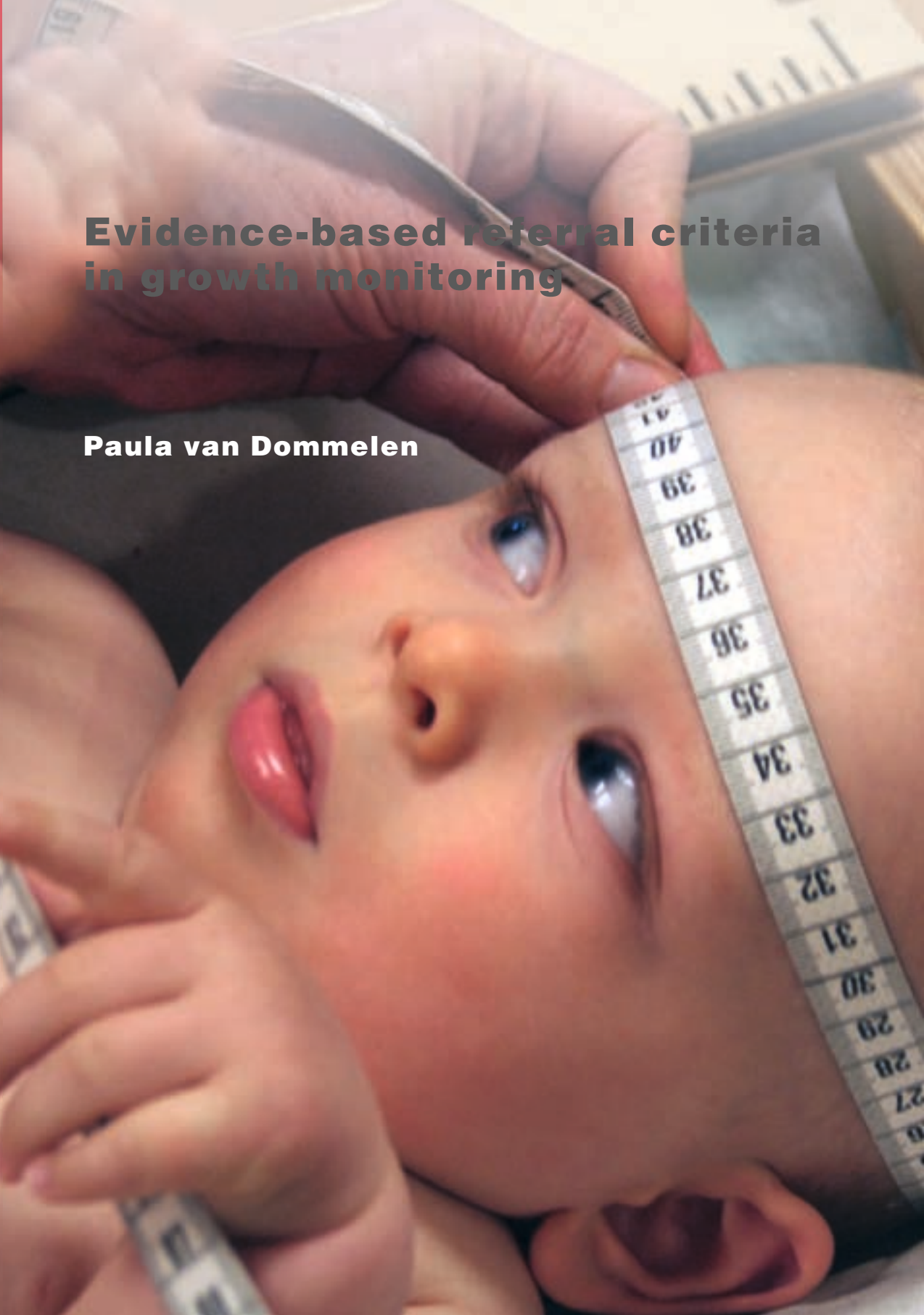
Evidence-based referral criteria in growth monitoring

Paula van Dommelen



Evidence-based referral criteria in growth monitoring

ISBN 978-90-393-4963-2

Evidence-based referral criteria in growth monitoring

Evidence-based verwijscriteria voor groei
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. J.C. Stoof,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 5 december 2008
des ochtends te 10.30 uur
door

Paula van Dommelen

geboren op 2 januari 1978, te Linschoten

Promotor: Prof.dr. S. van Buuren

Co-promotor: Dr. P.H. Verkerk

Members of the thesis committee:	Prof.dr. M.A. van den Hout
	Prof.dr. T.J. Cole
	Prof.dr. J.M. Wit
	Prof.dr. A.C. Hokken-Koelega
	Prof.dr. J.J. Hox

Table of contents

General introduction and outline of this thesis

1

Introduction

Growth monitoring in infancy and childhood has been part of preventive child health programs for more than a century in both developed and underdeveloped countries. It is a popular tool for defining health and nutritional status of children.[1] At the individual level, growth monitoring consists of measuring the individual's height, weight and head circumference and plotting these measurements on a growth chart. The position of the measurements on such a growth chart shows whether the growth pattern of the child deviates from that of the reference population. An important goal of growth monitoring is to identify, at an early stage, genetic disorders, diseases or other conditions that manifest themselves through an abnormal growth.[2,3] Examples of conditions that may be detected by growth monitoring include Turner's syndrome, growth hormone deficiency, juvenile hypothyroidism, psychosocial deprivation, skeletal abnormalities, multi-symptomatic syndromes, celiac disease, cystic fibrosis, hydrocephalus, (hypernatraemic) dehydration, malnutrition and obesity.[3] Notice that obesity is not a condition in itself, but it is a risk factor of early mortality and severe illnesses. At the population level, growth monitoring consists of studying surveys of growth data to assess health and nutritional status for purposes of program planning, implementation, and evaluation.

Despite the longstanding and wide acceptance of growth monitoring, not much is known about using growth monitoring as a screening program.[2-6] According to the United Kingdom (UK) national screening committee (NSC), screening is defined as: "a public health service in which members of a defined population, who do not necessarily perceive they are at risk of, or are already affected by a disease or its complications, are asked a question or offered a test, to identify those individuals who are more likely to be helped than harmed by further tests or treatment to reduce the risk of a disease or its complications."[7] The intention of screening is to identify the disease early, thus enabling timely intervention and management in the hope to reduce mortality and suffering from the disease. Growth monitoring can be considered as a screening program. However, in general there are three important differences between growth monitoring and other screening programs. First, a referral is often made on a combination of abnormal growth and other clinical symptoms, while the result of a conventional screening program usually only depends on the result of a test. Second, growth monitoring is aimed at identifying multiple diseases simultaneously, while a screening program is often aimed at identifying one disease. Third, growth monitoring is usually performed over a certain time period, while a screening program offers a test to the population at one moment in time or several tests within a short time period.

The NSC developed 22 quality criteria for appraising the viability, effectiveness and appropriateness of a screening program.[7] One important screening criterion of the NSC is that there should be a validated screening test. In growth monitoring, a screening test consists of one or several referral criteria based upon growth. For example, the lowest centile of the growth chart for height can be used as a screening test. The test is then considered positive if a child's height is below this centile, and negative if above. Furthermore, the screening test should be validated. In other words, a proper screening test should be able to detect, at an early stage, as many children with growth-related conditions as possible (high sensitivity) at the account of only a limited number of children with a false-positive result (high specificity). There is a clear trade-off between sensitivity and specificity: higher sensitivity will usually mean lower specificity (and hence more unnecessary referrals); higher specificity will usually mean lower sensitivity (and hence more missed children with growth-related conditions). Furthermore, sensitivity and specificity are related to age. It is usually easier to find the condition if we wait longer. However, early treatment often leads to better outcomes compared to treatment at an older age. Therefore, optimizing sensitivity, specificity and referral age is an interrelated problem.

In the Netherlands, a consensus meeting was held in 1996 on referral criteria to diagnose short stature in childhood. As a result of this meeting, a Dutch consensus guideline for short stature was published.[8] However, since the guideline was based on consensus, it lacked evidence. In 2004 it was shown that if the Dutch consensus would be followed strictly, an unacceptably high percentage (over 80%) of healthy children would have to be referred.[9] In order to improve the situation, we initiated this study with the following research question:

What is the validity of referral criteria in growth monitoring?

This research question is an important part in the investigation if growth monitoring can be used as a screening program. The 22 NSC quality criteria are subdivided into four groups:

1. the epidemiology of the condition,

2. the properties of the test,

3. any treatment options, and

4. the acceptability of the screening program.

We will now discuss several elements of these four groups in detail.

1. The first group deals with the epidemiology of the condition. The condition should be an important health problem. It is known that an abnormal growth can be an early sign of several conditions and that some of these conditions are important health problems. The rapid rise in childhood obesity is a good example.

2. The second group concerns the properties of the test or referral criteria. This implies that the referral criteria should be acceptable to the population, safe, validated, precise and simple. Growth monitoring is currently already widely implemented in practice and is generally considered as safe. The central question in this thesis is the validity of referral criteria in growth monitoring. Furthermore, the referral criteria should be simple. The simplest criterion consists of comparing a single anthropometric measurement to some norm. More advanced criteria may involve multiple measurements over time. Repeated height or weight measurements over time allow for calculation of a growth rate and can be used to define an abnormal increase or decrease in growth. Until now, it is unclear whether the use of multiple measurements forms a substantial improvement over a single measurement. Related to this are the timing of measurements. Repeated measurements of height and weight as part of scheduled visits at child health care centers have been suggested,[10] but until now there is little information with regard to the most cost-effective number and timing of visits. The current recommendation in the Netherlands is that visits should be organized according to the immunization schedule, with additional visits in periods of rapid growth. According to the Dutch protocol, anthropometric measurements are carried out at birth (weight), at 1, 2, 3, 4, 6, 9, 11, 14, 24 and 45 months and at approximately 5-6 years, 9-10 years and at 13 years of age (weight and height). Head circumference is only measured in the first year of life. This thesis presents referral criteria based on measurements from multiple visits ($\geq$3) where the timing of measurement may differ. Potentially, the diagnostic performance of a given referral criterion can be enhanced by including covariates that are known to have a significant effect on growth. Many studies have reported genetic and environmental covariates of (fetal) growth, such as parental height, ethnicity, age of mother, sex, social economic status of the parents, condition of the mother and smoking.[11,12] This thesis studies the contribution of genes and environment on the growth process during infancy. We include parental height as a parameter in the referral criteria, because this is a potentially important covariate of growth. Target height is the term used for the expected height of a child given the height of the (natural) parents.

3. An important element of the third group states that there should be an effective treatment or intervention for the children with conditions identified through

early detection, with evidence leading to better outcomes than late treatment. In this thesis, we focus on children with Turner's syndrome, celiac disease, cystic fibrosis, hypernatraemic dehydration and obesity. For most of these conditions, it is known that early treatment improves outcomes. Early treatment with growth hormone has proved its efficacy in the treatment of various conditions with short stature.[13] Girls with Turner's syndrome have higher disease risks, especially due to the risk of dissection of the aorta and other cardiovascular diseases, as well as the risk of type 2 diabetes, osteoporosis and thyroid disease.[14] Early detection of girls with Turner's syndrome enables the pediatrician and cardiologist to evaluate these children. Girls with Turner's syndrome should be screened for hypertension and electrocardiographic abnormalities in addition to anatomic anomalies. Blood pressure should be monitored on a regular basis.[15] For children with celiac disease, early detection and treatment with a gluten-free diet is required to improve the immediate quality of life of the patients and to decrease the long-term risks, including a higher prevalence of malignancies, adverse pregnancy outcome, neurological problems and osteomalacia.[16] For children with cystic fibrosis, treatment may include pulmonary therapy (treatments to maintain lung function) and nutritional therapy. For infants with hypernatraemic dehydration, early detection is needed to prevent serious complications, such as fits, disseminated intravascular coagulation, multiple cerebrovascular accidents, and even death. Evidence on the long-term effects of treating childhood obesity is still limited.[17]

4. The fourth group deals with the acceptability of the screening program. An important element of this group is that the screening program should be cost-effective. Or in other words, screening should be worth the money. The direct costs of growth monitoring include equipment, staff, training of staff and the costs associated with referral for further investigation. A more demanding test is more expensive, but may also have a higher yield. Studies on economic modeling suggest that growth monitoring is associated with health improvements and is cost-effective.[3] In general, for an adequate evaluation of the cost-effectiveness, we need insight into the diagnostic performance of various referral criteria and the costs that are associated with this. From a societal perspective, a high specificity for the referral criteria is desirable to minimize unwanted health costs, to free clinical practitioners from being overloaded by work, to evade unnecessary interventions and treatments for healthy children, and to reduce parental and child's anxiety.

In summary, many steps have to be taken to investigate if growth monitoring fulfils the criteria of a screening program. This thesis will focus on the properties of growth monitoring as a screening test.

What is already known about growth of children and referral criteria for growth, and what new contribution will this thesis make? We now deal with the topics 'growth of children' and 'referral criteria for growth' in some more detail.

Growth of children

Growth is commonly used as a term to present height or weight gain. To study an individual's growth rate, one needs multiple (longitudinal) measurements. Growth can also be studied by a single (cross-sectional) measurement per individual. Most research has been done on cross-sectional data, producing growth charts for the normal development, syndrome-specific growth charts and growth standards representing the 'healthy' growth of a population.

Growth charts for the normal development

A growth chart gives a graphic presentation of how children normally grow. In the Netherlands, four nation-wide surveys have been performed to establish reference growth charts.[18-21] The first nation-wide growth study took place in 1955 (N=16,910), the second in 1965 (N=54,776), the third in 1980 (N=42,000) and the fourth in 1997 (N=14,500). The fourth study also included children of Moroccan and Turkish origin living in the Netherlands.[22,23] Since these children were considerably shorter than Dutch children, specific reference growth charts for Moroccan and Turkish children were published. As the growth in preterm infants differs from term infants, prenatal (intrauterine) reference curve charts for Swedish preterm infants were published.[24] These charts are currently used in clinical practice in the Netherlands to correct for gestational age. It is known that the growth pattern of twins differs from that of singletons during infancy, especially during the first years of life. However, no reference growth charts for twins are available. This thesis presents reference growth charts for Dutch monozygotic and dizygotic twins. These charts may be helpful to detect pathologies in twins. Weight loss during the first weeks after birth is an important indicator for dehydration of the infant. However, there are no reference charts available for weight loss in the neonatal period. This thesis presents a reference chart for weight loss in breast-fed neonates to detect neonates with (hypernatraemic) dehydration.

Syndrome-specific growth charts

There are also growth charts available for children with a syndrome that is known to affect growth. These include Down's syndrome (short stature, overweight)[25], Turner's syndrome (short stature)[26], Noonan syndrome (short stature)[27], Prader-Willi syndrome (failure to thrive at infancy, short stature, obesity)[28], Silver-Russell syndrome (low birth weight, poor growth)[29], Cri-du-chat syndrome (low birth weight, poor growth)[30], Williams syndrome (short stature, low head circumference)[31] and achondroplasia (short stature, overweight).[32-34] These charts are useful for comparing individual children to other children with the same diagnosis, and, perhaps, to detect other pathologies within

this group. It is recommended that they should be used in combination with the nation-wide reference charts.

Standards and references

There is an important distinction between a growth reference and a growth standard. A reference describes its population without making any claims about the health of its population, whereas a standard represents 'healthy' growth of a population and suggests a target to achieve.[35-37] Most growth charts describe existing growth patterns and are for this reason references, not prescriptive standards. Disorder-specific growth charts are typically used as references. In 2006, the World Health Organization (WHO) published an international growth standard representative of children (birth to age five) being raised according to recommended health practices.[38-40] These conditions include: exclusive or predominant breastfeeding for four to six months, complementary foods by six months, continued breastfeeding for 12 months or more; an optimal environment without conditions that could limit growth (smoking, altitude >1500 meters); and optimal health care (immunizations, good routine pediatric care). Although the WHO outline how children should grow, there is some controversy whether the WHO growth chart applies to all countries of the world.[41]

Presentation of growth charts

The presentation of growth charts differs between countries. Some countries present their growth charts in percentiles, while others use Standard Deviation Scores (SDS). Percentiles indicate a child's position within the context of the reference population. As an example, if an infant has a length-for-age at the 3rd percentile, then 3% of the population who are at the same age and sex are shorter than that infant. This example shows that percentiles are easy to explain. A disadvantage of percentiles is that a measurement well below the 3rd percentile cannot accurately be defined on the growth chart. Another disadvantage is that the distances between the percentiles are not equally distributed: for example the difference in centimeters between the 10th and the 20th percentile is larger than between the 20th and the 30th percentile. The SDS is an alternative that expresses the measurement relative to a reference population in units of standard deviations above or below the median.[42] WHO growth charts are available that present the -3, -2, 0, +2, +3 SDS lines. The Dutch growth charts show the -2.5, -2, -1, 0, +1, +2 and +2.5 SDS lines. The 0 SDS line represents the median. At every age, 50% of children in the reference population fall above this line and 50% below it. Most children (95%) fall between the –2 and +2 SDS lines. In this thesis we construct referral criteria based on SDS.

Referral criteria for growth

It is well-known that some conditions affect growth. Therefore, a wide variety of referral criteria and guidelines have been proposed to detect the children with growth-related conditions. We now discuss some of the referral criteria.

Short stature

Two consensus guidelines have been published to refer children with short stature.[43,44] The UK consensus recommends referral based on a single height measurement (<0.4th centile) at or around time of school entry at the age of 5 years. The Dutch consensus uses referral criteria based on a single height measurement (<-2.5 SDS), a height measurement corrected for parental height (<-1.3 SDS and height >1.3 SDS below target height) and a slow height gain (deflection of >0.25 SDS per year or >1 SDS over several years) between birth and 10 years of age. This thesis deals with the validity of several referral criteria for short stature. The validity of referral criteria is based on children with Turner's syndrome, celiac disease, cystic fibrosis and children from the general population.

Failure to thrive

Failure to thrive (FTT) is a very general term, but it is mostly used for slow weight or height gain during infancy and early childhood. There is little consensus on the choice of criteria for FTT.[45] Some authors define FTT as weight or height falling below the third or fifth centile, or falling two major centiles of the standard National Center for Health Statistics (NCHS) growth chart. Others state that malnutrition (weight <80% of ideal body weight for age) should be present to state that a child is failing to thrive.[46,47] The most important cause for FTT is that children don't receive or are unable to take in, retain, or utilize the calories needed to gain weight. In this thesis we investigate the diagnostic performance of FTT in children with hypernatraemic dehydration, celiac disease, cystic fibrosis and children from the general population.

Body Mass Index

Body Mass Index (BMI: kg/m^2) has become the de facto standard for defining childhood overweight and obesity, but there is little consensus about the precise criteria. The International Obesity Task Force (IOTF) proposed a definition that uses childhood BMI centiles linked to adult cut-off points, with 25 kg/m^2 as adult overweight, and 30 kg/m^2 as adult obesity.[48] Children that have a BMI above this cut-off point are defined as overweight or obese. Another definition was developed by the Centers of Disease Control and prevention (CDC) and is based on the 85th (overweight) and 95th (obesity) centiles of childhood BMI on a nationally representative survey in the United States (US).[49] The UK uses IOTF related cut-off points and those are equal to the 91st (overweight) and 98th (obesity) centiles of the UK chart. In Germany overweight and obesity are defined as respectively the 90th and 97th centiles of their growth chart. Recently, the WHO has proposed a new definition as +1 SD (overweight) and +2 SD (obesity) of BMI on the new WHO standards.[50] This thesis assesses the prevalence of overweight and obesity based on BMI with IOTF cut-off points in children living in the Netherlands. A complicating factor is the height bias in BMI. With BMI, tall children are proportionally more overweight than short children.[51,52] Some have argued that the increased obesity

risk in tall children has a physiological cause, but the evidence is incomplete. This thesis will address the impact of height bias on overweight prevalence.

Outline of this thesis

Chapter 2 presents new growth charts for Dutch monozygotic and dizygotic twins aged 0-2.5 years relative to singletons. **Chapter 3** considers the growth process of an individual as the phenotypic expression of her or his genotype and the influence of environmental factors. The contributions of genetic and environmental factors on variation in length (height) and weight are estimated from birth to 2.5 years of age. **Chapter 4** reports on the diagnostic performance of growth monitoring in detecting girls with Turner's syndrome, a chromosomal disorders that occurs in about 1 of 2500 female live births and that leads to seriously retarded height. **Chapter 5** deals with the question whether a mixed model approach leads to a better detection of girls with Turner's syndrome than conventional referral criteria for growth monitoring. New referral criteria incorporate information on the parameters of the mixed model, on parental height and on gestational age. **Chapter 6** presents an evidence-based guideline for the referral of short stature. Several referral criteria are formulated and applied to longitudinal growth data from children with Turner's syndrome, celiac disease, cystic fibrosis and to several samples from the general population. **Chapter 7** and **Chapter 8** address the diagnostic performance of body weight for the detection of children with celiac disease and cystic fibrosis, respectively. Celiac disease, also known as gluten-sensitive enteropathy, is an illness in which failure to thrive may be the earliest signs of the disease. Cystic fibrosis is one of the most common life-threatening autosomal recessive diseases in the Caucasian population, where early diagnosis is of great importance. Both chapters present evidence-based referral criteria to detect children with celiac disease and cystic fibrosis. **Chapter 9** proposes an evidence-based guideline to detect breast-fed infants with hypernatraemic dehydration. This guideline is based on a newly developed reference chart for weight loss by age between postnatal days 2 and 11. **Chapter 10** assesses the prevalence of overweight and obesity in children living in the Netherlands, and compares the findings with the third and fourth nation-wide surveys carried out in 1980 and 1997, respectively. Overweight and obesity are defined by the IOTF cut-off points for BMI. **Chapter 11** is an account of the impact of height bias in childhood BMI on overweight prevalence. **Chapter 12** contains a summary of the results and recommendations.

References

1. de Onis M, Habicht JP. Anthropometric reference data for international use: recommendations from a World Health Organization Expert Committee. Am J Clin Nutr 1996;64:650-8.
2. Hall DM. Growth monitoring. Arch Dis Child 2000;82:10-5.
3. Fayter D, Nixon J, Hartley S, Rithalia A, Butler G, Rudolf M, Glasziou P, Bland M, Stirk L, Westwood M. Effectiveness and cost-effectiveness of height screening programmes during the primary school years: a systematic review. Arch Dis Child 2008;93:278–84.
4. Panpanich R, Garner P. Growth monitoring in children. Cochrane Database. Syst Rev 2000;2:CD001443.
5. de Onis M, Wijnhoven TM, Onyango AW. Worldwide practices in child growth monitoring. J Pediatr 2004;144:461-5.
6. Grote FK, Oostdijk W, de Muinck Keizer-Schrama SM, Dekker FW, Verkerk PH, Wit JM. Growth monitoring and diagnostic work-up of short stature: an international inventorization. J Pediatr Endocrinol Metab 2005;18:1031-8.
7. Gray JA. New concepts in screening. Br J Gen Pract 2004;54:292–8.
8. De Muinck Keizer-Schrama SM: [Consensus 'diagnosis of short stature in children.' National Organization for Quality Assurance in Hospitals]. Ned Tijdschr Geneeskd 1998;142:2519-25.
9. van Buuren S, Bonnemaijer-Kerckhoffs DJ, Grote FK, Wit JM, Verkerk PH. Many referrals under Dutch short stature guidelines. Arch Dis Child 2004;89:351-2.
10. Saunders NR, Shouldice M. Health Maintenance Visits: A Critical Review. In: Feldman W, ed. Evidence-Based Pediatrics. Hamilton: B.C. Decker Inc., 2000.
11. Gardosi J, Chang A, Kalyan B, Sahota D, Symonds EM. Customised antenatal growth charts. Lancet 1992;339:283-7.
12. Mongelli M, Gardosi J. Longitudinal study of fetal growth in subgroups of a low-risk population. Ultrasound Obstet Gynecol 1995;6:340-4.
13. Krysiak R, Gdula-Dymek A, Bednarska-Czerwińska A, Okopień B. Growth hormone therapy in children and adults. Pharmacol Rep 2007;59:500-16.
14. Hjerrild BE, Mortensen KH, Gravholt CH. Turner syndrome and clinical treatment. Br Med Bull 2008;86:77-93.
15. Bondy CA. Congenital cardiovascular disease in Turner syndrome. Congenit Heart Dis 2008;3:2-15.
16. Goddard CJ, Gillett HR. Complications of coeliac disease: are all patients at risk? Postgrad Med J 2006;82:705-12.

