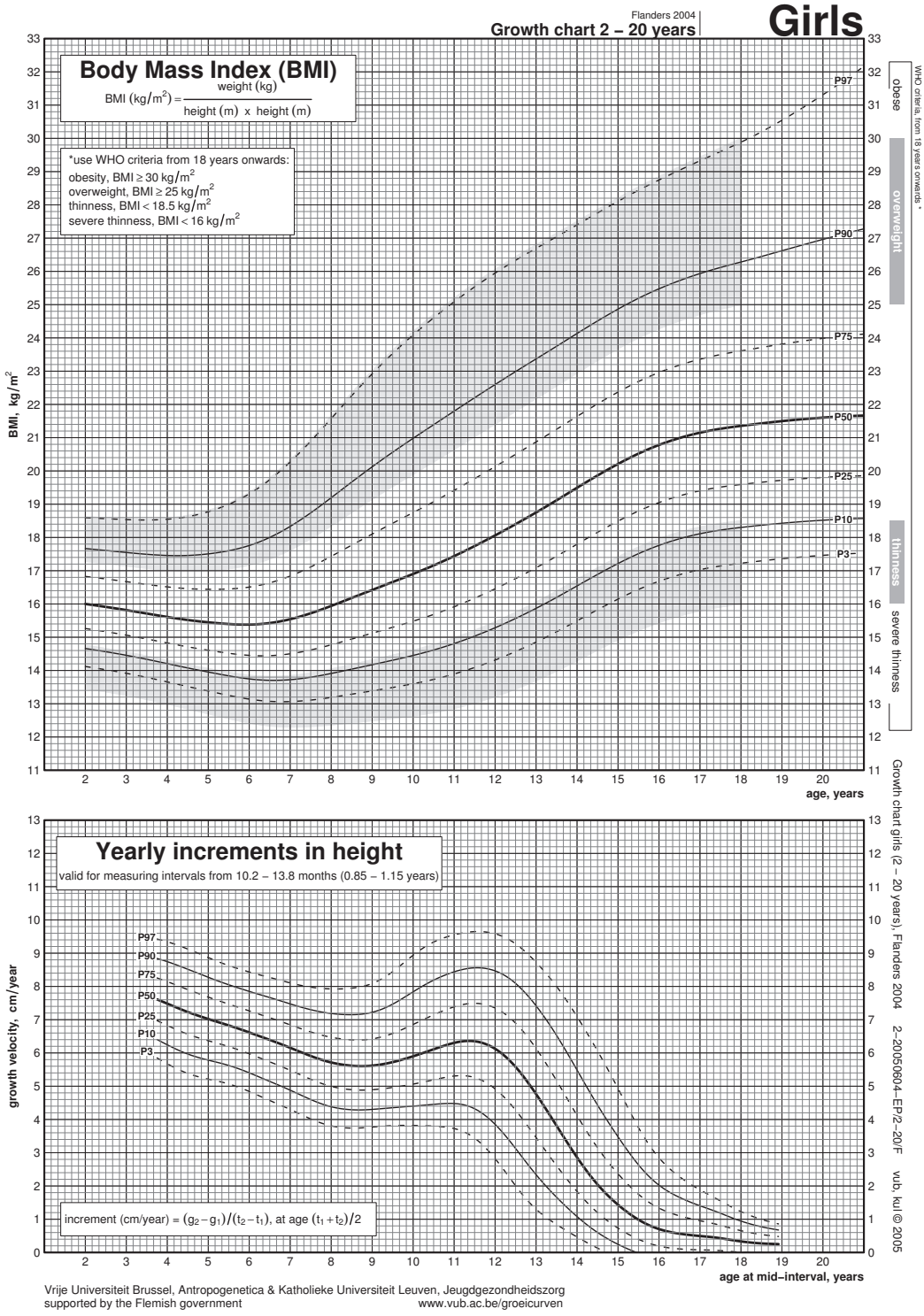
Vrije Universiteit Brussel, Antropogenetica & Katholieke Universiteit Leuven, Jeugdgezondheidszorg
supported by the Flemish Government www.vub.ac.be/groeicurven



Vrije Universiteit Brussel, Antropogenetica & Katholieke Universiteit Leuven, Jeugdgezondheidszorg
supported by the Flemish government www.vub.ac.be/groeicurven


Flanders 2004
Growth chart 1 – 5 years
Girls
4–20050604–EP/1–5/F
head circumference centimeter
length centimeter
weight kilogram
P97
P90
P75
P50
P25
P10
P3
+2 SD
+2.5 SD
–2 SD
–2.5 SD
length
height
Age, years and months
Growth chart girls (1 – 5 years), Flanders 2004 vub, kul © 2005
IR, Target Range (cm) = MPH ± 9 cm
Vrije Universiteit Brussel, Antropogenetica & Katholieke Universiteit Leuven, Jeugdgezondheidszorg
supported by the Flemish government
www.vub.ac.be/groeicurven


Flanders 2004
Growth chart 2 – 20 years
Girls
name / id
birthdate
height father
mother
MPH*
date
age
height
weight
BMI**
*MPH = height(father + mother)/2 – 6.5 **BMI = weight(kg)/height(m)²
Height
centimeter
+2.5 SD
+2 SD
P97
P90
P75
P50
P25
P10
P3
-2 SD
-2.5 SD
cm
kg
MPH, TR
TR, Target Range (cm) = MPH ± 9 cm
weight
kilogram
Pubic hair*
P5
P4
P3
P2
Breast development*
B5
B4
B3
B2
Menarche
M
P50
P10, 90
P3, 97
*Tanner pubertal stages
age, years
Growth chart girls (2 – 20 years), Flanders 2004 1–20050604–EP/2–20/F vub, kul © 2005
Vrije Universiteit Brussel, Antropogenetica & Katholieke Universiteit Leuven, Jeugdgezondheidszorg
supported by the Flemish government
www.vub.ac.be/groeicurven



Vrije Universiteit Brussel, Antropogenetica & Katholieke Universiteit Leuven, Jeugdgezondheidszorg
supported by the Flemish government www.vub.ac.be/groeicurven