17. Summerbell CD, Ashton V, Campbell KJ, Edmunds L, Kelly S, Waters E. Interventions for treating obesity in children. In The Cochrane Library 2005, Issue 3. Chichester: Wiley; 2005.
18. Roede MJ, van Wieringen JC. Growth diagrams 1980; Netherlands third nationwide survey. Tijdschrift voor Sociale Gezondheidszorg 1985;63(suppl.).
19. Fredriks AM, van Buuren S, Burgmeijer RJ, Meulmeester JF, Beuker RJ, Brugman E, Roede MJ, Verloove-Vanhorick SP, Wit JM. Continuing positive secular growth change in The Netherlands 1955-1997. Pediatr Res 2000;47:316-23.
20. de Wijn JF, de Haas JH. Groeidiagrammen van 1-25 jarigen in Nederland [Growth diagrams for ages 1-25 years in the Netherlands]. Netherlands Instituut voor Praeventieve Geneeskunde, 1960, Leiden.
21. van Wieringen JC, Wafelbakker F, Verbrugg HP, de Haas JH. Growth diagrams 1965 Netherlands. Nederlands Instituut voor Praeventieve Geneeskunde/Wolters-Noordhoff, 1971, Leiden/Groningen.
22. Fredriks AM, Van Buuren S, Sing RA, Wit JM, Verloove-Vanhorick SP.Height, weight, body mass index and pubertal development reference values for children of Turkish origin in the Netherlands. Eur J Pediatr 2003;162:788-93.
23. Fredriks AM, van Buuren S, Jeurissen SE, Dekker FW, Verloove-Vanhorick SP, Wit JM. Height, weight, body mass index and pubertal development references for children of Moroccan origin in The Netherlands. Acta Paediatr 2004;93:817-24.
24. Niklasson A, Albertsson-Wikland K. Continuous growth reference from 24th week of gestation to 24 months by gender. BMC Pediatr 2008;8:8.
25. Styles ME, Cole TJ, Dennis J, Preece MA.New cross sectional stature, weight, and head circumference references for Down's syndrome in the UK and Republic of Ireland. Arch Dis Child 2002;87:104-8.
26. Rongen-Westerlaken C, Corel L, van den Broeck J, Massa G, Karlberg J, Albertsson-Wikland K, Naeraa RW, Wit JM. Reference values for height, height velocity and weight in Turner's syndrome. Swedish Study Group for GH treatment. Acta Paediatr 1997;86:937-42.
27. Witt DR, Keena BA, Hall JG, Allanson JE. Growth curves for height in Noonan syndrome. Clin Genet 1986;30:150-3.
28. Grauer ML, Wollmann HA, Schulz V, Ranke MB. Reference values for height and weight in Prader–Willisyndrome based on 315 patients. Hormone Research 1997;48(Suppl. 2):54.
29. Wollman HA, Kirchner T, Enders H, Preece MA, Ranke MB. Growth and symptoms in Silver–Russell syndrome: review on the basis of 386 patients. Eur J Pediatr 1995; 154:958–68.
30. Marinescu RC, Mainardi PC, Collins MR, Kouahou M, Coucourde G, Pastore G, Eaton-Evans J, Overhauser J. Growth charts for cri-du-chat syndrome: an international collaborative study. Am J Med Genet 2000;94:153-62.

31. Martin ND, Smith WR, Cole TJ, Preece MA. New height, weight and head circumference charts for British children with Williams syndrome. Arch Dis Child 2007;92:598-601.
32. Hoover-Fong JE, McGready J, Schulze KJ, Barnes H, Scott CI. Weight for age charts for children with achondroplasia. Am J Med Genet 2007;143:2227-35.
33. Hunter AG, Hecht JT, Scott CI Jr. Standard weight for height curves in achondroplasia. Am J Med Genet 1996;62:255-61.
34. Horton WA, et al. Standard growth curves for achondroplasia. J Pediatr 1978;93:435-8.
35. World Health Organization. Working Group. Use and interpretation of anthropometric indicators of nutritional status. Bulletin of the WHO 1983;64:929-41.
36. Cole TJ. The Use of Growth References. In: Ulijaszek SJ JF, Preece MA, ed. The Cambridge Encyclopedia of Human Growth and Development. Cambridge, UK: Cambridge University Press, 1998:80-9.
37. Garza C, Frongillo E, Dewey KG. Implications of growth patterns of breast-fed infants for growth references. Acta Paediatr Suppl 1994;402:4-10.
38. Garza C, de Onis M. A new international growth reference for young children. Am J Clin Nutr 1999;70 (suppl):169S-172S.
39. Grummer-Strawn LM, Garza C, Johnson CL. Childhood growth charts. Pediatrics 2002;109:141-2.
40. WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards based on length/height, weight and age. Acta Paediatr 2006;450:76–85.
41. van Buuren S, van Wouwe JP. WHO Child Growth Standards in action. Arch Dis Child 2008;93:549-51.
42. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med 1992;11:1305-19.
43. Hall DMD, Elliman D. Health for all children, 4th edn. Oxord: Oxford University Press, 2003.
44. van Buuren S, Fredriks AM, Verkerk PH. Comment on ‘‘Consensus ‘Diagnostiek kleine lichaamslengte bij kinderen’’ [Comment on Dutch consensus guidelines for short stature]. Ned Tijdschr Geneeskd 1999;143:1585–6.
45. Steward DK, Ryan-Wenger NA, Boyne LJ. Selection of growth parameters to define failure to thrive. J Pediatr Nurs 2003;18:52-9.
46. Bithoney WG, Dubowitz H, Egan H. Failure to thrive/growth deficiency. Pediatr Rev 1992;13:453-60.
47. Bassali RW, Benjamin J. Failure to Thrive. Emedicine, August 11, 2004.
48. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ 2000;320:1240-3.

49. Barlow SE, Dietz WH. Obesity evaluation and treatment: Expert Committee recommendations. The Maternal and Child Health Bureau, Health Resources and Services Administration and the Department of Health and Human Services. Pediatrics 1998;102:E29.
50. de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. Bulletin of the World Health Organization 2007;85:660–7.
51. Freedman DS, Thornton JC, Mei Z, Wang J, Dietz WH, Pierson RN Jr et al. Height and adiposity among children. Obes Res 2004;12:846-53.
52. Buchan IE, Bundred PE, Kitchiner DJ, Cole, TJ. Body mass index has risen more steeply in tall than in short 3-year olds: serial cross-sectional surveys 1988-2003. Int J Obes 2007;31:23-9.

Growth references for height, weight and body mass index of twins aged 0-2.5 years

2

Paula van Dommelen[1,2], Mathisca C.M. de Gunst[1], Aad W. van der Vaart[1], Stef van Buuren[2,3] and Dorret I. Boomsma[1]

[1]VU University, Amsterdam, The Netherlands
[2]TNO Quality of Life, Leiden, The Netherlands
[3]University of Utrecht, Utrecht, The Netherlands

Acta Paediatr 2008;97:1099-104.

Abstract

Aim: To determine the size of the growth deficit in Dutch monozygotic and dizygotic twins aged 0-2.5 years as compared to singletons and to construct reference growth charts for twins.
Methods: Growth of twins was studied using longitudinal data on over 4000 twins aged 0-2.5 years of the Netherlands Twin Register. The LMS method was used to obtain growth references for length/height, weight, and body mass index (BMI) for twins.
Results: During the first 2.5 years of age, differences in length/height and weight between twins and singletons decrease but do not disappear. BMI of twins deviates less from that of singletons. Approximately half of the growth retardation from birth until 1.5 years of age was attributable to gestational age. Between 1.5 years and 2.5 years of age, this difference was reduced to one-third. Thus, a substantial part of the growth difference could not be explained by gestational age.
Conclusion: During the first 2.5 years of life, there is a difference in growth between twins and singletons. Twins catch up in their body size, i.e. they grow faster after birth, but do not yet achieve the same height and weight till they reach 2.5 years of age. We recommend the use of the growth references for twins.

Introduction

Monitoring children's height is a standard procedure in many countries. Children's height is measured in order to diagnose abnormal height growth rates and to monitor the results of any treatment for such conditions. Weight is another important growth parameter, which provides information about the individual's nutritional status. The Quetelet Index (body mass index (BMI) = kg/m^2) is used to identify cases of overweight or underweight and to monitor nutritional status. The increase in height and weight during infancy has a strong correlation with gestational age and with the growth during pregnancy. A study on Australian twins and singletons concluded that the twins exhibit slower growth in comparison to singletons from week 26 of pregnancy until birth.[1] In addition, twin pregnancies were about 3 weeks shorter than the singleton pregnancies, resulting in low birth weight. The same was found in Dutch twins, where the mean birth weight of twins was almost 1 kg less than that of singletons.[2] Belgian twins showed a reduction in weight gain from week 32 of pregnancy onwards. For height this effect was not seen until week 39.[3] The Belgian study signalled that height and weight of twins cannot be compared with those of singletons. Also, the American twins were found to lag at birth, both in terms of height and weight. They rapidly caught up in weight during the first 3 months, whereas height took much longer.[4] The literature cited above suggests that during infancy, the growth pattern of twins differs from that of singletons. Therefore,

there is need for special growth charts for twins. In order to fill this gap, we investigated the size of the growth deficit in Dutch monozygotic and dizygotic twins from birth to 2.5 years of age. We compared longitudinal data from Dutch twins with reference charts for the Netherlands dating from 1997.[5,6] Reference charts for twins were constructed. Abnormalities in the growth of twins can now be identified more effectively with these new charts than would be possible by using the standard references.

Patients and methods

The data were derived from the Netherlands Twin Register (NTR) at VU University, Amsterdam. Longitudinal length/height and weight measurements were obtained in post-natal clinics between birth and approximately 2.5 years of age of the twins born during the period 1986-1992.[7-9] A child was included if it was measured on at least one occasion between birth and 2.7 years of age and suffered no severe handicaps. Twins were analyzed as two separate individuals.[10]

For length/height, the data consisted of 1420 monozygotic boys (MZB), 1580 monozygotic girls (MZG), 2669 dizygotic boys and boys from boy-girl twin pairs (DZB) and 2623 dizygotic girls and girls from boy-girl twin pairs (DZG). For weight, we had 1428 MZB, 1583 MZG, 2677 DZB and 2630 DZG. For BMI, there were 1418 MZB, 1577 MZG, 2665 DZB and 2618 DZG. Most of the children were measured on 9-12 occasions.

The LMS method was used to determine the reference lines for length/height, weight and BMI.[11] The principle behind this method is that, following a suitable transformation, the data show a standard normal distribution.We refer to a transformed data point as a standard deviation score (SDS). In the LMS method, this transformation involves the use of three age-dependent curves. These are the skewness curve (L), the median curve (M), and the coefficient of variation curve (S). In order to obtain smooth and accurate L, M, and S curves, the method uses the standard likelihood function with a penalty term for lack of smoothness (maximum penalized likelihood).[12] Worm plots were used to check the normality of the SDS.[13] The LMS Pro program (version 1.16, dated 15 April 2002) was used for the calculations involved in the LMS method.[14] The worm plots were made by using S-plus 2000. For length/height and weight, age was scaled in the way it expanded during periods of rapid growth and compressed during periods of slow growth. For BMI, a power transformation was used, using 0.33 (for boys) and 0.25 (for girls) with zero offset.[13] Children with retarded growth are likely to visit post-natal clinics more often. In order to prevent short children from becoming over-represented in the LMS analyses a weighting factor was calculated for all measurements. This weighting factor was defined per child as the inverse of the number of occasions on which that child

was measured. When L, M and S references for twins are available, each measurement can be converted into SDS. SDS of measurement x is calculated as $((x/M)^{L}-1)/LS$ (when $L\neq0$) or $\ln(x/M)/S$ (when $L=0$). This SDS expresses the measurement in relation to twins in units of standard deviations above or below the median and is useful to detect trends in both mean and variability. Growth anomalies in twins were calculated in relation to the Dutch 1997 references.

To investigate the deficit of SDS corrected for gestational age, we applied the prenatal (intrauterine) reference curve according to Niklasson et al. in preterm infants.[15] This curve was used to express SDS till the age corresponding with 40 weeks of gestation. Between 40 and 42 weeks an interpolation between this curve and that of the 1997 Dutch references was used. From 42 weeks of gestation till the age of 2 years, SDS was calculated on ages corrected for gestation, using the 1997 Dutch references. The SDS for the term infants was based on the 1997 Dutch references.

Results

Table 1 contains mean length/height, weight and BMI on the SDS scale for various age groups. During the first 6 months , the mean length and weight deficit was equal to -1.3 to -1.4 SDS (10th percentile) for monozygotic twins in relation to the reference population of singletons. For dizygotic twins, length and weight deficit ranged from -1.2 to -1.3 SDS. Twins catch up part of the growth deficit in later life. Between 0.5 and 1.5 years, mean SDS had increased to -0.6 SDS (approximately the 25th percentile), and between 1.5 and 2.5 years it reached approximately -0.3 SDS (approximately the 35th percentile). At that point, dizygotic girls, in particular, had nearly reached the reference level. For BMI, mean SDS was substantially closer to the mean of the reference population than for height and weight (see Table 1).

Table 2 shows the magnitude of the contribution made by gestational age relative to the standard reference population. For the twins, whose age of gestation was known to range from 39 to 41 weeks, i.e. for term births, (954 children), mean SDS was calculated for three age groups. We observed a deviation of -0.6 to -0.7 SDS for the length and the weight throughout the first 6 months. One year later, that deviation was -0.3 to -0.4 SDS. That means that approximately half of the size of the deviation seen throughout the first 18 months can be attributed to gestational age. During the period from 18 months to 2.5 years, this was reduced to one-third. When applying the prenatal reference curve for the preterm infants and the 1997 Dutch references for the term infants, the mean (SD) length/height SDS corrected for gestational age was -0.52 (1.04) for age group <0.5 year, -0.25 (0.97) for age group 0.5-1.4 years and -0.17 (1.03) for age group 1.5-2.5 years, and for weight was -0.63 (1.08) for age group <0.5 year, -0.37 (0.97) for age group 0.5-1.4 years

and -0.20 (0.99) for age group 1.5-2.5 years. This is in agreement with the results shown in Table 2. Accordingly, a correction for premature birth alone will not be enough to render the growth of twins comparable to the growth of singletons.

Table 1 *Mean standard deviation score (SDS) of Dutch twins for length/height, weight and BMI relative to the 1997 Dutch references.*

	Monozygotic		Dizygotic		Total
	Boys	Girls	Boys	Girls	
Length/height					
< 0.5 year###	-1.37 (1.17)	-1.31 (1.26)	-1.20 (1.13)	-1.16 (1.18)	-1.24 (1.18)
0.5-1.4 year##	-0.59 (0.96)	-0.63 (1.01)	-0.56 (0.99)	-0.54 (1.01)	-0.57 (0.99)
1.5-2.5 year#	-0.33 (1.01)	-0.33 (1.07)	-0.33 (1.01)	-0.22 (1.04)	-0.30 (1.03)
Weight					
< 0.5 year###	-1.43 (1.12)	-1.37 (1.24)	-1.33 (1.09)	-1.25 (1.15)	-1.33 (1.15)
0.5-1.4 year	-0.66 (0.91)	-0.57 (1.00)	-0.65 (0.93)	-0.54 (0.96)	-0.60 (0.95)
1.5-2.5 year	-0.36 (0.95)	-0.26 (1.03)	-0.31 (0.96)	-0.22 (0.97)	-0.28 (0.98)
BMI					
< 0.5 year	-0.57 (1.01)	-0.56 (1.01)	-0.59 (1.02)	-0.53 (0.95)	-0.56 (0.99)
0.5-1.4 year	-0.35 (0.90)	-0.18 (0.94)	-0.36 (0.96)	-0.22 (0.88)	-0.28 (0.92)
1.5-2.5 year	-0.10 (0.96)	-0.03 (1.01)	-0.05 (0.99)	-0.08 (1.02)	-0.07 (1.00)

Statistically significant between monozygotic and dizygotic twins: ### p<0.005, ##p<0.001(girls), #p<0.01 (girls).
The standard deviation is shown in parentheses.

Table 2 *Mean standard deviation score (SDS) of Dutch twins whose age of gestation was known to range from 39 to 41 weeks, for length/height, weight and BMI in relation to the 1997 Dutch references. The standard deviation is shown in parentheses.*

	Length/height SDS	Weight SDS	BMI SDS
< 0.5 year	-0.59 (0.89)	-0.73 (0.89)	-0.39 (0.91)
0.5-1.4 year	-0.31 (0.93)	-0.39 (0.88)	-0.21 (0.86)
1.5-2.5 year	-0.10 (1.01)	-0.09 (0.93)	0.03 (0.98)

The difference in length and weight from birth to 6 months of age between monozygotic and dizygotic twins was small but statistically significant in both girls and boys (see Table 1). This difference in SDS varied between 0.10 and 0.17. This is a relatively small difference and therefore in clinical practice it is recommended to use the reference charts based on the dizygotic twins, as most twins are dizygotic (see Figs. 1-2). However, if a computer-based system is available in child health care, we recommend to use the L, M, S values for length and weight for both monozygotic and dizygotic twins, as length and weight of monozygotic twins are systematically lesser than that of dizygotic twins (see Table 3).

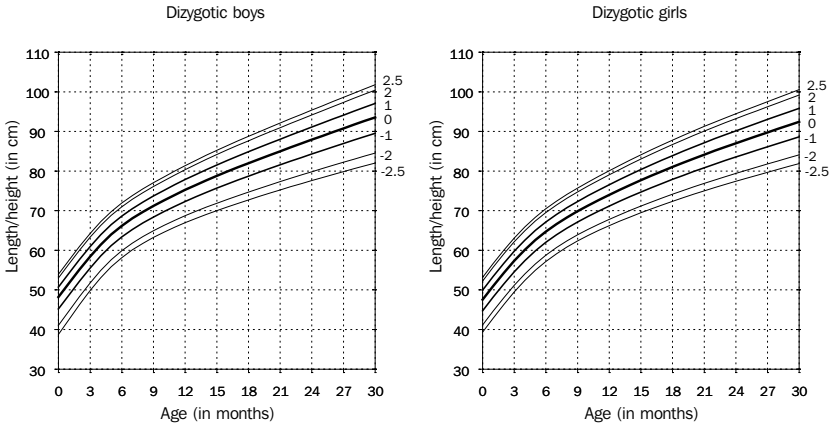


Figure 1 *Reference charts for length/height of twins in the Netherlands: dizygotic boys and girls from birth to 2.5 years of age; the following curves are shown -2.5 SDS (= P0.6), -2 SDS (= P2), -1 SDS (= P16), 0 SDS (= P50 = median), 1 SDS (= P84), 2 SDS (= P98) and 2.5 SDS (=P99.4).*

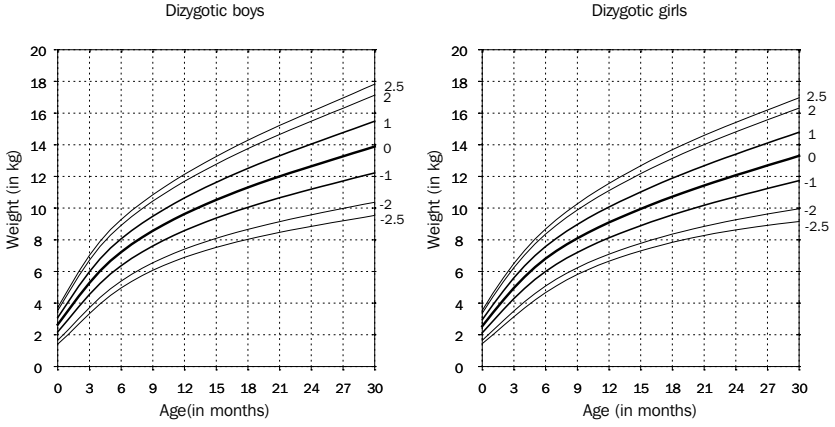


Figure 2 *Reference charts for weight of twins in the Netherlands: dizygotic boys and girls from birth to 2.5 years of age; the following curves are shown -2.5 SDS (= P0.6), -2 SDS (= P2), -1 SDS (= P16), 0 SDS (= P50 = median), 1 SDS (= P84), 2 SDS (= P98) and 2.5 SDS (= P99.4).*

It is noteworthy that our twin length references are skewed to the left (i.e. L > 1), while length references are usually normally distributed (e.g. the Dutch 1980 or 1997). Furthermore, this left skewness disappears with age in the monozygotic twins, but remains unchanged in the dizygotes. This might indicate that the monozygic twins who are born short are more likely to catch up over a period of time than the dizygotic twins.

We tested this by comparing the growth velocity of height between monozygotic and dizygotic twins born short (< -1 SDS using twin references) on a subpopulation of twins previously described in van Dommelen et al.[16] We found that monozygotic twins who were born short had a slightly higher (+0.29 to +0.40) growth velocity compared to dizygotic twins, although not statistically significant. Thus, the analysis did not provide evidence in favor of the suggestion that monozygtic twins are more likely to catch up. The difference in BMI between the type of twins was not statistically significant, and we have therefore constructed only one reference chart based on the dizygotic twins (see Fig. 3). The LMS values for BMI are shown in Table 4.

Table 3 *Reference values (LMS) for length/height and weight in monozygotic and dizygotic twins aged 0-30 months.*

		Monozygotic twins						Dizygotic twins					
		Length/height			weight			Length/height			weight		
Sex	Age	L	M	S	L	M	S	L	M	S	L	M	S
boys	0	5.2	47.61	0.055	1.6	2.51	0.175	4.2	48.21	0.057	1.3	2.65	0.170
	1	5.0	51.20	0.052	1.5	3.44	0.161	4.2	51.72	0.053	1.3	3.55	0.158
	2	4.7	54.71	0.049	1.4	4.36	0.148	4.2	55.18	0.049	1.3	4.44	0.146
	3	4.5	58.03	0.046	1.4	5.23	0.137	4.2	58.47	0.046	1.3	5.29	0.137
	4	4.3	61.03	0.044	1.3	6.00	0.128	4.2	61.43	0.044	1.3	6.03	0.129
	5	4.1	63.64	0.042	1.3	6.66	0.120	4.2	63.99	0.041	1.3	6.67	0.123
	6	3.9	65.90	0.040	1.2	7.22	0.115	4.2	66.18	0.040	1.3	7.23	0.118
	8	3.7	69.58	0.037	1.2	8.14	0.109	4.2	69.71	0.038	1.3	8.16	0.112
	10	3.4	72.55	0.036	1.1	8.91	0.106	4.2	72.62	0.037	1.3	8.94	0.109
	12	3.2	75.13	0.036	1.1	9.58	0.106	4.2	75.25	0.037	1.3	9.63	0.108
	18	2.7	81.96	0.036	0.9	11.26	0.109	4.2	81.98	0.038	1.3	11.30	0.109
	24	2.3	87.80	0.038	0.8	12.56	0.114	4.2	87.91	0.039	1.3	12.65	0.113
	30	1.9	93.08	0.040	0.8	13.67	0.119	4.2	93.56	0.040	1.3	13.89	0.118
girls	0	5.4	47.37	0.051	1.6	2.45	0.176	3.6	47.53	0.055	1.2	2.54	0.163
	1	5.1	50.71	0.049	1.4	3.30	0.164	3.6	50.95	0.051	1.3	3.38	0.152
	2	4.8	53.95	0.047	1.3	4.13	0.152	3.6	54.24	0.048	1.4	4.20	0.141
	3	4.5	57.02	0.045	1.2	4.91	0.142	3.6	57.33	0.046	1.5	4.96	0.132
	4	4.3	59.83	0.043	1.1	5.60	0.133	3.6	60.12	0.043	1.5	5.66	0.125
	5	4.0	62.31	0.042	1.0	6.22	0.127	3.6	62.58	0.041	1.5	6.27	0.120
	6	3.8	64.48	0.040	0.9	6.76	0.122	3.6	64.72	0.040	1.5	6.81	0.116
	8	3.4	68.10	0.039	0.7	7.67	0.116	3.6	68.31	0.038	1.3	7.71	0.111
	10	3.1	71.07	0.038	0.6	8.41	0.114	3.6	71.31	0.037	1.1	8.45	0.108
	12	2.8	73.72	0.037	0.5	9.04	0.114	3.6	73.98	0.036	1.0	9.09	0.107
	18	2.1	80.63	0.037	0.2	10.64	0.115	3.6	81.04	0.037	0.9	10.73	0.109
	24	1.5	86.70	0.039	0.0	11.96	0.116	3.6	87.02	0.038	1.1	12.07	0.112
	30	1.0	92.24	0.040	-0.3	13.18	0.118	3.6	92.44	0.039	1.4	13.30	0.116

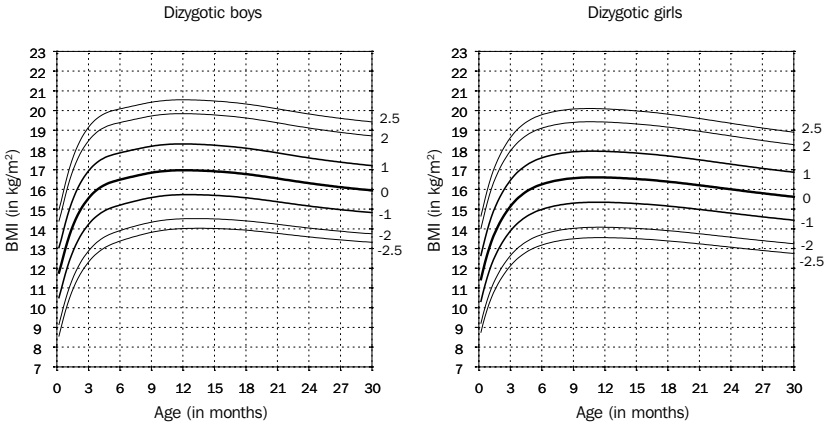


Figure 3 *Reference charts for Body Mass Index of twins in the Netherlands: dizygotic boys and girls from birth to 2.5 years of age; the following curves are shown -2.5 SDS (= P0.6), -2 SDS (= P2), -1 SDS (= P16), 0 SDS (= P50 = median), 1 SDS (= P84), 2 SDS (= P98) and 2.5 SDS (= P99.4).*

Table 4 *Reference values (LMS) for BMI for all twins aged 0-30 months.*

			Boys			Girls
Age (mo)	L	M	S	L	M	S
0	1.4	11.76	0.124	0.3	11.44	0.122
1	0.8	13.35	0.098	0.3	13.04	0.093
2	0.6	14.72	0.091	0.3	14.32	0.088
3	0.5	15.55	0.088	0.3	15.14	0.085
4	0.4	16.05	0.085	0.3	15.69	0.083
5	0.4	16.33	0.082	0.3	16.04	0.082
6	0.3	16.49	0.081	0.3	16.26	0.081
8	0.2	16.74	0.078	0.3	16.51	0.079
10	0.1	16.92	0.077	0.3	16.60	0.079
12	0.0	16.97	0.076	0.3	16.61	0.078
18	-0.2	16.78	0.075	0.3	16.39	0.078
24	-0.3	16.32	0.075	0.3	16.01	0.078
30	-0.5	15.95	0.075	0.3	15.62	0.078

Discussion

Differences in length/height, weight and BMI between twins and singletons decline during the first 2.5 years, but do not disappear completely. Part of these differences remains even after correcting for premature birth. Accordingly, there is a genuine need for special growth charts for twins. This study has developed growth references specifically for twins.

The new growth charts are based on Dutch twins. The WHO Multicentre Growth Study detected only small differences for height and weight in children up to the age of 2 years among different populations.[17] It seems likely that this would be similar to twins. Given the large growth deficit in twins during early age, we advise to use twin-specific references rather than reverting to the countries own reference for singletons with or without correction. We, therefore, recommend the twin references as presented here for application to twins in other populations. For East-Asian countries, we cannot give the same advice as Hur et al. reported that the total phenotypic variances of birthweight were about 45% larger in Caucasians than in East Asians.[18] Therefore, East-Asian twins might grow differently than Caucasian twins. Growth charts for Japanese twins are available and we advise to use these for countries in East Asia.[19] No LMS references for Japanese twins were obtained.

The standard reference population dates from 1997[5,6], while the twins in our study were born in the period from 1986 to 1992. In view of the fact that improvements in the availability and quality of food, health and hygiene can lead to an increase in the height-growth rate, various studies have been conducted to identify the difference.[5,6,20] These studies show that secular trend only becomes evident later in life. Since 1965, the height of individuals up to 3 years of age has remained virtually unchanged.[5] With regard to BMI, in the age group from birth to 2.5 years of age, no more than 13% of the population examined in 1997 passed the P90 for 1980, 54% the P50 and 90% the P10.[6] Furthermore, we examined secular trend between 1988/1989, which is part of the period in which the twin data collection took place, and 1997 by using a reference sample obtained from the Social Medical Survey of Children Attending Child Health Clinics cohort, a nationally representative cohort of 2151 children born in the Netherlands in 1988-1989.[21] For this cohort, mean length, weight and BMI SDS was equal to -0.12, -0.05 and 0.12 for age group <0.5 year, 0.01, -0.04 and -0.01 for age group 0.5-1.4 years, and 0.07, 0.04 and 0.05 for age group 1.5-2.5 years. These results show no systematic trend. Therefore, our results are unlikely to be affected by the differences in birth dates.

During the first 2.5 years of life, differences occur in growth between twins and singletons, even after correcting for gestational age. We recommend the use of reference growth charts for twins.

Acknowledgements

The authors gratefully acknowledge the valuable suggestions made by the reviewer.

References

1. Liu YC, Blair EM. Predicted birthweight for singletons and twins. Twin Res 2002;5:529-37.
2. Bleker OP, Oosting J, Hemrika DJ. On the cause of the retardation of fetal growth in multiple gestations. Acta Genet Med Gemellol (Roma) 1988;37:41-6.
3. Hennequin Y, Rorive S, Vermeylen D, Pardou A. Twins: interpretation of height-weight curves at birth. Rev Med Brux 1999;20:81-5.
4. Wilson RS. Twin growth: initial deficit, recovery, and trends in concordance from birth till nine years. Ann Hum Biol 1979;6:205-20.
5. Fredriks AM, Buuren S van, Burgmeijer RJF, Meulmeester JF, Beuker RJ, Brugman E, et al. Continuing positive secular growth change in the Netherlands 1955-1997. Pediatric Res 2000;47:316-23.
6. Fredriks AM, Buuren S van, Wit JM, Verloove-Vanhorick SP. Body index measurements in 1996-7 compared with 1980. Arch Dis Child 2000;82:107-12.
7. Boomsma DI, Orlebeke JF, Baal GCM van. The Dutch Twin Register. Growth data on weight and height. Behav Genet 1999;22:247-51.
8. Boomsma DI, Vink JM, Beijsterveldt TC van, Geus EJ de, Beem AL, Mulder EJ, et al. Netherlands Twin Register. A focus on longitudinal research. Twin Res 2002;5:401-6.
9. Boomsma DI, de Geus EJ, Vink JM, Stubbe JH, Distel MA, Hottenga JJ, et al. Netherlands Twin Register: from twins to twin families. Twin Res Hum Genet 2006;9:849-57.
10. Healy MJ. Reference values and standards for paired organs. Ann Hum Biol 1993;20:75-6.
11. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med 1992;11:1305-19.
12. Green PL. Penalized likelihood for general semi-parametric regression models. International Statistical Review 1987;55:245-59.
13. Buuren S van, Fredriks M. Worm plot: a simple diagnostic device for modelling growth reference curves. Stat Med 2001;20:1259-77.
14. Cole TJ, Pan H. LMS program version 1.16. A program for calculating age-related reference centiles using the LMS method. Londen:Institute of Child Health; 2002.
15. Niklasson A, Ericson A, Fryer JG, Karlberg J, Lawrence C, Karlberg P. An update of the Swedish reference standards for weight, length and head circumference at birth for given gestational age (1977-1981), Acta Paediatr Scand 1991;80:756-62.
16. Dommelen P van, Gunst MC de, Vaart AW van der, Boomsma DI. Genetic study of the height and weight process during infancy. Twin Res 2004;7:607-16.

17. WHO Multicentre Growth Reference Study Group. Assessment of differences in linear growth among populations in the WHO Multicentre Growth Reference Study. Acta Pediatrica Suppl 450:56-65.
18. Hur YM, Luciano M, Martin NG, Boomsma DI, Iacono WG, McGue M et al. A comparison of twin birthweight data from Australia, the Netherlands, the United States, Japan, and South Korea: are genetic and environmental variations in birthweight similar in Caucasians and East Asians? Twin Res Hum Genet 2005;8:638-48.
19. Ooki S, Yokoyama Y. Physical growth charts from birth to six years of age in Japanese twins. J Epidemiol 2003;13:333-41.
20. Hauspie RC, Vercauteren M, Susanne C. Secular changes in growth. Horm Res 1996;45(Suppl 2):8-17.
21. Herngreen WP, Reerink JD, van Noord-Zaadstra B.M., Verloove-Vanhorick SP, Ruys JH. The SMOCC-study: Design of a representative cohort of live-born infants in the Netherlands. Eur J Public Health 1992;2:117-22.

Genetic study of the height and weight process during infancy

3

Paula van Dommelen[1], Mathisca C.M. de Gunst[1], Aad W. van der Vaart[1] and Dorret I. Boomsma[1]

[1]VU University, Amsterdam, The Netherlands

Twin Res 2004;7:607-16.

Summary

Longitudinal height and weight data from 4649 Dutch twin pairs between birth and 2.5 years of age were analyzed. The data were first summarized into parameters of a polynomial of degree 4 by a mixed-effects procedure. Next, the variation and covariation in the parameters of the growth curve (size at one year of age, growth velocity, deceleration of growth, rate of change in deceleration [i.e., jerk] and rate of change in jerk [i.e., snap]) were decomposed into genetic and nongenetic sources. Additionally, the variation in the estimated size at birth and at 2 years of age interpolated from the polynomial was decomposed into genetic and nongenetic components. Variation in growth was best characterized by a genetic model which included additive genetic, common environmental and specific environmental influences, plus effects of gestational age. The effect of gestational age was largest for size at birth, explaining 39% of the variance. The differences between monozygotic and dizygotic twin correlations were largest for size at 1 and 2 years of age and growth velocity of weight, which suggests that these parameters are more influenced by heritability than size at birth, deceleration and jerk. The percentage of variance explained by additive genetic influences for height at 2 years of age was 52% for females and 58% for males. For weight at 2 years of age, heritability was approximately 58% for both sexes. Variation in snap height for males was also mainly influenced by additive genetic factors, while snap for females was influenced by both additive genetic and common environmental factors. The correlations for the additive genetic and common environmental factors for deceleration and snap are large, indicating that these parameters are almost entirely under control of the same additive genetic and common environmental factors. Female jerk and snap, and also female height at birth and height at 2 years of age, are mostly under control of the same additive genetic factor.

Introduction

The growth of an individual can be viewed as the phenotypic expression of his or her genotype and the influence of environmental factors. In this paper we estimate the influence of genetic and environmental factors on variation in height and weight during the first 2.5 years of age in a large sample of Dutch twin pairs born between 1986 and 1992.[1] To this aim a two-stage method was used. Firstly, the longitudinal measurements on height and weights of individuals were reduced to parameters of a polynomial of degree 4. This was done separately for mono- and dizygotic male and female twins. Height and weight at birth and at 2 years of age was estimated by interpolating the polynomial of the individuals. Secondly, a multivariate biometric analysis was performed on the fitted coefficients and the interpolated values by decomposing the variances of the parameter

values into genetic and environmental components. Because both gestational age and sex have been shown to be significant predictors for physical features of infants[2], these variables were included as explanatory variables in the second stage of the study.

Growth curve models can be used to describe growth at particular time-points as well as the process of growth over time. They are well suited to analyze longitudinal data when the times of measurement are irregularly spaced and differ for different individuals, as they describe growth with a limited number of interpretable parameters, such as height and weight at 1 year of age, growth velocity, deceleration of growth, rate of change in deceleration (i.e., jerk) and rate of change in jerk (i.e., snap). The uniform description through such growth parameters makes it possible to compare individuals. A number of growth curves have been suggested in the literature and have been shown to be representative at different periods of life.[3-5]

The present study was conducted to expand on previous research on the genetics of height and weight in young Dutch children, in particular the study of Baker et al.[6], In Baker et al.'s study, longitudinal data on height of a subset (996 twin pairs) of our sample (2701 pairs for height and 3477 pairs for weight) were summarized by parameters of the quadratic polynomial growth curve via a multiple regression procedure for each individual. These parameters were then subjected to a multivariate biometrical analysis.

A polynomial of degree 4 has been used as the growth curve, which turned out to be better suited than polynomials of degree 2 or 3 to describe growth of children from birth to 2.5 years of age. To estimate the growth curve parameters a mixed-effects model was used. Mixed-effects models are generally used to describe a relationship between a response variable and covariates in data from individuals that are grouped according to one or more classification factors. The grouping is reflected by the fact that each parameter in the model (size at 1 year of age, growth velocity, deceleration of growth, jerk and snap) is the sum of a fixed and a random component. The fixed components are the same for individuals in the same group, but may vary for the different groups. The random components are different for every individual, but for individuals of the same group originate from the same normal distribution. The individuals in one group are thus treated as a random sample from a population of similar individuals.[7] Mixed-effects models are designed to estimate the average behavior of an individual in the population, as well as the variability among and within individuals.[8] An advantage of using mixed-effects models instead of performing multiple regression is a reduction of the number of free parameters. Because the number of observations (i.e., the number of time instances at which measurements were taken) per individual can be small, the parameter estimates obtained by applying a multiple regression procedure to each individual separately are subject to a relatively large estimation error. By using a mixed-effects model we use data from all individuals to estimate the growth coefficients for each individual by taking into

account all other individuals. This will result in parameters that fluctuate less between individuals than if multiple regressions had been used. Of course, the procedure will be unreliable if the growth curves cannot be described well by a mixed-effects model.

In the second stage of the analyses, the variation in the growth curve parameters was decomposed into genetic and nongenetic components. Two series of multivariate analyses were carried out: one which simultaneously analyzed size at 1 year of age, growth velocity, deceleration of growth, rate of change in deceleration and rate of change in jerk. The second series of genetic models analyzed size at birth and at 2 years of age. The genetic model fitting was done separately for the height and the weight data.

Materials and Methods

Longitudinal growth data were obtained from the Netherlands Twin Register (NTR) at the VU University in Amsterdam, The Netherlands. Parents of twins responded to questionnaire items on twin similarity, gestational age, and height and weight as measured by the Youth Health Services up until the age of approximately 2.5 years. Parents were asked to indicate when height and weight had been measured.[9] Similarity items were used to obtain zygosity of same-sex twin pairs. The agreement between zygosity assigned by the replies to the questions and zygosity determined by DNA markers/blood typing was around 93%.[10]

We started with a total of 4649 twin pairs who were born between 1986 and 1992. After checking the data for suitability of inclusion in the analysis, there were 4137 twin pairs for the height analysis and 4154 pairs for the weight analysis. These twin pairs had a known zygosity and both the youngest and oldest of the twin pair had at least one measurement for height or weight. The median number of measurements was 9 (SD = 2.5) for height and 12 (SD = 2.8) for weight per child. The maximum number of measurements was 20. The twin pairs were divided into six zygosity groups, MZM (monozygotic, males), DZM (dizygotic, males), MZF (monozygotic, females), DZF (dizygotic, females), DOSmf (dizygotic opposite sex, male born first), and DOSfm (dizygotic opposite sex, female born first).

For each child, each individual growth pattern was summarized into the parameters of a polynomial of degree 4 (for descriptions see below). The measurements on height and weight were analyzed separately. To ensure good parameter estimates each child was required to have at least one measurement before the age of 3 months, at least one between 3 months and 1 year and 3 months, and at least one after the age of 1 year and 3 months. With these requirements we have a total of 472 MZM, 434 DZM, 528 MZF, 412 DZF, 447 DOSmf and 408 DOSfm for the height analysis. For the weight analysis there were 587 MZM, 546 DZM, 663 MZF, 543 DZF, 595 DOSmf and 543 DOSfm. The

sample of twin pairs with known gestational age is a reduced dataset which consists of 444 MZM, 415 DZM, 505 MZF, 395 DZF, 434 DOSmf and 394 DOSfm twin pairs for height and 550 MZM, 526 DZM, 636 MZF, 518 DZF, 573 DOSmf and 522 DOSfm twin pairs for weight.

Estimates of the polynomial growth parameters for each individual were obtained by first fitting mixed-effects models with maximum likelihood and next computing the estimated conditional modes of the random effects given the observations. In this step the data from DZM and DOS males, as well as from DZF and DOS females were combined, because no large differences in heights and weights between these groups were noticed.[11] The estimation procedure was therefore based on four groups of individuals, namely MZM, MZF, DZM and DOS males, and DZF and DOS females.

Splus 6.1 was used for the computations. Growth at birth and 2 years of age was estimated by interpolating the polynomial of the individuals. The individual sets of growth parameters and interpolated values were then subjected to further analyses.

The height and weight data were analyzed by mixed-effects models. A mixed-effects model assumes each growth parameter to be the sum of a fixed and a random component, where the fixed component is the same for every individual, and the random component is different but has the same normal distribution. Therefore, this model accommodates individual variations through the random effects, but ties the individuals together through the fixed effects and the covariance matrix of the random effects. The fixed effects represent the mean values of the parameters in the subpopulation of individuals. The random effects represent the deviations of the individual coefficients from their subpopulation average. Therefore, random effects contribute to the covariance structure of the data. These effects may introduce correlations between cases. In our situation each of the four groups of MZM, MZF, DZM/DOS males, and DZF/DOS females is viewed as a subpopulation with its own parameter values.

Let n be the number of children, t the age in years and $y_i(t)$ the height (in cm) or weight (in kg) of the i th child at age t. Then, for $i = 1,\ldots, n$, the dependency of the response variables height and weight on age is given by the following polynomial of degree 4 (centered at age 1):

$$y_i(t) = \alpha_1 + \alpha_2(t-1) + \alpha_3(t-1)^2 + \alpha_4(t-1)^3 + \alpha_5(t-1)^4 + \varepsilon_{it}$$

In this model, α_1 represents the height/weight at 1 year of age, α_2 the instantaneous rate of growth at 1 year (velocity), α_3 the amount of deceleration in the individual's growth curve, α_4 represents the rate of change in deceleration (jerk) and α_5 the rate of change in jerk (snap).

A mixed-effects model assumes each growth parameter to be the sum of a fixed and a random component, which is $\alpha_k = \alpha_{k0} + \alpha_{ki}$, with α_{k0} fixed effects and α_{ki} random effects, $k=1,...,5$. The measurement errors ε_{it} are assumed to be independent across individuals and to be normally distributed with mean zero and a common variance. For the mixed-effects procedure it is assumed that for different individuals the random effects have the same multivariate normal distribution with mean vector zero, and are independent of the measurement errors.

Growth parameters of children can be estimated with only three observations by a fourth-order polynomial mixed-effects model, as this model ties the individuals together through the fixed effects and the covariance matrix of the random effects. Therefore, the model borrows strength across individuals in estimating individual parameters. Thus with three observations, estimation with a fourth-order polynomial, the mixed-effects model is less of a problem than with the simpler method that estimates the parameters for each individual separately. It is also possible to estimate a child's growth parameters by a fourth-order polynomial mixed-effects model with only one or two observations, but in this study, each child was required to have at least three measurements, or the growth curve would have been smoothed too much towards the average curve.

Size at birth, β_1, and size at 2 years of age, β_2, was obtained by interpolating the polynomial of degree 4 with its estimated parameters.

Genetic Model Fitting

Pearson correlation coefficients were used to summarize twin resemblance for each of the growth parameters (α_1 through α_5 and β_1 and β_2). Pearson correlations were also used to quantify the relationship between gestational age and the growth parameters and size at birth and at 2 years. Variation between individuals in the growth parameters was analyzed as a function of additive genetic influences, common and specific environment, and gestational age. For each pair of twins, the gestational age, the set of growth parameters and interpolated values were collected in a vector, and multivariate modeling was carried out on the variance-covariance matrices of these vectors with the computer package Mx 1.52.[12] The unknown parameters of the multivariate model, which are denoted by the vector θ, were estimated by maximum likelihood under the assumption that the observational vectors, that is, for each pair of twins the vector containing gestational age and the growth parameters or the interpolated values, are sampled independently from a multivariate normal distribution, with the form of the covariance matrix depending on the zygosity group.

This corresponds to minimizing with respect to θ for the six twin groups simultaneously a distance function between the covariance matrix $\sum_i(\theta)$ of the form particular to the group ($i=1, 2,...,6$) and the sample covariance matrix S_i of the observations in the group.

In particular, the function

$$F_{ML} = \sum_{i=1}^{6} N_i \left(\log \det \Sigma_i(\theta) - \log \det S_i + \text{trace}(S_i \, \Sigma_i^{-1}(\theta)) - p \right)$$

was minimized where N_i denotes the number of twin pairs in zygosity group *i*, and *p* equals the dimension of the observational vectors (in the first series of analyses $p = 11$, i.e., 5 growth parameters for the oldest twin, 5 for the youngest twin, and gestational age; in the second series of analyses $p = 5$, i.e., 2 interpolated values for the oldest twin, 2 for the youngest twin and gestational age). The likelihood ratio statistic was used to determine the goodness-of-fit of the different models relative to the model with unrestricted covariance matrices (i.e., when the covariance matrices of the six twin groups are estimated by the sample covariance matrices S_i). As it may be difficult to obtain a good-fitting model with large datasets when the number of observations is large[13], the Normed Fit Index (NFI), Akaike's information criterion (AIC)[14] and Bayesian Information Criterion (BIC) were also calculated. Values of NFI close to 1 and negative values of AIC or BIC indicate that the model under consideration provides a good fit to the data relative to the corresponding model with unrestricted covariance matrices.

Variation in growth parameters was assumed to be the sum of additive genetic variance, and common and unique environmental variances. Two series of analyses were carried out for both height and weight. In the first series, size at 1 year of age, growth velocity, deceleration of growth, rate of change in deceleration and rate of change in jerk were simultaneously analyzed with a triangular decomposition. In the second set of analyses, size at birth and at 2 years of age were simultaneously analyzed. The estimated growth parameters (α_1 through α_5 and β_1 and β_2) were modeled as linear functions of the latent variables additive genetic effects (*A* male; *A*' female), common environment (*C* male, *C*' female), specific environment (*E* male, *E*' female) and the observed variable gestational age (*GA* male, *GA*' female).[12,15,16]

Figure 1 depicts the path diagram for DOSfm twin pairs for size at birth and 2 years of age. The vector (β_1, β_2) on the left contains the data for the female twin; the vector (β_1', β_2') on the right contains the corresponding values for the male twin. The vector (*ga,* β_1, β_2, β_1', β_2') is expressed linearly in the latent factors which are indicated in circles. The latent factors are represented by two additive genetic factors for the first twin and two additive genetic factors for the second twin. In twin pairs of opposite sex, these factors correspond with *A* and *A*'. Likewise, there are four common environmental factors and four specific environmental factors. The model also includes a gestational age factor. The female β_1 loads on A_1, C_1, E_1, *GA*, and the female β_2 on A_1, A_2, C_1, C_2, E_1, E_2, *GA*.

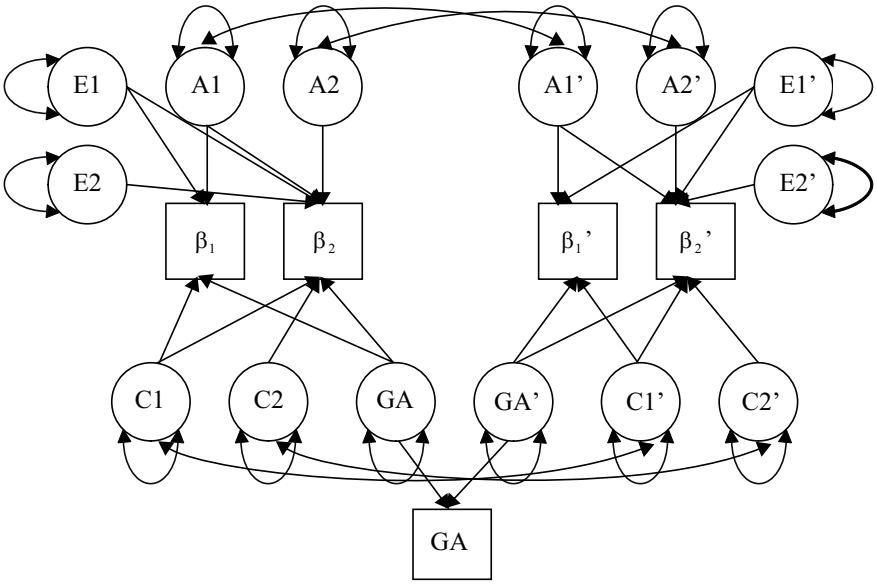


Figure 1 *Path diagram of the interpolated values; size at birth (β_1 and β_1') and size at two years of age (β_2 and β_2').*

The structure of the covariance matrices for the six twin groups follows from the model given in Figure 1 and can be conveniently described by writing the latent vectors as matrix products involving standard normal random vectors.

Let the *p*-dimensional observational vector for a given twin pair be denoted as (γ, ξ, ξ*) representing gestational age and the growth parameters or interpolated values of the youngest and oldest of the twin pair. Then it is assumed that

$$\begin{pmatrix} \gamma \\ \xi \\ \xi^* \end{pmatrix} = \begin{pmatrix} 0 \\ XA \\ X^*A^* \end{pmatrix} + \begin{pmatrix} 0 \\ YC \\ Y^*C^* \end{pmatrix} + \begin{pmatrix} V \\ S \\ S^* \end{pmatrix} GA + \begin{pmatrix} 0 \\ ZE \\ Z^*E^* \end{pmatrix}$$

where *X, X*, Y, Y*, Z* and *Z** are deterministic lower-triangular (*k* x *k*) matrices (where *p= 2k+1* and *k* equals 5 for the first series of analyses and 2 for the second series), *V* is a number and *S* and *S** are deterministic *k*-vectors; and *A, A*, C, C*, E, E*,* are *k*-dimensional standard normal random vectors and *GA* is a standard normal random variable. The elements of the matrices *X, X*, Y, Y*, Z, Z*, V, S* and *S** are called factor loadings, and are the unknown parameters θ that are estimated from the data. The model includes the possibility that the factor loadings depend on the sex of the individual. Factor

loadings are the same for all individuals of the same sex. For instance, for a monozygotic pair of twins we have $X=X^*$, where this matrix may be different for male and female monozygotic twin pairs. Furthermore, the four vectors in the decomposition on the right are assumed to be stochastically independent, all vectors E and E^* are assumed to be stochastically independent, reflecting different specific environments, and $C=C^*$, reflecting identical common environment for the two individuals in a pair of twins. Finally, it is assumed that $A=A^*$ for a monozygotic pair of twins, reflecting identical genetic make-up. The cross-covariance matrix between A and A^* is assumed to be 0.5 times the identity matrix for a dizygotic pair of twins.

The factor loadings or path coefficients which represent the influence of the latent factors on the observations are estimated, together with the unknown variance of gestational age, by maximum likelihood based on the joint distribution of the growth parameters and gestational age as indicated above. Next it is possible to compute for each of the factors the proportion of the variance that it contributes to the total variance of the observational vector. Refer to Neale and Cardon[16] for more details on the triangular or Cholesky decomposition.

Several submodels of the general model can be formed by setting appropriate sets of factor loadings equal to zero. In the 'Null Model' *EE*(GA)(GA*)* the loadings on both the additive genetic factors and the common environmental factors are assumed to be zero: $X=X^*=Y=Y^*=0$. In this model, any familial resemblance in growth can only arise because there is variation between twin pairs in *GA*. The additive genetic factors and common environmental factors are added separately in the models *AA'EE'(GA)(GA')*, and *CC'EE'(GA)(GA')* respectively, whereas *AA'CC'EE'(GA)(GA')* is the full model with all factors included. Finally, *ACE(GA)* denotes the model with all types of factors included, but with the factor loadings constrained to be identical for males and females. These submodels, versus the model in which the covariance matrix of the observational vector is an arbitrary positive definite matrix, can be tested through the likelihood ratio test.

Results

The estimates of the growth parameters of the polynomial of degree 4 and the residual variances of the mixed-effects models for the zygosity groups are given in Table 1.

Table 1 *Estimates of the fixed component along with the standard deviation of the random component, the residual variance and Akaike's criterion (AIC) of the polynomial of degree 4 mixed-effects model for different zygosity groups.*

Height parameters (N=2701 twin pairs)												
Zygosity	α_1		α_2		α_3		α_4		α_5		Residual variance	AIC
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD		
MZM	75.3	2.57	14.4	1.94	-4.99	2.58	6.52	1.85	-3.20	1.69	0.869	33058
MZF	73.8	2.55	14.9	1.86	-4.50	2.78	5.60	1.69	-2.95	2.06	0.799	35122
DZM & DOSm	75.3	2.57	14.2	1.99	-4.17	2.91	6.41	2.08	-3.65	2.44	0.895	61092
DZF & DOSf	74.1	2.46	14.8	1.85	-4.38	2.56	5.52	1.89	-2.90	2.13	0.860	57814
Weight parameters (N=3477 twin pairs)												
Zygosity	α_1		α_2		α_3		α_4		α_5		Residual variance	AIC
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD		
MZM	9.61	0.96	3.64	0.99	-1.17	1.48	1.57	0.85	-0.98	1.12	0.211	11407
MZF	9.11	1.01	3.60	0.92	-1.11	1.32	1.38	0.73	-0.83	1.02	0.186	9310
DZM & DOSm	9.65	0.97	3.73	1.05	-1.12	1.52	1.43	0.91	-0.95	1.20	0.205	21024
DZF & DOSf	9.15	0.94	3.62	0.92	-1.13	1.33	1.31	0.74	-0.75	0.98	0.193	16735

Note: α_1 = size at one year of age, α_2 = velocity, α_3 = deceleration, α_4 = jerk, α_5 = snap.

The height and weight curves based on the estimated fixed parameters of the mixed model for height and weight for female and male twin pairs are shown in Figure 2. The Pearson correlations among the growth parameters, the interpolated values and gestational age are shown in Table 2. Size at 1 year of age (α_1) correlates largely with size at 2 years of age (β_2). Also the correlation between deceleration (α_3) with snap (α_5) is large. The longer the gestation period, the larger the height and weight at birth (β_1). Also, the larger the height at birth or the longer the gestation period, the more slowing over the growth rate (α_2) can be seen. Large growth velocity implies a large and heavy child at the age of 2 and deceleration rate is changing rapidly.

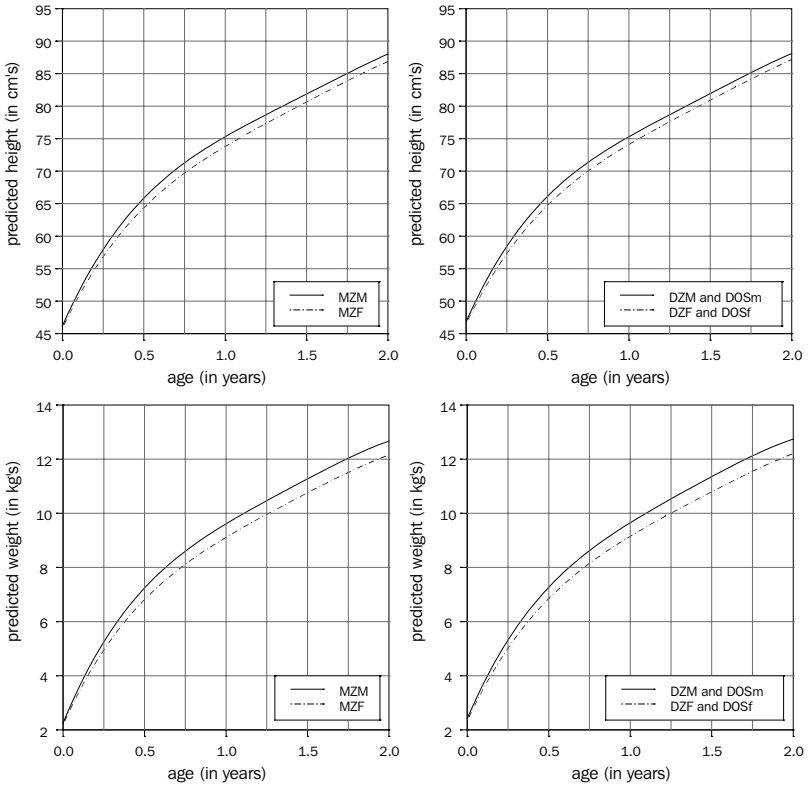


Figure 2 *The height and weight growth curves for monozygotic and dizygotic boys and girls based on the estimated fixed parameters of the polynomial of degree 4 mixed-effects model.*

Table 2 *Correlations among polynomial parameters, interpolated values and gestational age (upper=males, lower=females).*

	Height									Weight							
	α_1	α_2	α_3	α_4	α_5	β_1	β_2	GA		α_1	α_2	α_3	α_4	α_5	β_1	β_2	GA
α_1	1	0.26	-0.13	0.17	-0.01	0.50	0.91	0.21	α_1	1	0.39	-0.32	0.11	0.11	0.28	0.83	0.17
α_2	0.29	1	-0.31	-0.46	0.41	-0.13	0.50	-0.28	α_2	0.46	1	-0.08	-0.75	0.21	0.13	0.64	-0.06
α_3	-0.05	-0.26	1	-0.02	-0.71	0.42	0.05	0.41	α_3	-0.38	-0.13	1	0.09	-0.92	0.23	0.09	0.21
α_4	0.12	-0.47	-0.04	1	-0.65	-0.41	0.05	-0.16	α_4	0.10	-0.70	0.09	1	-0.36	-0.31	-0.05	-0.04
α_5	-0.17	0.32	-0.72	-0.60	1	0.07	-0.10	-0.13	α_5	0.15	0.22	-0.91	-0.37	1	-0.11	-0.17	-0.17
β_1	0.53	-0.11	0.46	-0.42	-0.06	1	0.43	0.62	β_1	0.30	0.08	0.23	-0.25	-0.12	1	0.27	0.62
β_2	0.92	0.52	0.13	0.00	-0.25	0.46	1	0.15	β_1	0.85	0.69	0.01	-0.05	-0.12	0.29	1	0.14
GA	0.26	-0.23	0.40	-0.16	-0.17	0.62	0.20	1	GA	0.18	-0.08	0.20	-0.02	-0.16	0.64	0.15	1

Note: α_1 - α_5 see Table 1, β_1 = size at birth, β_2 = size at two years of age.

To show the twin resemblance for the growth parameters and the interpolated values, the within-pair correlations corrected for gestational age are shown in Table 3. It can be seen that there is marked twin resemblance for the parameters, with the MZ correlations being significantly larger than the DZ correlations. This indicates that at least some degree of heritability exists. However, MZ correlations are not twice as high as DZ correlations, which points to an additional influence of the common environment. The differences between monozygotic and dizygotic twins are largest for size at 1 (α_1) and 2 years of age (β_2) and growth velocity of weight (α_2). This means that these parameters are more influenced by heritability than size at birth, deceleration, jerk and snap.

Table 3 *Within-pair correlations for parameters and interpolated values corrected for gestational age.*

	Height							Weight						
Zygosity	α_1	α_2	α_3	α_4	α_5	β_1	β_2	α_1	α_2	α_3	α_4	α_5	β_1	β_2
MZM	0.89	0.80	0.67	0.74	0.67	0.65	0.90	0.87	0.88	0.84	0.88	0.84	0.70	0.86
DZM	0.60	0.60	0.61	0.49	0.55	0.55	0.61	0.55	0.58	0.68	0.70	0.69	0.55	0.53
MZF	0.89	0.79	0.72	0.69	0.69	0.65	0.90	0.84	0.89	0.83	0.86	0.84	0.70	0.87
DZF	0.66	0.57	0.52	0.54	0.54	0.53	0.66	0.57	0.62	0.66	0.66	0.66	0.59	0.55
DOSmf	0.55	0.57	0.55	0.65	0.59	0.63	0.56	0.55	0.60	0.65	0.64	0.65	0.55	0.54
DOSfm	0.50	0.55	0.52	0.62	0.54	0.56	0.47	0.54	0.57	0.56	0.63	0.59	0.47	0.55

Note: α_1 - α_5 see Table 1, β_1 - β_2 see Table 2.

Goodness-of-fit tests were performed step-wise from the Null model (*EE'(GA)(GA')*) which contains specific environmental factors and gestational age for males and females, to the final model *(AA'CC'EE'(GA)(GA'))* which contains all factors of interest in this study. All growth models had the final model as best goodness-of-fit. The results from the Null model to the final model are summarized in Table 4. This table shows that all BIC of the final models are negative, which means that the Cholesky decomposition provides a good fit relative to the model with unrestricted covariance matrix.

Table 4 *Goodness of fit tests of different Cholesky decompositions varying from the Null model in which the resemblance of the halves of a twin pair is due to gestational age effect, model without a common environmental component, model without an additive genetic component, model with sex-limitation to the model with additive genetic and common environmental components without sex-limitation.*

Measure	Model	χ^2	df	p	AIC	BIC	NFI
Height	*EE'(GA)(GA')*	9715	355	<0.001	9005	6925	0.00
α_1, α_2, α_3, α_4, α_5	*AA'EE'(GA)(GA'*	2072	325	<0.001	1422	-482	0.81
	CC'EE'(GA)(GA')	3036	325	<0.001	2386	482	0.71
	ACE(GA)	2691	340	<0.001	2011	19	0.75
	AA'CC'EE'(GA)(GA')	1514	295	<0.001	924	-804	0.87
Weight	*EE'(GA)(GA')*	12448	355	<0.001	11738	9569	0.00
α_1, α_2, α_3, α_4, α_5	*AA'EE'(GA)(GA')*	1410	325	<0.001	760	-1225	0.91
	CC'EE'(GA)(GA')	2251	325	<0.001	1601	-384	0.84
	ACE(GA)	1215	340	<0.001	535	-1542	0.93
	AA'CC'EE'(GA)(GA')	611	295	<0.001	21	-1781	0.97
Height	*EE'(GA)(GA')*	3535	79	<0.001	3377	2914	0.00
β_1, β_2	*AA'EE'(GA)(GA')*	330	73	<0.001	184	-244	0.93
	CC'EE'(GA)(GA')	707	73	<0.001	561	133	0.82
	ACE(GA)	184	76	<0.001	32	-413	0.97
	AA'CC'EE'(GA)(GA')	123	67	<0.001	-11	-404	0.98
Weight	*EE'(GA)(GA')*	4302	79	<0.001	4144	3661	0.00
β_1, β_2	*AA'EE'(GA)(GA'*	333	73	<0.001	187	-259	0.94
	CC'EE'(GA)(GA')	745	73	<0.001	599	153	0.84
	ACE(GA)	199	76	<0.001	47	-417	0.97
	AA'CC'EE'(GA)(GA')	135	67	<0.001	1	-408	0.98

Note: α_1 - α_5 see Table 1, β_1 - β_2 see Table 2.

The maximum likelihood estimates of the additive genetic, common environmental, gestational age and the specific environmental variance proportions under the best-fitting mixed model for females and males are given in Tables 5a and 5b, respectively. From these tables it can be seen that variation in height and weight at birth is to a large extent determined by gestational age (38% to 40% explained variance). For weight at birth, 21% to 27% of the variance is explained by common environmental factors. Size at 1 year (except for female height) and 2 years of age is mostly influenced by additive genetic factors (55% to 74%, and 52% to 59%, respectively). The largest differences between the sexes are in the height at 1 year of age. Males show a much larger proportion

of variance due to additive genetic influences. The weight velocity parameter (57% to 63%) is also mainly determined by additive genetic factors. Height growth velocity, deceleration (except for female height) and rate of change in deceleration are explained by both additive genetic and common environmental factors. Deceleration of female height is mainly determined by common environmental factors (44%), specific environmental factors (28%) and partly by gestational age (16%).

The correlations for the additive genetic and common environmental factors for deceleration and snap are large, indicating that these parameters are almost entirely under control of the same additive genetic and common environmental factors. Female jerk and snap, and also female height at birth and height at 2 years of age are mostly under control of the same additive genetic factor.

Table 5a *Proportions of variance (diagonal) and correlations (off-diagonal) of the additive genetic component (A), common environmental component (C), gestational age (GA) and specific environmental component (E) of the ACE without sex-limitation Cholesky decomposition for the female parameters of the polynomial and the interpolated values.*

Females		A	C	GA	E
Height	α_1	0.44	0.39	0.07	0.10
	α_2	0.51 0.36	0.36 0.38	0.06	0.01 0.20
	α_3	0.42 0.63 0.12	0.10 0.02 0.44	0.16	0.07 0.13 0.28
	α_4	0.17 -0.59 -0.62 .31	0.24 -0.45 0.20 0.42	0.02	-0.03 -0.62 -0.02 0.25
	α_5	0.07 -0.54 -0.77 0.95 0.33	0.29 -0.07 -0.69 0.48 0.36	0.04	-0.04 -0.31 -0.81 0.56 0.28
Weight	α_1	0.64	0.17	0.03	0.16
	α_2	0.64 0.57	0.10 0.31	0.01	0.35 0.11
	α_3	0.51 0.15 0.45	0.37 0.12 0.35	0.04	0.30 -0.06 0.16
	α_4	0.07 -0.65 0.04 0.39	0.20 -0.83 -0.23 0.48	0.00	0.16 -0.68 -0.06 0.13
	α_5	-0.26 -0.18 -0.91 0.23 0.37	-0.15 -0.34 -0.91 0.53 0.43	0.03	-0.09 -0.02-0.92 0.30 0.17
Height	β_1	0.10	0.30	0.38	0.22
	β_2	0.92 0.52	0.17 0.34	0.04	0.43 0.10
Weight	β_1	0.14	0.27	0.40	0.20
	β_2	0.34 0.58	0.06 0.25	0.02	0.47 0.15

Note: α_1 - α_5 see Table 1, β_1 - β_2 see Table 2.

Table 5b *Proportions of variance (diagonal) and correlations (off-diagonal) of the additive genetic component (A), common environmental component (C), gestational age (GA) and specific environmental component (E) of the ACE without sex-limitation Cholesky decomposition for the male parameters of the polynomial and the interpolated values.*

Males		A	C	GA	E
Height	α_1	0.74	0.13	0.04	0.09
	α_2	0.54 0.44	0.05 0.32	0.07	0.08 0.17
	α_3	0.38 0.33 0.33	0.16 0.06 0.32	0.15	0.05 0.25 0.20
	α_4	0.30 -0.49 -0.19 0.48	0.18 -0.55 0.08 0.30	0.03	-0.07 -0.60 -0.03 0.19
	α_5	0.00 -0.41 -0.78 0.74 0.60	0.00 -0.29 -0.71 0.60 0.25	0.01	-0.09 -0.49 -0.73 0.68 0.14
Weight	α_1	0.55	0.31	0.03	0.12
	α_2	0.71 0.63	-0.15 0.26	0.01	0.28 0.11
	α_3	0.56 0.26 0.34	0.27 -0.08 0.48	0.05	0.08 -0.17 0.14
	α_4	-0.24 -0.77 -0.17 0.43	0.56 -0.79 -0.05 0.47	0.00	0.11 -0.74 -0.09 0.11
	α_5	-0.35 -0.31 -0.90 0.39 0.32	0.00 -0.16 -0.93 0.37 0.50	0.03	0.05 0.04 -0.94 0.30 0.15
Height	β_1	0.15	0.27	0.39	0.20
	β_2	0.49 0.58	0.43 0.29	0.02	0.48 0.10
Weight	β_1	0.24	0.21	0.38	0.17
	β_2	0.16 0.59	0.26 0.25	0.02	0.47 0.14

Note: α_1 - α_5 see Table 1, β_1 - β_2 see Table 2.

Discussion

The height and weight process during infancy has been described by summarizing longitudinal data into parameters of a growth model. Several growth models were tried, namely the Jenss-Bayley growth curve[17], the first component of the Infancy-Childhood-Puberty model, polynomials with a maximum degree of 4, and the polynomial of degree 4 was chosen. It had the best fit (i.e., the smallest AICs and residual variances) without being overfitted and has parameters which are straightforwardly interpreted in terms of growth. The second best model is the Jenss-Bayley model[17], of which the AICs are between 13 to 451 larger for height, and between 2088 to 4504 larger for weight compared to the polynomial of degree 4.

When estimating the growth parameters by a multiple regression procedure instead of a mixed-effects model, several problems were encountered: the normality assumption for

the parameters, which is needed for the biometric analyses, is violated, the fluctuation of parameters is strong because of the variability in number of observations per individual, and some individuals had to be removed from the data as the multiple regression fitting procedure cannot handle a small number of observations per individual. Therefore, it was chosen to estimate the growth parameters by a mixed-effects model.
The estimated growth parameters, together with the interpolated values for size at birth, were modeled by several multivariate genetic models. The fit of the multivariate genetic models is reasonably good, as all BIC are less than zero. The results of these models are the proportions of variance and correlation explained by additive genetic, common environmental, gestational age and specific environmental factors. The correlations indicate which parameters are under control of the same additive genetic and common environmental factors.

Statistically significant additive genetic variance was found for variation in height at birth, at 1 and at 2 years of age. In the first year, the additive genetic component for height increased from 0.10-0.15 to 0.44-0.74 and for weight from 0.14-0.24 to 0.55-0.64. Similar results were obtained by Levine et al. for American twins.[18] Baker et al.[6] used a subset of our dataset (n = 996 twin pairs out of 2701) and found that the additive genetic component for height varies between 0.25 and 0.45 at 1 year of age.

In common with Baker et al.[6], it was found that the models with sex-limitation fit the data better than models constraining equality across sexes. It was also concluded that deceleration of height in females is largely determined by common environmental factors. Common environmental factors explain 21% to 27% of the variance for weight at birth. As the common effect of the mother and a more general common environmental effect cannot be separated in this design, the variance is likely to be due to maternal effects. Vandenberg et al.[19] concluded that genetic factors appear to be of paramount importance for the deceleration of the growth rate based on a polynomial of degree 2 centered at birth. For velocity, it was found that both common environmental factors and additive genetic factors are important, while Baker et al.[6] concluded that additive genetic factors are more important. These differences may be due to the fact that a polynomial of degree 4 instead of degree 2 was used, that the growth parameters were estimated by a mixed-effects model instead of by a multiple regression procedure, and that the data set was larger.

When observing the correlations obtained from the multivariate genetic models, it was concluded that deceleration and snap are almost entirely under control of the same additive genetic and common environmental factors. Female jerk and snap, and also female height at birth and height at 2 years of age, are mostly under control of the same additive genetic factor.

Literature surveys show that there are several factors that could explain part of additive genetic or common environmental effects on the growth parameters and the interpolated values.[20-22] Examples of genetic and environmental factors include mother's educational level, family income, smoking, alcohol, caffeine and parity. Further investigation into this would be of value in the future.

Acknowledgements

We thank C.E.M. van Beijsterveldt for her help with the computer package Mx and S. van Buuren for his useful comments on the growth models and fitting procedures.

References

1. Boomsma DI, Orlebeke JF, van Baal GC. The Dutch Twin Register: growth data on weight and height. Behav Genet 1992;22:247-51.
2. Ben-Amitai D, Livshits G, Levi I, Kobyliansky E, Merlob P. The relative contribution of birth weight and gestational age to physical traits of newborn infants. Early Hum Dev 1990;22:131-44.
3. Karlberg J, Engström I, Karlberg P, Fryer JG. Analysis of linear growth using a mathematical model. I. From birth to three years. Acta Paediatr Scand 1987;76:478-88.
4. Karlberg J, Fryer JG, Engström I, Karlberg P. Analysis of linear growth using a mathematical model. II. From 3 to 21 years of age. Acta Paediatr Scand Suppl 1987;337:12-29.
5. Thissen D, Bock RD. Linear and nonlinear curve fitting. In A. von Eye (Ed.), Statistical methods in longitudinal research, Volume II: Time series and categorical longitudinal data. Orlando, FL: Academic Press, 1990; pp 289-318.
6. Baker LA, Reynolds C, Phelps E. Biometrical analysis of individual growth curves. Behav Genet 1992;22:253-64.
7. Lindstrom ML, Bates DM. Nonlinear mixed effects models for repeated measures data. Biometrics 1990;46:673-87.
8. Pinheiro JC, Bates DM. Mixed-effects models in S and S-PLUS, statistics and computing. New York: Springer-Verlag, 2000.
9. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. Nat Rev Genet 2002;3:872-82.
10. Rietveld MJ, van Der Valk JC, Bongers IL, Stroet TM, Slagboom PE, Boomsma DI. Zygosity diagnosis in young twins by parental report. Twin Res. 2000;3:134-41.
11. van Dommelen P, De Gunst MC, Van Der Vaart AW, Van Buuren S, Boomsma DI. Growth charts for height, weight and body-mass index of twins during infancy. Ned Tijdschr Geneeskd 2004;148:1345-50.

12. Neale MC, Boker SM, Xie G, Maes HH. Mx: Statistical modeling (5th ed.). Richmond, VA: Department of Psychiatry, Medical College of Virginia, 1999.
13. Hair JF, Anderson RE, Tatham PL, Black WC. Multivariate data analysis. Upper Saddle River, NJ: Prentice-Hall, 1998.
14. Akaike H. Factor analysis and AIC. Psychometrika 1987;52:317-332.
15. Boomsma DI, Vink JM, van Beijsterveldt TC, de Geus EJ, Beem AL, Mulder EJ, Derks EM, Riese H, Willemsen GA, Bartels M, van den Berg M, Kupper NH, Polderman TJ, Posthuma D, Rietveld MJ, Stubbe JH, Knol LI, Stroet T, van Baal GC. Netherlands Twin Register: A focus on longitudinal research. Twin Res 2002;5:401-6.
16. Neale MC, Cardon LR. Methodology for Genetic Studies of Twins and Families. Dordrecht, the Netherlands: Kluwer Academic, 1992.
17. Jenss RM, Bayley N. A mathematical method for studying the growth of a child. Hum Biol 1937;9:556-63.
18. Levine RS, Hennekens CH, Jesse MJ. Genetic variance of weight and length in infant twins. Am J Epidemiol 1987;126:929-35.
19. Vandenberg SG, Falkner F. Hereditary factors in human growth. Hum Biol 1965; 37:357-65.
20. Brooke OG, Anderson HR, Bland JM, Peacock JL, Stewart CM. Effects on birth weight of smoking, alcohol, caffeine, socioeconomic factors, and psychosocial stress. BMJ 1989;25;298:795-801.
21. Guimarães LV, Latorre MD, Barros MB. Risk factors in the occurrence of short stature of preschool children. Cad Saude Publica 1999;15:605-15.
22. Herngreen WP, van Buuren S, van Wieringen JC, Reerink JD, Verloove-Vanhorick SP, Ruys JH. Growth in length and weight from birth to 2 years of a representative sample of Netherlands children (born in 1988-89) related to socioeconomic status and other background characteristics. Ann Hum Biol 1994;21:449-63.

Towards evidence-based referral criteria for growth monitoring

4

Stef van Buuren[1], Paula van Dommelen[1], Gladys R.J. Zandwijken[2], Floor K. Grote[3], Jan Maarten Wit[3] and Paul H. Verkerk[1]

[1]TNO Quality of Life, Leiden, The Netherlands
[2]Dutch Growth Research Foundation, Rotterdam, The Netherlands
[3]Leiden University Medical Center, Leiden, The Netherlands

Arch Dis Child 2004;89:336-41.

Abstract

Aims: To evaluate the performance of growth monitoring in detecting diseases. Turner's syndrome (TS) is taken as the target disease.
Methods: Case-control simulation study. Three archetypal screening rules are applied to longitudinal growth data comparing a group with TS versus a reference group from birth to the age of 10 years. Main outcome measures were sensitivity, specificity, and median referral age.
Results: Clear differences in performance of the rules were found. The best rule takes parental height into account. Combining rules could improve diagnostic accuracy.
Conclusion: Growth monitoring is useful to screen for TS. A combined rule that takes absolute height SDS, parental height, and deflection in height velocity into account is the best way to do this. Similar research is needed for other diseases, populations, and ages, and the results should be synthesized into evidence-based referral criteria.

Introduction

Monitoring child growth and development is a routine part of child health care in many countries. In a typical scenario, the health care worker plots heights and weights on a reference diagram, and assesses whether the growth pattern of the child deviates from that of the reference population. If so, closer examination of the child might be needed. An important goal is to identify diseases and conditions that manifest themselves through abnormal growth. Examples include Turner's syndrome (TS), growth hormone deficiency, celiac disease, malnutrition, as well as many rare diseases. In contrast to its widespread use, current knowledge about the diagnostic performance of growth monitoring is incomplete.[1,2] Growth diagrams define the specificity of a single height measurement. The sensitivity of a single height measurement is unknown for even the most frequent diseases. Also, the sensitivity and specificity of measures involving two or more repeated observations, such as height gain, are unknown. The current state of affairs unfortunately precludes an informed discussion about referral criteria. Referral criteria have been evaluated,[3-7] but these studies have not prevented the appearance of widely different guidelines. For example, the recent UK guideline is based on just one universal height measurement at age 5.[8] In contrast, the Dutch consensus guidelines consist of multiple referral criteria covering infancy, childhood, and adolescence.[9] All in all, current practice differs among practitioners, and practices are not founded on evidence.

In order to make progress, we propose that all applications of growth monitoring should be judged along the conventional Wilson-Jungner criteria for screening tests.[10] Measures of diagnostic performance include sensitivity, specificity, and median referral time. The latter measure is essential to account for the temporal aspect of the problem. Of all diseases that might be detected by monitoring growth, TS is one of the most frequent, occurring in 1:2500 female live births. Only 20-40% of the affected individuals, usually the ones with typical clinical features and somatic abnormalities, are diagnosed in the newborn period.[11,12] Diagnosis of the remaining patients is made during childhood (usually because of growth retardation) or later (because of lack of pubertal development[12]). This makes growth retardation the most important referral criterion in the screening process of TS. The average adult height of untreated women is about 20 cm lower than the mean of the population.[13] Early detection of TS permits the clinician to counsel the family about the consequences of TS, such as an increased risk for cardiac, renal, thyroid, and auditory abnormalities associated with TS. Early detection also allows for the initiation of treatment with growth hormone, which increases final height substantially if started at a young age.[14,15] However, the diagnosis is often made too late,[11,16] so that the results of growth hormone treatment are less favorable. Some work has been done to identify girls with TS earlier using height velocity,[17] but the diagnostic value appeared limited. The goal of the present study is to gain insight into the diagnostic performance of a broader set of referral criteria for auxological screening for TS in the open population.

Method

Screening rules

We investigated screening rules that are suitable for application within the setting of the child health care system. A child that is "screened in" will be referred to a physician for further investigation, eventually leading to the diagnosis of TS. We formulated three archetypal screening rules: an absolute height standard deviation score rule (HSDS), a parental height corrected rule, and a deflection rule (DHSDS). Based on the *absolute HSDS* rule a child is referred if HSDS is lower than some criterion value. The *parental height corrected* rule takes genetic height potential into account by comparing the HSDS of the child to its target height SDS. The target height (TH) is the expected adult height given the heights of the biological parents and corrected for secular trend. For Dutch girls, the relevant formulas are TH=(maternal height + paternal height - 13)/2 + 4.5 and THSDS=(TH - 170.6)/6.5.[18] The *deflection* rule signals whether an abnormal deflection in height occurs in terms of a change in HSDS per year. Table 1 gives the precise definition of each rule, the description of the free parameters, and default values of the parameters as used in the Dutch guidelines.[9]

Table 1 *Three archetypal screening rules for growth monitoring with their definition, scenario parameters, interpretation, default parameter values according to the Dutch consensus guidelines[9], and the parameter values used in the simulation.*

Screening rule	Definition	Parameter	Interpretation	Default value*	Simulation values
Absolute height SDS	For ages 0 to *p* years, refer if SDS < *a*.	*a*	SDS referral level before age *p*	-2.5	-1.5, -2, -2.5, -3, -3.5, -4
	For ages p to 10 years, refer if SDS < *b*.	*b*	SDS referral level after age *p*	-2.5	-1, -1.5, -2, -2.5, -3
		p	Age (in years) at which the referral level changes	Unspecified	1, 2, 3
Parental height corrected	For ages *q* to 10 years, refer if	*c*	SDS cut off level below which SDS must lie	-1.3	-1, -1.3, -1.5, -2, -2.5
	SDS < c, AND	*d*	Difference between target height SDS and SDS	-1.3	-1, -1.3, -1.5, -2, -2.5
	SDS - THSDS < *d*.	*q*	Age (in years) after which the rule is effective	0, 3**	1, 2, 3
Deflection	For any pair SDS_1 and SDS_2 measured at ages X_1 and X_2 (in years), refer if				
	$r \leq X_1 < X_2 < 10$, AND	*e*	Minimal interval (in years) between X_1 and X_2	1***	1, 2, 3
	$X_2 - X_1 \geq e$, AND	*f*	SDS cut off level below which SDS_2 must lie	Unspecified	10, 0, -1, -2, -2.5
	$SDS_2 < f$, AND	*g*	Height velocity change in SDS per year	-0.25	-0.20, -0.25,-0.33, -0.50
	$(SDS_2 - SDS_1)/(X_2-X_1) < g$	*r*	Age (in years) after which the rule is effective	0, 3**	3

* According to Dutch consensus guidelines

** The Dutch consensus guidelines are ambiguous

*** The Dutch consensus guidelines require that three measurements should have been taken, each at least one half a year apart

Each screening rule was implemented in a computer program written in S-Plus,[19] and each rule was applied to longitudinal height data of children with and without TS. For each screening rule, we computed the sensitivity, specificity, and median referral age for specific scenarios. A scenario is a combination of parameters. We defined scenarios by all possible parameter combinations. We first studied the properties of each screening rule separately. Given these results, we defined scenarios that combined the most promising elements of the separate rules, and computed the outcomes for combined scenarios.

Material

Longitudinal height curves from 777 girls with TS were collected from three sources. The National Registry of Growth Hormone Treatment in Children of the Dutch Growth Foundation contains data of all children in the Netherlands receiving growth hormone (GH) treatment. From this registry, 316 girls with TS, born between 1968 and 1996 were selected. In addition, data from 87 girls with TS, born between 1973 and 1988 from the Sophia Children's Hospital and the data of 374 Dutch girls described by Rongen and colleagues[13] were used. The first two sources contain data of girls that were treated with GH and other growth promoting treatment. For this analysis we used only height measurements before treatment. Karyotype, date of diagnosis TS, the presence of congenital anomalies and/or dysmorphic features and parental height were collected when available. The average numbers of measurements per year per child during the first 10 years were 2.2, 0.7, 0.6, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, and 0.5, respectively.

A reference sample of longitudinal height data was retrospectively obtained for a cohort of all girls (n=489) born in 1989 and 1990 in the municipality of Landgraaf, located in the south of the Netherlands. Data were collected from the records of the local child health care centre. These are routinely collected data, and they thus include all measurement errors that are being made in practice. The modal number of observations per girl was 17. Data were collected in 2001, so the oldest girls were about 11 years. The average numbers of visits per year per child during the first 10 years were 8.1, 2.2, 0.8, 0.8, 0.7, 0.7, 0.2, 0.1, 0.1, and 0.8. Table 2 contains additional information about the samples.

Table 2 *Summary statistics for the Turner and reference samples.*

	Turner		Reference	
	Count/Mean	S.D.	Count/Mean	S.D.
Total sample size (n)	777		489	
Total number of measurements	9660		7319	
Mean number of measurements per girl	12.4		15.0	
Mean height for age SDS (ages 0-10 year)	-2.44	1.13	-0.31	1.05
Mean weight for age SDS (ages 0-10 year)	-1.74	1.28	-0.12	1.05
Mean weight for height SDS (ages 0-10 year)	-0.06	1.29	0.12	1.04
Mean BMI for age SDS (ages 0-10 year)	-0.19	1.27	0.11	1.04
Height of both father and mother known	357		203	
Height of only one parent known	3		10	
Height of both parents unknown	417		276	
Mean father's height (cm)	179.4	7.50	178.4	7.57
Mean mother's height (cm)	166.4	6.29	166.7	7.38
Target height (cm)	169.7	5.89	170.6	5.70
Target height SDS	0.06	0.82	0.01	0.88
Gestational age (weeks)	38.9	2.07	39.7	1.61
Dysmorphic features (%) (N=145)				
Cubitus valgus	31			
Large inter-nipple distance	29			
Low hair implantation	21			
Webbed neck	19			
Karyotype (%) (N=327)				
45,X	62			
46,X,iX or 46,X,idic(X)	5			
45,X and 46,XX	5			
45,X and (46,X,iX or 46,X,idic(X))	12			
Other	16			
Median age of diagnosis of TS (years) (N=46)				
45,X (N=27)	6.9	4.94		
Other (N=19)	10.4	4.81		

Statistical analysis

HSDS was calculated with respect to the Dutch height reference data.[20] Parental heights were frequently missing (55% of the Turner group, 58% in the reference group). Deleting incomplete records would not only be wasteful, but would also lead to a selective subsample. Mean HSDS of girls with TS was -3.24 for the subsample with missing parental heights, compared to -2.53 for the subsample with known parental heights, but no such differences were found in the reference group. We imputed these data under the assumption that the data are missing at random[21] using MICE.[22] The method created multivariate imputations by applying sequential linear regressions, where each

incomplete variable was imputed conditional on all other variables in an iterative fashion. The imputation model consisted of the last known HSDS, weight SDS, weight/height SDS, BMI SDS, age, and the height of the other parent. The number of iterations was set to 15. Predictive mean matching was used to create parental heights imputations. The imputation method possesses important properties: it includes parameter uncertainty, preserves the multivariate structure in the data, and has good coverage properties.[23] Figure 1 plots father's height against mother's height separately for the real and artificial data. It shows that the distribution is similar in both groups.

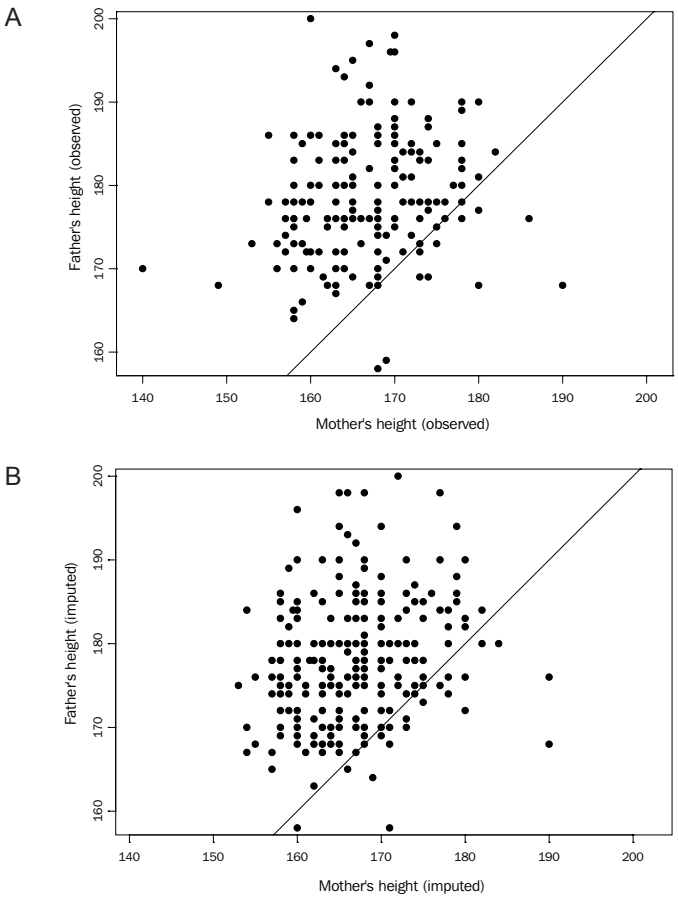


Figure 1 *Father's height plotted against mother's height in the reference sample. (A) Observed data from group in which both parental heights are known (n = 203). (B) Imputed (artificial) data for cases where at least one parental height is missing (n = 286). The reference line indicates the location of equal parental heights.*

We assumed that a child is referred the first time the growth pattern meets the criteria of a given screening rule. Multiple referrals by different rules were treated depending on the type of scenario under investigation. As long as we dealt with separate rules, the same child could be referred according to each rule - that is, as if the screening rules operated in isolation, but in any combined scenario, multiple referrals were counted as one. The screening age of children that were not referred before the age of 10 years was taken as 10 years. TS girls with a referral age of 10 under a given scenario are missed, so the proportion of such girls is the false negative rate (1 - sensitivity). The age of 10 years was chosen because treatment of TS, if indicated, could best be started before that age.

Finally, we synthesized our results by fitting linear regression models to the main outcome variables. These models can be used to predict sensitivity, specificity, and median referral age (MRA) in intermediate cases that were not part of the simulation design.

Results

Sensitivity and specificity

Figure 2A is the ROC plot of scenarios under the absolute HSDS rule. Only scenarios with a true positive rate (sensitivity) of at least 40%, a false positive rate (1 - specificity) of at most 15%, and with cut off age p=3 are plotted. Under the default scenario (-2.5, -2.5) children are referred that have an HSDS <-2.5 (a=-2.5, b=-2.5, p=3). Scenario (-2.5, -2.5) has a sensitivity of 70.2% and a specificity of 93.1%. Scenarios (-3, -2), (-3.5, -2), and (-4, -2) have better sensitivity and specificity for detecting TS. Specificity is, however, still on the low side for screening purposes (95-97%), thus these scenarios might lead to substantial numbers of false positives. Scenarios (-3.5, -2.5) and (-4, -2.5) cut down the number of false positives, at the expense of a loss of sensitivity. The influence of p on sensitivity and specificity was limited.

Performance of the parental height corrected rule was generally better (fig 2B). The current Dutch guideline (-1.3, -1.3) pairs a high sensitivity of 93.5% with a specificity of 95.9%. Rules using more stringent cut off points reduce the number of false positive referrals at the expense of sensitivity. Examples of interesting scenarios are (-1.5, -1.5), (-2, -2), and (-2.5, -2.5). Note that for these cases c=d. The difference with the absolute SDS rule is the extra requirement that THSDS > (c-d) - that is, THSDS > 0 or taller than average parents.

Screening based on the deflection of the growth curve has low sensitivity for rules with a specificity of at least 85% (fig 2C). Though not very sensitive, some deflection rules are highly specific. For example, the rule with e=3, f=-2, and g=-0.25 (not in fig 2C) pairs a sensitivity of 23% with the maximal specificity of 100%. It can be efficient to use such rules in conjunction with more sensitive rules.

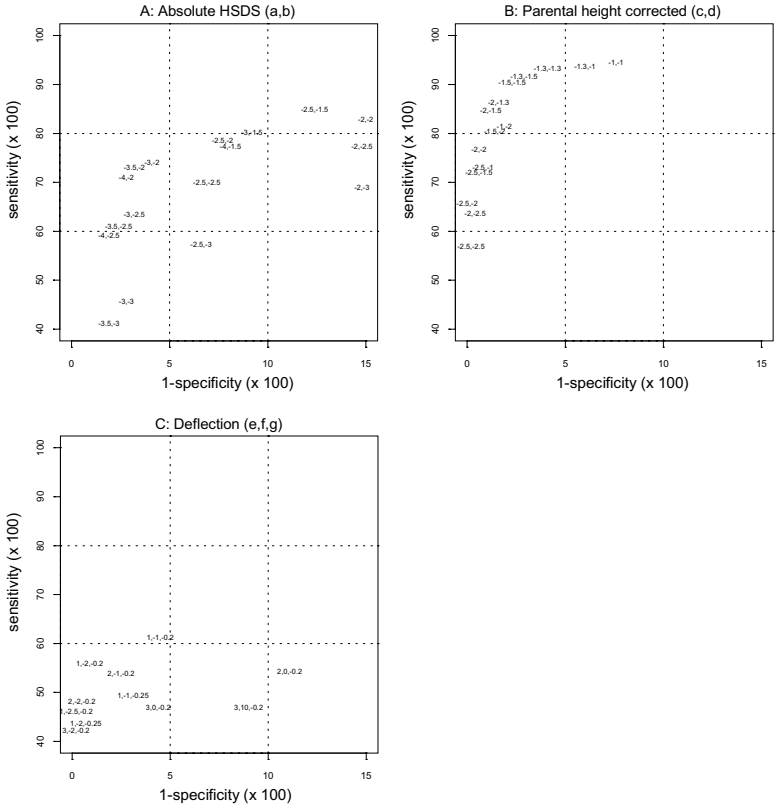


Figure 2 *ROC plot of different scenarios under three archetypal rules. Each scenario is labeled by its parameter values according to table 1. For example, the label (-3, -2) in the left plot indicates the scenario with a =-3 and b =-2. Only scenarios with p = q = r = 3 (cf table 1) are plotted.*

Median referral age

Median referral age in the Turner group generally did not exceed 6 years under the absolute HSDS or the parental height corrected rule (fig 3). Median referral age tends to be lower for higher sensitivity and lower specificity. Thus, more cases imply younger cases. This is especially true under scenarios that correct for parental height. The absolute HSDS rule provides the fastest detection of TS, primarily due to the fact that this is the

only rule that takes measurements during infancy into account. Earlier detection of TS is possible at the expense of specificity, especially if done through the parental height rule.

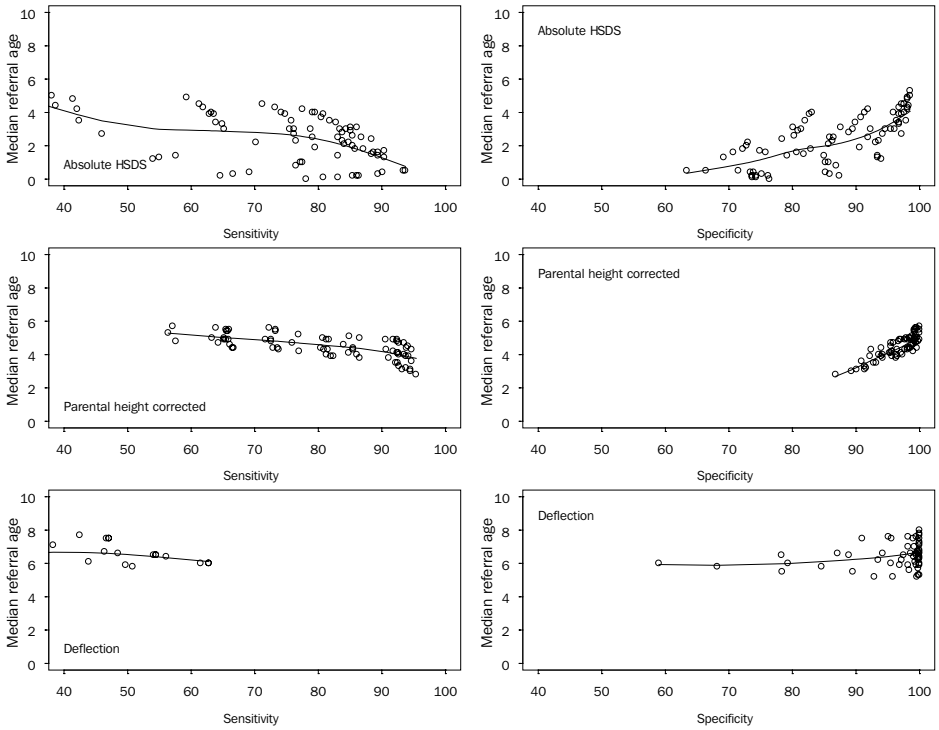


Figure 3 *Median referral age of girls with TS as a function of sensitivity and specificity under each rule. Every dot corresponds to a scenario.*

Predicting sensitivity, specificity, and median referral age

Table 3 contains a synthesis of the results. It gives estimated coefficients of the linear regression for all outcomes. As the proportion of explained variance is generally high, the regression equations can be used to generate fairly accurate predictions for intermediate scenarios not listed in the table. As an example, the estimated sensitivity for scenario (-3, -2) for the absolute HSDS rule is equal to 97.5-3.62*a*-0.60*b*-6.04*ab*=73.3%. The observed values vary between 74.2% (for p=3) and 76.5% (for p=1). In this case, the differences between these observed and predicted values fall within one standard deviation of the residual variation (4.97%).

Table 3 *Regression equations for predicting sensitivity, specificity and median referral age of the absolute HSDS (sds), parental height corrected (phc) and deflection (def) screening rules, the residual standard error, and the proportion of explained variance (r^2).*

Rule	Outcome	Predictive equation	Resid s.e.	r^2
sds	Sensitivity	$97.5 - 3.62a - 0.60b - 6.04ab$	4.97	0.89
	Specificity	$58.4 - 3.55a - 2.32b + 1.20p + 2.00ab$	4.56	0.79
	Median referral age	$1.45 + 0.33a + 1.13b - 0.04p + 0.59ab - 0.32ap + 0.20bp$	0.38	0.94
phc	Sensitivity	$145 + 21.6c + 26.8d + 5.95cd$	3.66	0.91
	Specificity	$67.0 - 11.0c - 11.3d + 4.82q - 3.41cd + 1.06cq - 1.15dq$	0.68	0.95
	Median referral age	$-1.09 - 1.76c - 1.92d + 1.26q - 0.62cd + 0.26cq + 0.32dq + 0.11cdq$	0.10	0.98
def	Sensitivity	$89.9 - 8.59e + 0.40f + 145g$	7.46	0.86
	Specificity	$82.9 + 3.48e - 1.05f - 21.1g$	5.47	0.57
	Median referral age	$5.63 + 0.82e - 0.02f + 2.28g$	0.28	0.87

Combining rules

A child will be referred if he or she meets any of the rules. Sensitivity of a combined rule will be higher than that of its components, while its specificity will be lower.[24] Thus in order to create highly specific combinations, the component rules must have high specificity to start with.

Table 4 shows the diagnostic properties of two combinations. Combining the parental height corrected rule (-2, -2) with the absolute height corrected rule (-3.5, -3) increases sensitivity from 76.9% to 82.4%, decreases specificity from 99.4% to 97.5%, and lowers median referral age to 4.7 years. Observe that this combined rule is inferior to the parental height corrected rule (-2, -1.5) in terms of sensitivity and specificity. The story is different for the combination of the absolute rule with the deflection rule (3, -2, -0.25), which refers children with a HSDS below -2 and a deflection of at least 0.25 SDS per year during at least three years. While this rule detects only 23% of the TS group, there is not a single child in the reference group with this growth pattern. The rule picks up a few new cases. Sensitivity increases from 76.9% to 79.2%, whereas specificity remains at 99.4%. This combined rule is better than comparable parental height corrected rules.

Table 4 *Combining rules using a high specificity strategy. Rows 1-3 list a parental height corrected (phc), an absolute sds (sds) and their combined (phc-sds) rule. Rows 4-6 list a parental height corrected (phc), a deflection (def) and their combined (phc-rule) rule. Row 7 is a single parental height corrected rule that is better than row 3 but not preferable to row 6. MRA = Median Referral Age.*

Row	Rule	Scenario parameters							Sensitivity	Specificity	MRA
		a	*b*	*c*	*d*	*e*	*f*	*g*	(*100)	(*100)	
1	phc			-2.0	-2.0				76.9	99.4	5.2
2	sds	-3.5	-3.0						41.4	98.1	4.8
3	phc-sds	-3.5	-3.0	-2.0	-2.0				82.4	97.5	4.7
4	phc			-2.0	-2.0				76.9	99.4	5.2
5	def					3	-2.0	-0.25	23.3	100.0	7.7
6	phc-def			-2.0	-2.0	3	-2.0	-0.25	79.2	99.4	5.3
7	phc			-2.0	-1.5				84.9	98.8	5.1

Discussion

Growth monitoring is important for detecting TS, but until now no evidence has been available about the diagnostic quality of possible screening procedures. We estimated sensitivity, specificity, and median referral age of TS for three screening rules, and for combinations of these rules. We found that these rules had different performance in discriminating TS. Rules that correct for parental height could identify TS better than rules using the absolute HSDS or rules based on the deflection of growth curves. Combining rules improved performance in particular cases.

The children in our control sample live in the southern part of the Netherlands, and are shorter on average (-0.31 HSDS) than the Dutch reference population. This means that the specificity for the Dutch reference population might be more favorable than estimates based on the shorter population. The equations in table 3 can be used to estimate the size of the effect. For example, setting a=-2, b=-2.5, and p=1 yields a predicted specificity of 82.5%. Had the group been -0.31 shorter, then substituting a=-1.69, b=-2.19, and p=1 predicts a specificity of 78.1% for that group. So the actual specificity for a group that is 0.31 HSDS shorter is here 4.4% lower. In order to eliminate such biases, we added 0.31 HSDS to the measurements of the reference group. The existence of regional height differences implies that the actual false-positive rates can vary across the country. Using the equations in table 3, it is straightforward to compute the effect of regional differences on sensitivity and specificity. Region specific screening rules can be created if the effect is substantial. Similar considerations apply to ethnic minority groups.

Diagnosis of TS is often unnecessarily delayed. Excluding the 20-40% of the patients identified in infancy, the median age of diagnosis is somewhere between 10 and 12 years.[11,16] Including 30% of the early cases into the calculation would lower the median age of diagnoses to the range of 7-8.4 years. By the time of diagnosis, patients were extremely short (mean -3.0 HSDS). We found that the median referral age of most screening rules studied here is between 4 and 6 years. Some rules even identify 50% or more of the cases within the first year. The current policy in the Netherlands is that GH treatment in girls with TS is applied if HSDS < -1.5 and if the child is older than 6 years, but preferably younger than 9. Before the age of 6 years treatment is only started if HSDS < -2.5. Our results suggest that systematic growth monitoring is able to find the large majority of cases in time.

The occurrence of missing parental heights complicated the analysis. It is inappropriate to simply ignore the records with incomplete parental heights because the shorter TS girls drop out more frequently. This leads to sensitivity estimates that are too low. The effect is substantial. For example, using just the complete cases in scenario c=d=-1.3 and q=3 results in a sensitivity estimate of 88.7%, compared to 93.5% based on the imputed sample. As it would be unfair to exclude the incomplete cases only for the parental height rule, sensitivity estimates for other rules would also be affected. Imputation yields unbiased estimates for the TS group as a whole. The precision of these estimates is lower than found in the hypothetical case in which we would have had complete data, but it is higher than obtained in the inappropriate complete-case analysis just discussed.

Our results enable informed decisions about specific choices in screening rules for identifying TS. Although growth charts are also used to detect other anomalies, like growth hormone deficiency or celiac disease, growth monitoring should at least be able to detect TS. If monitoring cannot pick up TS, then it almost certainly will fail in more complicated cases where the effects on growth are less pronounced. It is likely that repeating our study for other diseases will lead to different estimates for sensitivity and specificity. Additional complexities will surface, for example, the lack of a gold standard for diagnosis of growth hormone deficiency. However, such studies would probably not lead to a different ranking among the three rules. We expect that rules that take parental height into account are generally preferable to rules that do not.

The findings appear to be only partially in harmony with published guidelines and proposals.[7,9] As anticipated,[17,25] we found that centile crossing has low sensitivity and specificity, and in this sense, the Dutch guidelines may need re-evaluation. Marked differences occur with respect to the correction for parental height. Hall and Elliman[8] dismissed a correction for parental target height on practical grounds, whereas we found that it represents a substantial improvement, in line with earlier observations by Massa and Vanderschueren-Lodeweyckx.[16]

We conclude that growth monitoring is useful to screen for TS. The parental height corrected rule will refer 60-77% of the girls with TS before the age of 10 at tolerable levels of false positives-that is, at a maximum of 1%. We recommend the use of the combined rule "phc-def" listed in table 4. This rule refers children older than age three if HSDS is below -2 and if either HSDS is more than 2 SD below the target HSDS, or HSDS shows a deflection of 0.25 SDS per year or more during a period of at least three years. This rule picks up almost 80% of the girls with TS, while it refers only 0.6% of the non-TS population. We also recommend that similar research should be done for other diseases, populations, and ages. The results should be synthesized into general evidence-based referral criteria.

References

1. Panpanich R, Garner P. Growth monitoring in children. Cochrane Database Syst Rev 2000;2:CD001443.
2. Garner P, Panpanich R, Logan S. Is routine growth monitoring effective? A systematic review of trials. Arch Dis Child 2000;82:197-201.
3. Hindmarsh PC. Monitoring children's growth: abnormal growth should also be defined by the crossing of height centiles [letter]. BMJ 1996;312:122.
4. Paterson BA. A review of growth monitoring in Aboriginal school children in Australia. BMJ com, electronic letter, 5 March 1999.
5. Mulligan J, Voss LD, McCaughey ES, et al. Growth monitoring: testing the new guidelines. Arch Dis Child 1998;79:318-22.
6. Voss LD. Changing practice in growth monitoring. BMJ 1999;318:344-5.
7. Hall DMB. Growth monitoring. Arch Dis Child 2000;82:10-5.
8. Hall DMD, Elliman D. Health for all children, 4th edn. Oxord: Oxford University Press, 2003.
9. De Muinck Keizer-Schrama SMPF. Consensus 'Diagnostiek kleine lichaamslengte bij kinderen' [Dutch consensus guidelines for short stature]. Ned Tijdschr Geneeskd 1998;142:2519-25.
10. Wilson JM, Jungner G. Principles and practice of screening for disease. Public Health Paper Number 34. Geneva: World Health Organisation, 1968.
11. Sävendahl L, Davenport ML. Delayed diagnoses of Turner's syndrome: proposed guidelines for change. J Pediatr 2000;137:455-9.
12. Gicquel C, Gaston V, Cabrol S, et al. Assesment of Turner's syndrome by molecular analysis of the X chromosome in growth retarded girls. J Clin Endocrinol Metab 1998;83:1472-6.

13. Rongen-Westerlaken C, Corel L, van den Broeck J, et al. Reference values for height, height velocity and weight in Turner's syndrome. Swedish Study Group for GH treatment. Acta Paediatr 1997;86:937-42.
14. Sas TC, de Muinck Keizer-Schrama SM, Stijnen T, et al. Normalization of height in girls with Turner syndrome after long-term growth hormone treatment: results of a randomized dose-response trial. J Clin Endocrinol Metab 1999;84:4607-12.
15. Reiter EO, Blethen SL, Baptista J, et al. Early initiation of growth hormone treatment allows age-appropriate estrogen use in Turner's syndrome. J Clin Endocrinol Metab 2001;86:1936-41.
16. Massa G, Vanderschueren-Lodeweyckx M. Age and height at diagnosis in Turner syndrome: influence of parental height. Pediatrics 1991;88:1148-52.
17. Van den Broeck J, Hokken-Koelega A, Wit JM. Validity of height velocity as a diagnostic criterion for idiopathic growth hormone deficiency and Turner syndrome. Horm Res 1999;51:68-73.
18. Fredriks AM, van Buuren S, Burgmeijer RJF, et al. Groeidiagrammen (2e, herziene en vermeerderde druk) [Growth diagrams, second edition]. Houten: Bohn Stafleu van Loghum, 2002.
19. Insightful Corp. S-Plus 2000 user's guide. Seattle, WA: Insightful, 2000.
20. Fredriks AM, van Buuren S, Burgmeijer RJ, et al. Continuing positive secular growth change in The Netherlands 1955-1997. Pediatr Res 2000;47:316-23.
21. Little RJA, Rubin DB. Statistical analysis with missing data, 2nd edn. Wiley: New York, 2002.
22. Van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med 1999;18:681-94.
23. Brand JPL, van Buuren S, Groothuis-Oudshoorn CGM, et al. A toolkit in SAS for the evaluation of multiple imputation methods. Stat Neerl 2003;57:36-45.
24. Macaskill P, Walter SD, Irwig L, et al. Assessing the gain in diagnostic performance when combining two diagnostic tests. Stat Med 2002;21:2527-46.
25. Van Buuren S, Fredriks AM, Verkerk PH. Comment on Consensus 'Diagnostiek kleine lichaamslengte bij kinderen' [Comment on Dutch consensus guidelines for short stature]. Ned Tijdschr Geneeskd 1999;143:1585-6.

Individual growth curve models for assessing evidence-based referral criteria in growth monitoring

5

Paula van Dommelen[1,2], Stef van Buuren[1], Gladys R.J. Zandwijken[2] and Paul H. Verkerk[1]

[1]TNO Quality of Life, Leiden, The Netherlands
[2]Dutch Growth Foundation, Rotterdam, The Netherlands

Stat Med 2005;24:3663-74.

Summary

The goal of this study is to assess whether a growth curve model approach will lead to a more precise detection of Turner's syndrome (TS) than conventional referral criteria for growth monitoring. The Jenss-Bayley growth curve model was used to describe the process of growth over time. A new screening rule is defined on the parameters of this growth curve model, parental height and gestational age. The rule is applied to longitudinal growth data of a group of children with TS (n=777) and a reference (n=487) group. The outcome measures are sensitivity, specificity and median referral age. Growth curve parameters for TS children were different from reference children and can therefore be used for screening. The Jenss-Bayley growth model, which uses all longitudinal measurements from birth to a maximum age of 5 years with at least one measurement after the age of 2, together with parental height and gestational age can achieve a sensitivity of 85.2 per cent with a specificity of 99.5 per cent and a median referral age of 4.2 (the last measurement between the age of 2 and 5 of each child is considered to be the moment of referral). Sensitivity increases by 2 percentage points when decreasing the specificity to 99 per cent. The Jenss-Bayley growth model from birth to a maximum age of 8 years with at least one measurement after the age of 2, together with parental height results in a sensitivity of 89.0 per cent with a specificity of 99.5 per cent and a median referral age of 6.1. For a specificity of 98 per cent, we obtain a sensitivity of 92.3 per cent. In comparison to conventional rules applied to the same data, sensitivity is about 11-30 percentage points higher at the same level of specificity for the Jenss-Bayley growth rule. We conclude that from the age of 4, growth curve models can improve the screening on TS to conventional screening rules.

Introduction

Measuring height and weight is a routine part of child health care. The goal is to assess whether growth patterns of individual children deviate from the reference population so as to identify diseases and conditions that manifest themselves through abnormal growth. An example is Turner's syndrome (TS), a chromosomal disorder that occurs in about 1 of 2500 female live births and that leads to seriously retarded height. There is an increased risk for cardiac, renal, thyroid and auditory abnormalities associated with TS. Until recently, no evidence-based referral rules existed in growth monitoring. However, recently Van Buuren et al.[1] investigated the diagnostic performance of three conventional rules to detect TS. The first rule is based on the absolute height standard deviation score (absolute HSDS rule), which transforms height into the number of standard deviations above or below the median. The second rule takes genetic height potential into account by comparing the height SDS (HSDS) of the child to its target height (TH) SDS (parental

height corrected rule) and the third rule signals whether an abnormal deflection in height occurs in terms of a change in HSDS/year (deflection rule). In terms of sensitivity and specificity, the absolute HSDS rule and deflection rule appeared to be inferior to the parental height corrected rule. For children with height from birth to the age of 10, the application of the parental height corrected rule will refer 77 per cent of the girls with TS at a specificity of 99.4 per cent. Combining the parental height corrected rule and deflection rule increases sensitivity to almost 80 per cent with a specificity of 99.4 per cent. The median referral age of the parental height corrected rule and the combined parental height and deflection rule is 5.2 and 5.3 years, respectively.[1] The present study extends this work with referral criteria that are based on fitted individual growth curves, parental height and gestational age.

Growth curve models describe growth over time. They are well suited to analyze longitudinal data when the times of measurements are irregularly spaced. The models include parameters that can be estimated from individual longitudinal data. Some parameters correspond to interpretable quantities such as growth at birth, growth velocity, growth acceleration or deceleration. A number of growth curve models have been suggested in the literature and have been shown to be representative at different periods of life.[2,3] We considered several of such models, and used the well-known Jenss-Bayley (JB) growth curve for our data. The JB model describes growth of children from birth to 8 years of age. It was successfully applied by Deming and Washburn[4], Manwani and Agarwal[5], Berkey[6] and Dwyer[7]. Other studies investigated the use of growth curve fitting to compare groups. Rarick et al.[8] compared the growth pattern of normal children and those with Down's syndrome. Nagai et al.[9] studied the growth curves for Japanese patients with Prader-Willi syndrome. Fitted growth curve parameters have also been used as data for analysis of hereditary factors in growth and development.[10,11]

Davenport et al.[12] noticed that for children with TS growth retardation starts during the first year. We expect that such differences in growth of children with TS and without TS will be captured by the parameters of the JB model. The goal of this study is to assess whether a growth curve model approach will lead to a more precise detection of TS than conventional referral criteria for growth monitoring. Our strategy is to estimate the effect of each JB growth parameter on the probability of having TS given the observed growth data (prognostic score). Several thresholds for the prognostic score (PS) are simulated to determine its sensitivity and specificity.

Materials and Method

Material

Longitudinal heights from 777 girls with TS were collected from three sources. The National Registry of Growth Hormone Treatment in Children of the Dutch Growth Foundation contains data of all children in the Netherlands receiving growth hormone (GH) treatment. From this registry, all girls with TS (n=316) were selected. These patients were born between 1968 and 1996. In addition, data from 87 girls with TS, born between 1973 and 1988 from the Sophia Children's Hospital and the data of 374 Dutch girls described by Rongen et al.[13] were used. The first two sources contain data of girls that were treated with GH and other growth promoting treatment. For this analysis we used only height measurements before treatment.

A reference sample of longitudinal height data was obtained retrospectively for a cohort of all girls (n=487) born in 1989 and 1990 in the municipality of Landgraaf, located in the south of the Netherlands. Data were collected from the records of the local child health care. These are routinely collected data, and they thus include all measurement errors that are being made in practice. The modal number of observations/girl was 17. Data were collected in 2001, so the oldest girls were about 11 years old. The data are the same as in van Buuren et al.[1]

Models and statistical analyses

The advantage of the JB growth model compared with conventional referral rules is that all individual growth data are used in the referral criteria. The approach consists of two steps. Step 1 reduces the number of measurements into four interpretable parameters by the JB mixed-effects model. Step 2 consists of the application of heteroscedastic models fit by discriminant analysis which estimate the effect of the JB growth parameters, parental height and gestational age in order to estimate the PS. We will compare the results of the JB rule to the best conventional referral rules, which are the parental height corrected rule and the combination of the parental height and the deflection rule.[1]

Step 1: JB mixed-effects model

Height was modeled by the non-linear JB model. The parameters of this model were estimated by a mixed-effects model. A mixed-effects model assumes that each growth parameter is the sum of a fixed and a random component, where the fixed components are the same for every individual, and the random components may differ between individuals according to a normal distribution. Therefore, this model accommodates individual variations through the random effects, but ties the individuals together through the fixed effects and the covariance matrix of the random effects. A particular advantage of the mixed model is that it borrows strength across individuals in estimating individual parameters. Thus, having few observations in mixed models is less of a problem

compared to the simpler method that estimates the parameters for each individual separately. The random effects represent the deviations of the individual coefficients from their subpopulation average.

First, we studied whether the growth pattern of TS children differs from reference children. In this situation, each of the two groups (reference and TS individuals) is viewed as a subpopulation with its own set of parameter values. Second, we studied the possibility of screening according to the JB rule. If we want to determine whether a new child has TS, we have to choose in which group we estimate the growth parameters for that child. As the prevalence of TS is small, most children are reference children. Therefore, we most likely assume that each child is a reference child.

For each TS girl, we estimated her growth parameters by fitting her height together with the height of all reference children in a mixed-effects model. To obtain good estimates of the parameters of the growth curve for each TS girl, we assume the following minimal data conditions. The girls have to have at least one measurement between birth and 3 months of age, at least one between 3 months of age and 2 years, and at least one between 2 years and, respectively, 5 or 8 years (depending on the age-stopping-point). We have a total of 182 TS girls.

Let n be the number of children, t the age in years and $y_i(t)$ the height (in cm's) of the i th child at age t with $i = 1,...,n$. According to Jenss and Bayley[14] the height of the i th child can be modeled as:

$$y_i(t) = a_i + b_i t - \exp\{c_i + d_i t\} + \varepsilon_{it}$$

where a_i, b_i, c_i, and d_i are unknown parameters at the individual level and ε_{it} is the measurement error at age t.

In addition, we require that the parameters follow a multivariate normal distribution across individuals. Then, for $i = 1,...,n$, the two types of dependencies of the response variable height on age that were used are given by the following model.

$$y_i(t) = \alpha_1 + \alpha_2 t - \exp\{\alpha_3 + \alpha_4 t\} + \varepsilon_{it}$$

with $\alpha_k = \alpha_{k0} + \alpha_{ki}$, α_{k0} fixed effects and α_{ki} random effects, for $k = 1,...,4$.

This model has a linear component $\alpha_1 + \alpha_2 t$ in which the parameter α_2 determines infant growth velocity, and an exponential component $\exp\{\alpha_3 + \alpha_4 t\}$, which determines the decreasing growth rate shortly after birth.[15] The height at birth is represented by $\alpha_1 - \exp(\alpha_3)$. The measurement errors ε_{it} are assumed to be independent across individuals and to be normally distributed with mean zero and a common variance.

For the non-linear mixed-effects procedure it is assumed that the random effects have a multivariate normal distribution with mean vector zero and are independent of the measurement errors. The calculations were performed with the function nlme() in S-plus version 6.1.

Step 2: Discriminant analysis

Discriminant analysis can be used to create a model that explains the grouping of the reference and TS children. Unlike the JB mixed-effects models, which use a weighting process to control the influence of each individual to the estimates by taking into account the number of measurements, the model fit by discriminant analysis considers each individual to contribute equally. This means that children with a small number of measurements, and therefore a lack of information, will be treated the same way as children with a large number of measurements. This can be solved by only including children with a large number of measurements. A disadvantage is that sample selection may occur. As the main results of the mixed-effects model are the parameters (mean, standard error and covariance matrix) of the multivariate normal distribution for the control group and the TS group, we simulated growth parameters from these two multivariate distributions for 1000 individuals/group to overcome the problem of sample selection. We extended the parameters of the two multivariate normal distributions by adding the mean and standard deviation of parental height and gestational age, and adding the correlation between these variables and the growth parameters. With these extended multivariate distributions, we simulated parental height and gestational age for the 1000 simulated individuals/group. The growth parameters, parental height and gestational age of the 1000 simulated individuals are the predictor variables in the discriminant analysis.

As the TS and reference group have different covariance matrices, we used a heteroscedastic discriminant model, which leads to a quadratic discriminant function of the form:

$$d_i(\bar{x}) = \beta_{i0} + \beta_{i1}\bar{x} + \bar{x}^T\beta_{i2}\bar{x},$$

$$\text{where } \beta_{i0} = -\frac{1}{2}(p\log|\Sigma_i| + \bar{\mu}_i^T\Sigma_i^{-1}\bar{\mu}_i),\ \beta_{i1} = \bar{\mu}_i^T\Sigma_i^{-1},\ \text{and } \beta_{i2} = -\frac{1}{2}\Sigma_i^{-1}$$

with Σ_i the covariance matrix of group i and p-variate normal random variables $N_p(\bar{u}_i, \Sigma_i)$ for $i = 1,2$ (TS and reference group) and $p = 4$ (four parameters of the JB growth model).

In this way, relationships among predictor variables with respect to the grouping variable can be expressed by their mean values and their variance-covariance matrices. The results are the probabilities of having TS given the observed growth data, also called the

PS. The PS may differ by age range, the number of growth parameters, parental height and gestational age. The calculations were performed with the function discrim() and predict.discrim() in S-plus version 6.1.

Screening rules based on JB model

The new screening rule uses the PS and several thresholds (h). The PS is obtained under three discriminant models, namely the model with only the JB parameters as predictor variables, the model that adds parental height and the model that adds both parental height and gestational age. Each model was applied to both age groups. This results in 6 outcomes, which are named "0-5 JB screening rule", "0-8 JB screening rule", "0-5 JB parental screening rule", "0-8 JB parental screening rule", "0-5 JB corrected screening rule" and "0-8 JB corrected screening rule".

We formulated the screening rule as follows:

$PS>h$, $h \in (0,1)$

The larger the *PS*, the more likely the individual will have TS. A child with a large *PS* will be eligible for referral to a physician for further investigation. Sensitivity was obtained by the number of TS children who have a $PS>h$, divided by the total number of TS children. Specificity was calculated by the number of reference children who have a $PS \leq h$, divided by the total number of reference children.

Screening rules based on conventional criteria

The best conventional screening rules are the "parental height corrected rule" and the combination of the parental height corrected rule and the deflection rule.[1] The parental height corrected rule takes genetic height potential into account by comparing the HSDS of the child to its THSDS. The TH is the expected adult height given the heights of the biological parents and corrected for secular trend. For Dutch girls, the relevant formulas are TH=(maternal height+paternal height-13)/2+4.5 and THSDS=(TH-170.6)/6.5.[16] The parental height corrected rule is defined as follows:

For ages q to 10 years, refer if $SDS < c$ and $(SDS - THDS) < d$, with c the SDS cut-off level below which SDS must lie, d the difference between THSDS and SDS and q the age (in years) after which the rule is effective. Simulation values are $q = 3$, $c \in \{-2, -2.5\}$, and $d \in \{-2, -2.5\}$.

The deflection rule signals whether an abnormal deflection in height occurs in terms of a change in HSDS/year. In formula: For any pair SDS_1 and SDS_2 measured at ages X_1 and X_2 (in years), refer if $r \leq X_1 < X_2 < 10$ and $X_2 - X_1 \geq e$ and $SDS_2 < f$ and $(SDS_2 - SDS_1)/(X_2 - X_1) < g$, with e the minimal interval (in years) between X_1 and X_2, f the SDS cut-off level below which SDS_2 must lie, g the height velocity change in SDS/year and r the age (in years) after which the rule is effective. For the combination “parental height

corrected and deflection rule", we simulated the values c =-2, d=-2, e =3, f =-2, g =-0.25 and q, r are 3. This rule refers children older than age 3 if HSDS is below -2 and if either HSDS is more than 2 SD below the target HSDS, or HSDS shows a deflection of 0.25 SDS/year or more during a period of at least 3 years.

Imputation

Parental height and gestational age were frequently missing (55 per cent of the TS group and 58 per cent in the reference group for parental height and 73 per cent of the TS group and 3 per cent in the reference group for gestational age). We imputed parental height and gestational age under the assumption that the data are missing at random using multivariate imputation by chained equations (MICE).[17] The method created multivariate imputations by applying sequential linear regressions, where each incomplete variable was imputed conditional on all variables in an iterative fashion. The imputation model consisted of age, height SDS, weight SDS, BMI SDS, weight/height SDS, the height of the other parent and gestational age. The number of iterations was set to 15. Predictive mean matching was used to create parental heights imputations. The imputation method includes parameter uncertainty, preserves the multivariate structure in the data and has good coverage properties.[18] The distribution of father's height against mother's height for the real and artificial data is similar in both groups.

Correction

The children in our reference group live in the southern part of the Netherlands, and are on average -0.31 HSDS shorter than the Dutch reference population. This means that in our method the specificity estimate for our Dutch reference sample would become too low. In order to eliminate this bias, we added 0.31 HSDS to the measurements of our reference sample. With the new HSDS, we estimated the new heights for the measurements for the reference children. The outcome measures are based on the new heights.

Results

JB mixed-effects model

We chose to fit the JB mixed-effects model separately for each group. The mixed-effects model assumes each growth parameter to be the sum of a fixed and a random component. Table 1 shows the least squares estimate of the fixed component along with the standard error of the random component for each group. The fit is represented by the residual variances, Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC). All parameters have significant differences in means between the TS group and the reference group ($p<0.01$). Hardly any differences occur between the residual variances of the TS and the reference group, so the JB mixed-effects model fits the TS and reference groups equally well.

Table 1 *Results of the JB mixed-effects model. The least squares estimate of the fixed component, the standard error of the random component, residual variances, number of observations, number of children, AIC and BIC are presented for each group.*

JB Mixed-effects Model		0-5 years		0-8 years
	TS	Reference	TS	Reference
α_1	71.7 (0.38)	72.6 (0.29)	74.3 (0.34)	76.7 (0.29)
α_2	5.78 (0.088)	8.39 (0.072)	5.17 (0.048)	7.16 (0.053)
α_3	3.19 (0.016)	3.08 (0.013)	3.28 (0.012)	3.24 (0.011)
α_4	-1.34 (0.034)	-1.58 (0.028)	-1.12 (0.025)	-1.21 (0.018)
Residuals	1.071	1.073	1.059	1.188
# of measurements	2960	5806	4172	6332
N	525	476	580	484
	ACI=11910, BIC=12000	AIC=21140, BIC=21240	AIC=16494, BIC=16589	AIC=24117, BIC=24218

Discriminant analysis

The discriminant analysis yields the number of true-positives and false-negatives. The 0-5 JB growth model, which uses all longitudinal measurements from birth to 5 years of age, can separate the TS girls from the reference girls with a sensitivity of 84.5 per cent and a specificity of 100 per cent. After the age of 8, the sensitivity is equal to 91.3 per cent with a specificity of 100 per cent. Including parental height and gestational age and waiting for 5 years result in a sensitivity of 94.7 per cent with a specificity of 100 per cent. Note that these values are fitted from a screening perspective. The parameters of the growth curves are fitted separately for the TS group and the reference group and the group allocation in the discriminant analysis was known. In an actual screening context, the information as to which group each case belongs is not present. The following step corrects for this.

JB growth parameters for screening

For 182 TS girls we fitted each TS girl with the 1000 simulated reference children in a mixed-effects model and calculated the sensitivity by using the same discriminant function (based on the simulated values) as before. The results are shown in Table 2. For the JB corrected rule with 0.5-2 per cent false-positives, we obtained a sensitivity of between 85.2 and 87.4 per cent for growth from birth till 5 years of age and 89.0-92.3 per cent from birth till 8 years. Doubling the amount of false-positives from 1 to 2 per cent hardly improves sensitivity for growth from birth till 5 years of age.

Table 2 *Sensitivity, specificity and median referral age for the JB screening rule, the JB parental screening rule and the JB corrected screening rule from birth till 5 and 8 years, respectively.*

	0-5 years			0-8 years		
	JB screening	JB parental	JB corrected	JB screening	JB parental	JB corrected
Specificity (per cent)	Sensitivity	Sensitivity	Sensitivity	Sensitivity	Sensitivity	Sensitivity
99.5	83.0	85.2	85.2	86.3	89.0	89.0
99	84.1	86.7	87.2	89.0	90.1	90.1
98	84.6	86.8	87.4	90.7	92.3	92.3
Median referral age	4.2			6.1		

The median referral age for the 0-5 JB rule is 4.2 and for the 0-8 JB rule is 6.1. Note that the last measurement between the age of 2 and 5 or 8 of each child is considered to be the moment of referral. Figure 1 shows the receiver operation characteristic (ROC) curves for the JB rules.

Comparison with conventional screening rules

We applied the parental height corrected rule to the same 182 TS girls. A total of 85 per cent (to 5 years of age) to 87 per cent (to 8 years of age) have at least one measurement after the age of 3 and are presented in the following sensitivity and specificity. The parental height corrected rule has a maximum sensitivity of 57.1 per cent with a specificity of 99.8 per cent from birth till 5 years of age and a sensitivity of 69.6 per cent with a specificity of 99.4 per cent from birth till 8 years of age. The best JB rule from birth till 5 years of age has a sensitivity of 88.3 per cent with a specificity of 99.8 per cent and from birth till 8 years of age has a sensitivity of 90.5 per cent with a specificity of 99.4 per cent. The best conventional rule (for a high specificity) is the combined parental height and deflection rule. As this rule starts at the age of 3 and must have a minimum period of 3 years, we can only apply this rule to children older than 6 years of age. For the TS children from birth till 8 years of age, the sensitivity is equal to 74.7 per cent with a specificity of 99.9 per cent. The best JB rule has an 11 percentage point higher sensitivity with equal specificity.

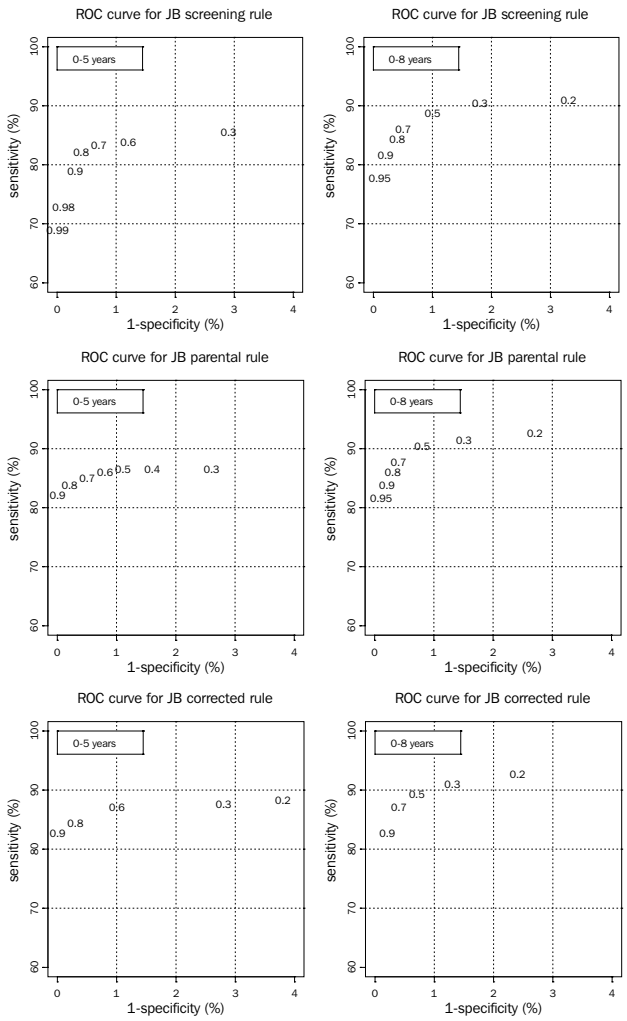


Figure 1 *The ROC curves for the JB screening rule (based on growth parameters), the JB parental screening rule (based on growth parameters and parental height) and the JB corrected screening rule (based on growth parameters, parental height and gestational age) from birth till 5 and 8 years, respectively. Several thresholds for the prognostic score (h) are given.*

Discussion

The use of individually fitted growth curves for detecting TS leads to better results in sensitivity and specificity than the conventional screening rules. Sensitivity increases by up to 30 per cent by specificities near 100 per cent. This improvement may be caused

by the fact that conventional screening rules only use part of the information of a growth curve. The JB rule incorporates all available information of the growth process. When it is possible to wait for 4 years, as is typically the case for TS, our results suggests that growth monitoring according to the JB rule generally improves upon the conventional screening rules. The JB rule and the conventional screening rule were developed and tested using the same sample. However, this sample is not representative of the larger population of Dutch girls. Therefore, the absolute value of sensitivity and specificity may be different for our population. However, we consider it to be very unlikely that our conclusion that the JB rule is superior to the conventional rules will be different in another sample.

The results show that the four JB parameters taken together are very effective at separating the two populations, but it would be interesting to know which of the four are most important. Table 1 compares the two sets of parameters, and expressing the differences between them in terms of their standard errors. This shows that α_2 is more than 4 times as important as α_1, α_3 and α_4 (t=48 versus t=3, -10 and -10). Therefore, the main growth defect appears to be in the linear part of the JB growth curve model. This means biologically that TS growth appears as constant centile crossing downwards or negative deflection, which suggests that a simpler approach than the JB growth curve model may be equally effective. Therefore, we repeated the analysis using HSDS instead of height, and summarized each child's growth as a linear trend with two parameters; HSDS at birth and a slope. Sensitivity increases from 79 to 89 per cent for the linear HSDS rule from birth to both 5 and 8 years of age with 0.5-2 per cent false-positives. This means that sensitivity is large for the linear HSDS model. However, sensitivity is less optimal than the JB growth curve model. The results of the linear HSDS rule with the combined parental height deflection rule are almost similar, which is not surprising as both rules investigate the deflection of HSDS in combination with maternal and paternal height. The differences lie in the starting point of the linear regression and the number of measurements used to obtain the slope (i.e. the combined rule compares two HSDS measurements successively while the linear HSDS rule takes all measurements into account).

The estimation of specificity and sensitivity for the JB rule is obtained from the same sample of children for which we developed the model. This estimation would be more convincing if they were obtained from a validation sample of children. However, at present we do not have access to a suitable validation data set. A split-sample technique, in which one half of the sample is used to develop the model and the other half is used to measure its performance, was contemplated but the number of children in each group would become too low. Obviously, independent validation and replication would further enhance the credibility of our results.

Requirements of the JB rule as applied here are to have a least two measurements before age 2 and having at least one measurement after the age of 2 to obtain good estimates of the growth parameters. We do not recommend using the JB rule when the minimal data condition are not met, as the growth curve of a TS child will be smoothed too much toward the average curve for the reference population, which makes it more difficult to distinguish TS from reference children (low sensitivity). When our requirements are met, we see that the predicted curves for TS children will not be smoothed so much toward the reference population. The size of residuals is a good predictor of the smoothness toward a reference population. When the residuals are small, we obtain a good estimation of the growth parameters and not so much pulling toward the reference population. The standard deviation of the residuals of the 182 predicted curves for all TS cases is equal to 0.79. This is less than 1 cm which is small considering the fact that the height range varies between 40 and 130 cm. When the data conditions are not met, conventional screening rules are recommended. More work is needed to determine fruitful combinations of both types of rules.

The median referral age for the JB rule from birth to 5 years of age is 4.2 years. A decrease in median referral age might be obtained by applying the JB rule from birth to the first measurement after the age of 2. Choose a threshold with a large specificity. Some TS girl will be referred soon after the age of 2 and all other TS girls have to wait until the age of at most 5 years. In this case, the median referral age can be minimized while sensitivity and specificity will stay the same.

In this paper, we used the JB model, but we also considered the first two components of the infancy-childhood-puberty (ICP) growth curve model. The ICP growth curve decomposes linear growth mathematically into three additive and partly superimposed components-infancy, childhood and puberty.[3] The starting point of the childhood component represents the age of onset. Most healthy infants show an abrupt increase in growth rate.[3] Finding the exact age at onset of the childhood phase can only be determined when the interval between measurements is small. As a great number of children do not satisfy this condition, we decided to fix the age of onset at 9 months. Due to computational problems (i.e. convergence problems) of the first two components of the ICP growth curve in the mixed-effects model, we decided to choose the JB growth curve. We also fitted the count model, but this model had convergence problems as well. We applied a polynomial of degree four. To make a comparison between the fit of the JB and a polynomial of degree four (P4), we compared the residuals, AIC and BIC. JB had a better fit than P4.

Application of the JB rule requires a computer system to perform the calculations. Child health care is slowly adopting the use of computers to record the biometrical data. Where this is done, we think that the JB rule can be implemented without too much effort.

Acknowledgements

We thank Ine Bonnemaijer for her co-operation in obtaining the reference group data. We thank Sabine de Muinck Keizer-Schrama for making available data from the Sophia Kinderziekenhuis. We thank Ciska Ronger Westerlaken for the use of the data from the Turner reference study.

References

1. Buuren van S, Dommelen van P, Zandwijken GRJ, Grote FK, Wit JM, Verkerk PH. Towards Evidence Based Referral Criteria for Growth Monitoring. Arch Dis Child 2004;89:336-41.
2. Thissen D, Bock RD. Linear and nonlinear curve fitting. In A. von Eye (Ed.), Statistical methods in longitudinal research, Volume II: Time series and categorical longitudinal data. Orlando: Academic press, 1990; pp 289-318.
3. Karlberg J, Engstrom I, Karlberg P, Fryer JB. Analysis of linear growth using a mathematical model. I. From birth to three years. Acta Paediatr Scand 1987;76:478-88.
4. Deming J, Washburn AH. Application of the Jenss curve to the observed pattern of growth during the first eight years of life in forty boys and forty girls. Hum Biol 1963;35:484-506.
5. Manwani AH, Agarwal KN. The growth pattern of Indian infants during the first year of life. Hum Biol 1972;45:341-9.
6. Berkey CS. Comparison of two longitudinal growth models for preschool children. Biometrics 1982; 38:221-34.
7. Dwyer JT, Andrew EM, Berkey C, Valadian I, Reed RB. Growth in "new" vegetarian preschool children using the Jenss-Bayley curve fitting technique.American Journal of Clinical Nutrition 1983;37:815-27.
8. Rarick GL, Wainer H, Thissen D, Seefeldt V. A double logistic comparison of growth patterns of normal children and children with Down's syndrome. Hum Biol 1975;2:339-46.
9. Nagai T, Matsuo N, Kayanuma Y, Tonoki H, Fukushima Y, Ohashi H, Murai T, Hasegawa T, Kuroki Y, Niikawa N. Standard growth curves for Japanese patients with Prader-Willi syndrome. Am J Med Genet 2000;95:130-4.
10. Baker LA, Reynolds C, Phelps E. Biometrical analysis of individual growth curves. Behav Genet 1992;22:253-64.
11. Van Dommelen P, de Gunst MC, Van der Vaart AW, Boomsma DI. Genetic study of the height and weight process during infancy. Twin Res 2004;7:607-16.

12. Davenport ML, Punyasavatsut N, Stewart PW, Gunther DF, Savendahl L, Sybert VP. Growth failure in early life: an important manifestation of turner syndrome. Horm Res 2002;57:157-64
13. Rongen-Westerlaken C, Corel L, van den Broeck J, Massa G, Karlberg J, Albertsson-Wikland K, Naeraa RW, Wit JM. Reference values for height, height velocity and weight in Turner's syndrome. Swedish Study Group for GH treatment. Acta paediatr 1997;86:937-42.
14. Jenss RM and Bayley N. A mathematical method for studying the growth of a child. Hum Biol 1937;9:556-63.
15. Marubini E, Milani S. Approaches to the analysis of longitudinal data. In: Falkner F., Tanner J.M., eds. Human Growth Vol.3 : Methodology, ecological, genetic, and nutritional effects on growth. 2nd ed. New York and London, Plenum Press, 1986; pp 79-94.
16. Buuren S van, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med 1999;18:681-94.
17. Fredriks AM, Buuren S van, Burgmeijer RJF, Verloove-Vanhorick SP, Wit JM. Groeidiagrammen. Handleiding bij het meten en wegen van kinderen en het invullen van groeidiagrammen. Houten: Bohn Stafleu van Loghum, 2002. 2e herz. dr.
18. Brand JPL, Buuren S van, Groothuis-Oudshoorn CGM, Gelsema ES. A toolkit in SAS for the evaluation of multiple imputation methods. Statistica Neerlandica 2003;57:36-45.

Developing evidence-based guidelines for referral for short stature

6

Floor K. Grote[1#], Paula van Dommelen[2#], Wilma Oostdijk[1], Sabine M.P.F. de Muinck Keizer- Schrama[3], Paul H. Verkerk[2], Jan Maarten Wit[1] and Stef van Buuren[2,4]

[1]Leiden University Medical Center, Leiden, The Netherlands

[2]TNO Quality of life, Leiden, The Netherlands

[3]Erasmus MC - Sophia Children's Hospital, Rotterdam, The Netherlands

[4]University of Utrecht, Utrecht, The Netherlands

[#]These authors contributed equally to this work

Arch Dis Child 2008;93:212-7.

Abstract

Objective: To establish evidence-based guidelines for growth monitoring on a population basis.

Study design: Several auxological referral criteria were formulated and applied to longitudinal growth data from four different patient groups, as well as three samples from the general population.

Results: Almost 30% of pathology can be detected by height standard deviation score (HSDS) below -3 or at least two observations of HSDS below -2.5 at a low false-positive rate (<1%) in 0-3-year-old infants. For 3-10-year olds, a rule concerning distance to target height of >2 SD in combination with HSDS <-2.0 has the best predictive value. In combination with a rule on severe short stature (<-2.5 SDS) and a minor contribution from a rule on "height deflection", 85.7% of children with Turner's syndrome and 76.5% of children who are short because of various disorders are detected at a false-positive rate of 1.5-2%.

Conclusions: The proposed guidelines for growth monitoring show high sensitivity at an acceptably low false-positive rate in 3-10-year-old children. Distance to target height is the most important criterion. Below the age of 3 years, the sensitivity is considerably lower. The resulting algorithm appears to be suitable for industrialized countries, but requires further testing in other populations.

Introduction

Growth monitoring in infancy and childhood has been part of preventive child health programs for more than a century, and short stature or growth retardation is regarded as a relatively early sign of poor health. Despite this longstanding and wide acceptance of growth monitoring, there is little evidence for its effectiveness and efficiency.[1] In developing countries, growth monitoring is primarily aimed at detecting malnutrition. In industrialized countries, the major purpose of growth monitoring is early detection of growth disorders, such as Turner's syndrome (TS), growth hormone deficiency and celiac disease (CD).

For early identification of children with abnormal growth, one requires good growth-monitoring systems as part of preventive child health programs, well-defined and accurate referral criteria, and good diagnostic work-up after referral. Although most industrialized countries have a child health program that includes regular growth monitoring, there is a wide diversity in protocols used for growth monitoring and diagnostic work-up of growth disorders, and a virtual absence of experimental studies on the efficacy of these screening and diagnostic procedures.[2] Few guidelines have been published on referral criteria and diagnostic work-up for children with impaired growth, and these are based

on consensus meetings rather than experimental evidence.[3,4] In the few experimental studies on growth monitoring, various referral criteria have been used.[5-7]

In the Netherlands, a consensus meeting was held in the mid-1990s to establish auxological referral criteria.[3] Three auxological parameters were chosen: height standard deviation score (HSDS), change in HSDS (HSDS deflection), and distance between height and target height SDS. Additional criteria included clinical signs (disproportion or dysmorphism), specific symptoms (such as those associated with emotional deprivation), or previous history of low birth weight and/or length (small for gestational age, SGA). Thereafter, however, it was shown that application of these auxological criteria would lead to far too many unnecessary referrals.[8]

Consequently we started a project aimed at producing evidence-based guidelines for growth monitoring, with a high positive predictive value at an acceptable false-positive rate. We previously studied the predictive value of various auxological criteria for the detection of TS[9], and evaluated the auxological parameters of patients with various causes of growth failure referred to pediatric clinics. In this report, we describe the performance of the best screening rules in terms of sensitivity and specificity in four groups of patients with growth disorders and in three reference samples, and propose that these can be used in growth-monitoring protocols.

Methods

Materials

Longitudinal height and weight data from four different patient groups and three reference populations were used. Each group was analyzed separately. For the patient groups, only measurements before or at age of diagnosis or start of diet (CD cohort) were taken into account.

The first group of patients consisted of 777 girls with TS, collected from three sources and previously described by van Buuren et al.[9] The second group contained new patients referred for short stature to the outpatient clinics of the general pediatric departments of two hospitals (Erasmus MC - Sophia Children's Hospital, Rotterdam and Spaarne Hospital, Haarlem) in 1998-2002. Of 542 children referred to the clinic, 27 were found to have a pathology (mainly growth hormone deficiency (n=7), CD (n=7) and TS (n=3)). Only these 27 children were included in the analyses. The third group consisted of patients with cystic fibrosis (CF) collected from three major CF clinics in the Netherlands: Erasmus MC - Sophia Children's Hospital in Rotterdam (n=166), University Hospital Maastricht (n=30) and Juliana Children's Hospital in The Hague (n=20). The last group contained patients with CD consisting of two separate subgroups: (1) a retrospective

study described by Damen et al,[10] in which they studied catch up growth in patients with celiac disease; (2) a prospective study on catch up growth by Boersma et al.[11]

The first reference sample was obtained from the Social Medical Survey of Children Attending Child Health Clinics (SMOCC) cohort, a nationally representative cohort of 2151 children born in The Netherlands in 1988-1989, consisting of length and weight data for children up to the age of 2.5 years.[12] The second reference population was a cohort of all children born in the years 1989 and 1990 in Landgraaf and Kerkrade, located in the southern part of the Netherlands (”Limburg”, n =970).[8] The third population was a sample of children born in 1985-1988, attending school doctors between 1998 and 2000 in Leiden and Alphen aan den Rijn, the Netherlands (‘‘ZHN’’, n = 400).[12a]

Screening rules

By combining evidence found in previous studies, three auxological referral criteria were formulated. Only criteria of practical value for child health programs were considered.

The first rule takes genetic height potential into account by comparing the HSDS of the child with its target height in combination with a HSDS below a certain cut-off. In our earlier study on TS,[9] as well as in a study on a mixed population of short children (unpublished work), we found that this combination offers the best predictive value. We calculated the test characteristics for a distance between HSDS and target height of more than 2 SDS, with cut-off points for height SDS of -2, -1.5 or -1.0 SDS. This rule was labeled “short for target height”.

The second rule concerns HSDS. HSDS is generally considered one of the most important referral criteria, especially when parental height is not available.[2,4] To keep the percentage of false-positives low, we chose, for historical and pragmatic reasons, a cut-off of -2.5 (~0.6th centile), as it is the lowest line on various growth charts. This rule was labeled “very short”.

The third rule applies to a deviation from the expected growth channels, expressed as either height velocity (cm/year) or SDS for age or a change in HSDS. The change in HSDS is thought to be more suitable, because it better reflects the deviation from canalization of the growth curve, and because height velocity depends not only on age but also on HSDS position. Although the usefulness of low growth velocity for growth screening appears limited,[7,13] it has long been considered the most important growth parameter, and many clinicians can show examples of cases where deflection of the growth curve is the only indication of a growth disorder-for example, an acquired growth hormone deficiency caused by a brain tumor, or primary hypothyroidism caused by Hashimoto disease. Van Buuren et al[8] found that a ”height deflection” of more than 0.25 SDS per year would lead to a large number of false-positives. The predictive value of a deflection can be

improved if one demands a continuous deflection over 3 years (e.g., 0.25 SDS/year over at least 3 years),[9,13] a larger deflection over an undefined time interval (e.g., a deflection of >1.0 SDS), or in combination with an absolute HSDS <-2. In the present analysis, we combined various expressions of "height deflection" (per year or cumulative) with various cut-off points for HSDS (<-2.0, <-1.5 or <-1.0). This decision rule was labeled "height deflection". We decided that deflection with a cut-off of 1.0 SDS over an undetermined time interval would be most practical, as this should detect both a slow and fast bend in the growth curve, and several growth reference diagrams include lines with a distance of 1 SDS.[14] In countries where a distance of 0.67 SD is used, a deflection of 1 SD can be easily assessed by multiplying the 0.67 SD interval by 1.5.

Analytic procedure

Length, height, weight, target height, body mass index and weight for length or height were expressed as SDS, using recent Dutch, Turkish and Moroccan reference data.[15-18] All criteria were first analyzed for all age groups. As growth curves in the first 3 years can cross SDS lines when birth length SDS is far from target height SDS, and length measurements are less accurate, specificity of the various rules is expected to be lower than in later years, leading to too many referrals.[8] We therefore performed separate analyses in two age groups (0-3 and 3-10 years), and calculated test characteristics for different cut-off values (HSDS -3.0, -2.5, -2.0, -1.5 and -1.0) and other additive parameters.

Data on parental height were often (4-58%) missing from the various datasets. We imputed these data under the assumption that data were missing at random using multivariate imputation by chained equations (MICE).[19,20] The imputation model consisted of the last known HSDS (except for the CF population, where we chose the HSDS closest to the age of 5 years instead because in most children catch-up growth has resulted in a normal height at this age[21]), HSDS, weight SDS, weight for height SDS, body mass index SDS, sex (except for the TS group as these were all girls), HSDS of the father and/or HSDS of the mother (if available), ethnicity (except for the TS and Limburg cohort) and, for the CF and CD cohorts, age at diagnosis or start of diet. The number of iterations was set to 15. Predictive mean matching was used to create parental height imputations.

Target height

Target Height (TH) was calculated by Tanner's method with an additional correction for secular trend:

TH(boys) = ((FH + MH+ 13)/2) + 4.5

TH(girls) = ((FH + MH - 13)/2) + 4.5

where FH is father's height and MH is mother's height. The target height standard deviation score (THSDS) was calculated as THSDS(boys) = (TH(boys)-184)/7.1 and THSDS(girls) = (TH(girls)-170.6)/6.5.

Calculations were based on the assumption that a child is referred if the growth pattern meets the criteria of a given screening rule for the first time. If a child only has one measurement, the child cannot comply with criteria concerning deflection or repetition and is therefore considered as non-referred. All rules were analyzed separately as well as in combination with the others. A false positive rate of <1% for the separate rules and <2% for the combined rules was assumed to be acceptable from the perspective of preventive child health care.

Results

Table 1 shows the number of children per age group and the mean number of measurements. Applying the three auxological criteria separately to all age groups resulted in a high number of referrals in the general population (presumably false-positives) (data not shown). This was primarily due to referrals in the 0-3 year group, the "height deflection" and "short for target height" rules producing a high false-positive rate. Extra criteria were added and the cut-off points were varied for children under the age of 3 years. The performance of the different rules was then tested in the two age groups.

Table 1 *Number of children (N) and mean number of measurements (n) per child in each group.*

		Limburg n=970	ZHN n=400	SMOCC n= 2151	Turner's syndrome n= 777*	Short stature due to pathology n= 27	Cystic Fibrosis n=216	Celiac disease n=102
Age group	Number of measurements	N (n)	N (n)	N (n)	N (n)	N (n)	N (n)	N (n)
0-3	≥1 AND at least 1 weight measurement before 0.1 years	931 (11)	341 (11)	1942 (8)	353 (4)	23 (6)	89 (5)	86 (7)
	≥2 with 0.5-1 year interval AND at least 1 weight measurement before 0.1 years	810 (12)	321 (14)	1835 (9)	158 (8)	15 (9)	32 (10)	66 (12)
3-10	≥1	958 (3)	361 (4)	0	524 (5)	17 (3)	25 (2)	22 (4)
	≥2	893 (4)	339 (4)	0	472 (6)	13 (3)	14 (3)	16 (5)

*492 children had measurements under the age of 3 years.
CD, celiac disease; CF, cystic fibrosis; Limburg, all children born in the years 1989 and 1990 in Landgraaf and Kerkrade; SMOCC, Social Medical Survey of Children Attending Child Health Clinics; SSP, short stature due to pathology; TS, Turner's syndrome; ZHN, children born in 1985-1988, attending school doctors between 1998 and 2000 in Leiden and Alphen aan den Rijn.

Table 2 shows scenarios with the best test performance, and tables 3 and 4 show the yield of these best scenarios in terms of sensitivity (true-positives) and 1-specificity (false-positives), respectively.

Table 2 *Referral criteria with the best test characteristics.*

Rule	Criteria	Rule No
0-3 years		
Repeatedly very short: at least twice a length SDS < -2.5	$HSDS_1$ < -2.5 and $HSDS_2$ < -2.5 AND $0.5 \leq Age_2 - Age_1 < 1$ year AND [birth weight >=2500 grams or if no birth weight available than first measurement within 0.1 year (5 weeks) with weight SDS ≥ -2, and gestational age ≥ 37 weeks (or not available)]	1
Extremely short: at least once a length SDS < -3	HSDS < -3 AND [birth weight ≥ 2500 grams or if no birth weight available than first measurement within 0.1 year (5 weeks) with weight SDS ≥ -2, and gestational age ≥ 37 weeks (or not available)]	2
Combination of rule 1+2		3
3-10 years		
Short for target height	HSDS - THSDS < -2 AND HSDS < -2	1
Very short: length SDS < -2.5	HSDS < -2.5	2
Height deflection	Delta HSDS < -1 AND HSDS < -2	3
Combination of rule 1, 2 and 3		4

For children under the age of 3 years, the true-positive rate for pathology is modest, if the false-positive rate has to be kept low. The best rule consists of a HSDS <-2.5 at least twice within 1 year (very short repeated) or a HSDS <-3 (extremely short), confined to infants born at or after 37 weeks of gestational age (or when information on gestational age is not available) and born with a weight ≥ 2500 g (if birth weight was not available, the first measurement within 0.1 year (5 weeks) with a weight SDS ≥ -2 was used). With this rule, 14.7% of the children with TS can be detected, at a false-positive rate of <1%. This is probably an underestimation, because the value of 7.1% for a repeated HSDS < -2.5 increased to 15.8% when only the subgroup of children with more than two measurements was assessed. The "short for target height" rule did not result in acceptable test characteristics.

Above the age of 3 years, 85.7% of the children with TS and 76.5% of the children with mixed pathology could be detected by the combination of the "short for target height" rule, the "very short" rule and the "height deflection" rule.

If a stepwise approach is taken for 3-10-year-old children, the "very short" rule would add 42 patients (7.7%) to the 76.9% of girls with TS who complied with the "short for target height" rule. For the group of children with short stature due to mixed pathology, three cases (17.7%) would be added to the 58.8% of children who complied with the "short for target height" rule. The addition of this rule would increase the false-positive

rate by 0.3% (one child) in the ZHN cohort and 0.7% (seven children) in the Limburg cohort. Applying the ”height deflection” rule after the two other rules would only add a few extra patients (four patients (0.8%) for TS, none for the children with mixed pathology), and the false-positive rate would increase by 0.6% (two children).

Table 3 *Sensitivity of several auxological rules for four different patient groups (true-positives).*

Rule		TS (%)	SSP (%)	CF (%)	CD (%)
0-3 years	Repeatedly very short*	7.1	14.8	0.0	1.2
	Extremely short	13.0	26.1	6.7	4.7
	Combination	14.7	26.1	6.7	4.7
3-10 years	Short for target height	76.9	58.8	8.0	27.3
	Very short	74.0	58.8	4.0	18.2
	Height deflection**	13.4	17.6	0.0	18.2
	Combination	85.7	76.5	8.0	27.3

If a child has only 1 measurement, the child cannot be referred according to the “repeatedly very short rule” and the absolute “height deflection” rule.
*In the subgroup with ≥2 measurements, sensitivity would be 15.8% for TS, 26.7% for mixed pathology, and 1.5% for CD.
**In the subgroup with ≥2 measurements, sensitivity would be 14.8% for TS, 23.1% for mixed pathology, and 25.0% for CD. CD, celiac disease; CF, cystic fibrosis; SSP, short stature due to pathology; TS, Turner’s syndrome.

Table 4 *Estimated percentages of referrals in three reference populations (false-positives).*

Rule		Limburg	ZHN	SMOCC
0-3 years	Repeatedly very short*	0.2	0.0	0.4
	Extremely short	0.2	0.6	0.7
	Combination	0.3^	0.6	0.9^
3-10 years	Short for target height	0.7	1.1	NA
	Very short	0.9	0.8	NA
	Height deflection**	0.1	0.8	NA
	Combination	1.5#	1.9#	NA

If a child has only 1 measurement, the child cannot be referred according to the “repeatedly very short” rule and the absolute “height deflection” rule. *Based on subgroup with ≥2 measurements specificity is 0.2% for Limburg and 0.4% for SMOCC **Based on subgroup with ≥2 measurements specificity is 0.1% for Limburg and 0.9% for ZHN ^No significant difference between Limburg and SMOCC for the combined rule 0-3 years (χ^2 (1)=2.79, p=0.10) #No significant difference between Limburg and ZHN for the combined rule 3-10 years (χ^2 (1)=0.38, p=0.54) NA, not available; SMOCC, Social Medical Survey of Children Attending Child Health Clinics; ZHN, children born in 1985-1988, attending school doctors between 1998 and 2000 in Leiden and Alphen aan den Rijn.

Discussion

We have established evidence-based guidelines for growth monitoring on a population basis. In 0-3-year-old infants, after exclusion of babies born preterm and with a low birth weight, we found that a HSDS <-3 or at least two observations of a HSDS <-2.5 within 1 year gives the best performance at a low false-positive rate. However, only 14.7% of the children with TS and 26.1% of the children with other growth disorders could be detected with these rules. For 3-10-year-old children, the “short for target height”

rule in combination with the "very short" rule and a minor contribution of the "height deflection" rule detected 85.7% of children with TS and 76.5% of children who were short because of various disorders at a low false-positive rate.

The low efficacy and efficiency of growth monitoring between 0 and 3 years of age, particularly for rules involving target height and length deflection, is probably mainly caused by the low correlation between length and mid-parental height at birth, which rapidly increases during the first 3 years of life.[22] Crossing over SDS lines in this age period is therefore not unusual. This is in line with our observation that referral based on a low length velocity or a large distance to target height would lead to too many referrals in this age group, and confirms our earlier data.[8] For this age group, the only useful referral rule was based on an extremely low or repeatedly low HSDS. Only 15-26% of the growth disorders studied was detected, and even fewer infants with CF or CD. This is in concurrence with our studies on CF and CD, in which we found that body mass index is a better auxological tool than length at this young age.

In concurrence with our earlier observations on TS,[9] we found that, also in a mixed set of growth disorders diagnosed in a pediatric clinic, the best decision rule for detecting children older than 3 years with pathology is the "short for target height" rule. This result contrasts with earlier speculations that this parameter might be too inaccurate because of the uncertainty of parental height.[4] From the preventive health care perspective, the ''height deflection'' rule is of little use. We propose to keep this rule in the algorithm, as it is important that the rare cases of growth deflection due to acquired growth disorders are detected in good time. To keep the false-positive rate low, we combined HSDS deflection with a HSDS <-2.0, but a severe deflection irrespective of the HSDS reached should be considered as an alarm signal.

Not only auxological rules are important, but also a number of clinical symptoms and signs. If medical history reveals that birth weight and/or length was low, and HSDS is <-2.0 from the age of ~3 years, the diagnosis of persistent short stature after SGA can be made. It is known that ~10% of children born SGA do indeed remain short and do not achieve normal adult height.[23] Referral to a growth clinic is needed for further diagnostic tests and for the decision on growth hormone treatment. As catch-up can occur within the first 2 years, but sometimes it occurs between the age of 2 and 3, we set the age limit for catch-up at 3 years. Figure 1 is a graphical representation of the algorithm. Furthermore, it is important in the medical history to check for symptoms of emotional deprivation (psychosocial short stature) but fortunately this is a rare finding.[24-26] Obviously, a thorough physical examination should be carried out, and special attention should be given to body proportions and dysmorphic features. Abnormal body proportions are important signs of skeletal dysplasia, and dysmorphic features can direct attention to various primary growth disorders (''syndromes'').

Concern has been raised about the applicability of target height, as the height of the father is often missing. One can either ignore the height of the mother altogether and not correct for parental height, or one can assume that the father's height is the same as the mother's with a correction of 13 cm (the mean difference in adult height between men and women). It is not known which option is better, but we favor the latter. A similar approach can be taken if one of the parents is known to have a pathological growth disorder.

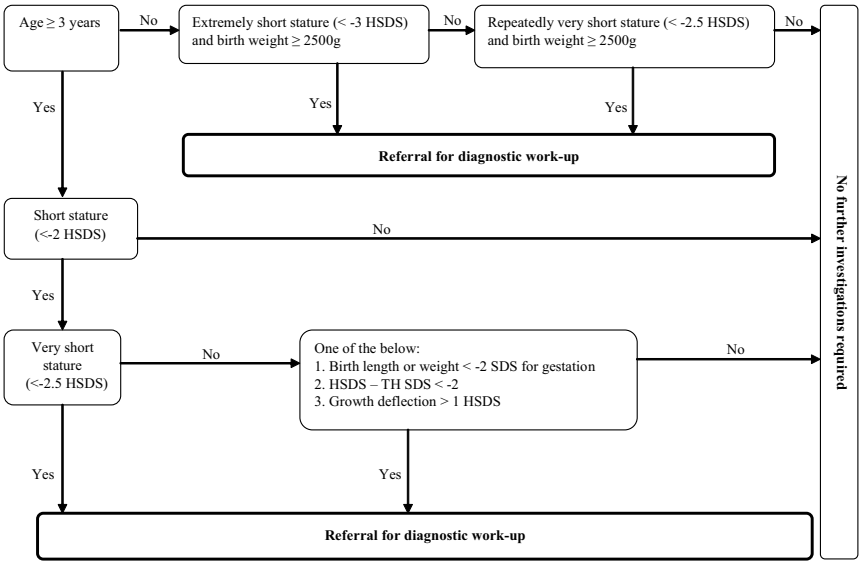


Figure 1 *Flow diagram of proposed criteria for referral of children with growth disorders. These guidelines are proposed for screening purposes only. In the case of an unusual growth pattern, certainly if associated with clinical symptoms or signs (such as disproportion and/or dysmorphic features, emotional deprivation), even if it did not comply with the rules for referral or the recommendations, doctors should still be free to follow their clinical judgment.*
HSDS, height standard deviation score; THSDS, target height standard deviation score.

The UK90 standards use an inter-centile bandwidth of 0.67 SDS instead of 0.5 or 1 SDS, so that the two lower centiles are the 0.4th and 2.3th centiles, equivalent to -2.67 and -2.0 SDS. If the 0.4th centile (-2.67 SDS) was used instead of -2.5 SDS (0.6th centile), specificity would be slightly higher and sensitivity slightly lower than calculated for a height SDS of -2.5. With respect to the "deflection", crossing an interval of 1 SD is equal to 1.5 times the interval between two reference lines on the UK charts (or 50% of the interval between the P50 and P2.3). For a more accurate estimate, the first SDS and the second SDS can be calculated and then subtracted.

In conclusion, the proposed guidelines for growth monitoring show a high sensitivity at an acceptably low false-positive rate in 3-10-year-old children. Distance to target height is the most important criterion. Below the age of 3 years, the guidelines can only detect a small percentage of pathology at an acceptably low false-positive rate, and are therefore of limited use. Besides auxological rules, clinical information taken from the medical history and physical examination can offer important guidance in taking the decision to refer patients for further tests. Finally, no algorithm can fully replace clinical judgment, and, in the case of an unusual growth pattern, even if it does not comply with the rules for referral, doctors should be encouraged to follow their clinical judgment.

Acknowledgements

This study was possible with a grant from ZonMw (No 2100.0050) and supported by an educational grant from Pfizer.

References

1. Garner P, Panpanich R, and Logan S. Is routine growth monitoring effective? A systematic review of trials. Arch Dis Child 2000;82:197-201.
2. Grote F.K., Oostdijk W, de Muinck Keizer-Schrama SMPF, et al. Growth monitoring and diagnostic work-up of short stature: an international inventorisation. J Pediatr Endocrinol Metab 2005;18:1031-8.
3. Muinck Keizer-Schrama SM. [Consensus 'diagnosis of short stature in children.' National Organization for Quality Assurance in Hospitals]. Ned Tijdschr Geneeskunde 1998;142:2519-25.
4. Hall DM. Growth monitoring. Arch Dis Child 2000;82:10-5.
5. Ahmed ML, Allen AD, Sharma A , et al. Evaluation of a district growth screening programme: the Oxford Growth Study. Arch Dis Child 1993;69:361-5.
6. Frindik JP, Kemp SF, Kearns FS , et al. Growth screening. A positive medical experience. Clin Pediatr (Phila) 1992;31:497-500.
7. Voss LD, Mulligan J, Betts PR , et al. Poor growth in school entrants as an index of organic disease: the Wessex growth study. BMJ 1992;305:1400-2.
8. Van Buuren S, Bonnemaijer-Kerckhoffs DJ, Grote FK , et al. Many referrals under Dutch short stature guidelines. Arch Dis Child 2004;89:351-2.
9. van Buuren S, van Dommelen P, Zandwijken GR , et al. Towards evidence based referral criteria for growth monitoring. Arch Dis Child 2004;89:336-41.
10. Damen GM, Boersma B, Wit JM , et al. Catch-up growth in 60 children with celiac disease. J Pediatr Gastroenterol Nutr 1994;19:394-400.
11. Boersma B, Houwen RH, Blum WF , et al. Catch-up growth and endocrine changes

in childhood celiac disease. Endocrine changes during catch-up growth. Horm Res 2002;58 Suppl 1:57-65.
12. Herngreen WP, Reerink JD, van Noord-Zaadstra B.M., et al. The SMOCC-study: Design of a representative cohort of live-born infants in the Netherlands. Eur J Public Health 1992;2:117-22
12a Grote FK, Oostdijk W, De Muinck Keizer-Schrama SM, et al. Referral patterns of children with poor growth in primary health care. BMC Public Health 2007;7:77.
13. Van den Broeck J, Hokken-Koelega A.C.S., Wit J.M. Validity of height velocity as a diagnostic criterion for idiopathic growth hormone deficiency and Turner syndrome. Horm Res 1999;51:68-73.
14. Use and interpretation of anthropometric indicators of nutritional status. WHO Working Group. Bull World Health Organ 1986;64:929-41.
15. Fredriks AM, van Buuren S, Burgmeijer RJ , et al. Continuing positive secular growth change in The Netherlands 1955-1997. Pediatr Res 2000;47:316-23.
16. Fredriks AM, van Buuren S, Wit JM , et al. Body index measurements in 1996-7 compared with 1980. Arch Dis Child 2000;82:107-12.
17. Fredriks AM, van Buuren S, Jeurissen SE , et al. Height, weight, body mass index and pubertal development reference values for children of Turkish origin in the Netherlands. Eur J Pediatr 2003;162:788-93.
18. Fredriks AM, van Buuren S, Jeurissen SE , et al. Height, weight, body mass index and pubertal development references for children of Moroccan origin in The Netherlands. Acta Paediatr 2004;93:817-24.
19. van Buuren S, Boshuizen HC, and Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med 1999;18:681-94.
20. van Buuren S, Brand JPL, Groothuis-Oudshoorn CGM , et al. Fully conditional specification in multivariate imputation. Journal of Statistical Computation and Simulation 2006;76:1049-64.
21. Karlberg J, Kjellmer I, and Kristiansson B. Linear growth in children with cystic fibrosis. I. Birth to 8 years of age. Acta Paediatr Scand 1991;80:508-14.
22. Gibson AT, Carney S, Cavazzoni E , et al. Neonatal and post-natal growth. Horm Res 2000;53 Suppl 1:42-9.
23. Hokken-Koelega AC, De Ridder MA, Lemmen RJ , et al. Children born small for gestational age: do they catch up? Pediatr Res 1995;38:267-71.
24. Doeker B, Simic S, Hauffa BP , et al. [Psychosocially stunted growth masked as growth hormone deficiency]. Klin Padiatr 1999;211:394-8.
25. Gohlke BC, Khadilkar VV, Skuse D , et al. Recognition of children with psychosocial short stature: a spectrum of presentation. J Pediatr Endocrinol Metab 1998;11:509-17.
26. Gohlke BC, Frazer FL, and Stanhope R. Growth hormone secretion and long-term growth data in children with psychosocial short stature treated by different changes in environment. J Pediatr Endocrinol Metab 2004;17:637-43.

Screening rules for growth to detect celiac disease: a case-control simulation study

7

Paula van Dommelen[1#], Floor K. Grote[2#], Wilma Oostdijk[2], Sabine M.P.F. de Muinck Keizer-Schrama[3], Bart Boersma[4], Gerard Damen[5], Cassandra G. Csizmadia[2], Paul H. Verkerk[1], Jan Maarten Wit[2] and Stef van Buuren[1,6]

[1]TNO Quality of life, Leiden, The Netherlands

[2]Leiden University Medical Center, Leiden, The Netherlands

[3]Erasmus MC - Sophia Children's Hospital, Rotterdam, The Netherlands

[4]Medical Center Alkmaar, Alkmaar, The Netherlands

[5]University Medical Center St Radboud, Nijmegen, The Netherlands

[6]University of Utrecht, Utrecht, The Netherlands

[#]These authors contributed equally to this work

BMC Pediatr 2008;8:35.

Abstract

Background: It is generally assumed that most patients with celiac disease (CD) have a slowed growth in terms of length (or height) and weight. However, the effectiveness of slowed growth as a tool for identifying children with CD is unknown. Our aim is to study the diagnostic efficiency of several growth criteria used to detect CD children.
Methods: A case-control simulation study was carried out. Longitudinal length and weight measurements from birth to 2.5 years of age were used from three groups of CD patients (n=134) (one group diagnosed by screening, two groups with clinical manifestations), and a reference group obtained from the Social Medical Survey of Children Attending Child Health Clinics (SMOCC) cohort (n=2,151) in The Netherlands. The main outcome measures were sensitivity, specificity and positive predictive value (PPV) for each criterion.
Results: Body mass index (BMI) performed best for the groups with clinical manifestations. Thirty percent of the CD children with clinical manifestations and two percent of the reference children had a BMI Standard Deviation Score (SDS) less than -1.5 and a decrease in BMI SDS of at least -2.5 (PPV=0.85%). The growth criteria did not discriminate between the screened CD group and the reference group.
Conclusions: For the CD children with clinical manifestations, the most sensitive growth parameter is a decrease in BMI SDS. BMI is a better predictor than weight, and much better than length or height. Toddlers with CD detected by screening grow normally at this stage of the disease.

Background

One of the goals of growth monitoring in developed countries is the detection of undiagnosed illnesses. Nevertheless, there is little consensus on which referral criteria for children with growth retardation are appropriate.[1] Recently we reported on the predictive value of various growth criteria for the detection of Turner's syndrome.[2] The focus of that study was on short stature and slowed growth for length or height, as short stature is the main common physical characteristic of Turner's syndrome. Growth retardation, however, may also imply failure to thrive in terms of slowed growth for weight and BMI.

Celiac disease (CD), also known as gluten-sensitive enteropathy, is characterized by subtotal villous atrophy of the small intestine, intra-epithelial lymphocytosis and crypt hyperplasia, and is associated with a variable mode of presentation. The classical presentation is characterized by failure to thrive, diarrhoea, irritability, vomiting, anorexia, foul stools, abdominal distension and muscle wasting. However, many infants, toddlers and children with celiac disease present with few or no signs and symptoms.[3-6]

The prevalence of the classical presentation of CD decreased in the past decade, while the prevalence of non-classical presentations increased.[3-5] Growth failure in terms of length (or height) or weight may be the earliest sign of the disease.[7] In 1994, the reported incidence of clinically diagnosed CD in the Netherlands was 0.54 per 1000 live births.[8] However, screening studies using detection of anti-endomysium antibodies have shown a much higher prevalence (1:300 to 1:100). The ratio of clinically diagnosed versus CD detected by screening varies between 1:7 and 1:14.[9] Early detection and treatment with a gluten-free diet is required to improve the immediate quality of life of the CD patients and to decrease the long-term risks, including reduction in adult height, a higher prevalence of malignancies, adverse pregnancy outcome, neurological problems and osteomalacia.[10]

Mass screening for CD using specific antibodies is unlikely to be performed, because of the uncertainty concerning the cost-benefit ratio. As there is a high incidence of CD (1.7 to 8.3%) in children with growth retardation without gastrointestinal symptoms and even higher (up to 59.1%) when other (endocrine) causes for short stature are excluded[7], a substantial proportion of infants and children with CD may be detected through growth monitoring.

In the Netherlands, nearly every child is monitored for height and weight from birth till the age of 16-18 years. Children with abnormal growth are referred to secondary health care providers according to certain criteria.[11] As no pathologic causes for short stature were detected in most of these referred children, we recently revised the referral criteria. These criteria minimize the unnecessary referrals and are aimed at not missing important diseases such as CD, Turner's syndrome and endocrine abnormalities.[12]

So far, it is generally assumed that most CD patients have a slowed growth in terms of length (or height) and weight.[13] However, the effectiveness of slowed growth as a tool for identifying children with CD is unknown. The aim of this study is to establish optimal referral criteria based on abnormal growth for detecting asymptomatic and symptomatic children with CD.

Methods

Patients

Longitudinal length and weight data of patients with CD were collected from three different studies. The first study was a prospective screening study using blood tests in unrecognized CD in children aged 2-4 years, visiting the Community Child Health Care Centers in the Dutch province of Zuid (South)-Holland.[9] In this study, 32 children with CD were detected between May 1997 and June 1998. The second study was a retrospective study on catch up growth in patients with CD.[13] A written questionnaire

including their symptomatology, duration of complaints before diagnosis, age at diagnosis, associated diseases in the past and parental heights was sent to all members of the Dutch Celiac Society in the early nineteen eighties. Growth data were collected from 74 children younger than 16 years. The third study was a prospective study on catch up growth.[14] All newly diagnosed childhood CD patients from two separate pediatric departments were included between April 1994 and September 1995 (n=28). The children in the second and third study presented with a full range of classical symptoms. We used all growth data before and at the start of the gluten-free diet, till the age of 2.5 years. The data was gathered retrospectively from child welfare clinics, pediatricians and general practitioners. Additional growth information of these children was obtained from physicians in the Regional Child Health Care Centres. The diagnosis of CD was confirmed by histology for all patients, although in the retrospective study we were dependent on the information provided by patient reports. In total, we included 134 children: 32 children from the first study, 74 children from the second study and 28 children from the third study. Exclusion criteria were: an unknown date of starting the gluten-free diet and no measurement between birth and 2.5 years of age. After excluding such cases, 122 children were eligible for further analyses: 26 children from the first study and 96 children from the second and third study (see Figure 1).

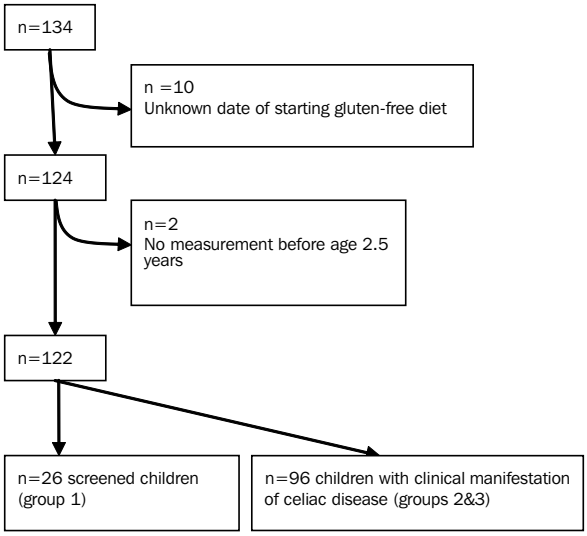


Figure 1 *Flow chart of children with CD used in the study.*

The first CD group was asymptomatic or featured symptoms that were not signalled by the parents or the general practitioners. Therefore, this group was analyzed separately (screened group). The second and third CD groups were clinically diagnosed and we reasoned that these two groups could be pooled (symptomatic group).

Reference sample

A reference sample was obtained from the Social Medical Survey of Children Attending Child Health Clinics (SMOCC) cohort, a nationally representative cohort of 2,151 children born in the Netherlands during 1988-1989.[15] Of this cohort longitudinal data of length and weight of children from birth to 2.5 years of age were available. The length and weight from birth to two years of these children were previously described by Herngreen et al.[16]

Power analysis

For an estimated sensitivity of 50% we obtained a 95% confidence interval (95%-C.I.) of +/- 19% with the 26 screened CD children and +/- 10% with the 96 symptomatic CD children. For an estimated specificity of 98% we obtained a 95%-C.I. of +/-0.6% with the 2,151 reference children.

Screening rules

We formulated several screening rules for growth that could serve as criteria for referral to specialist care (Table 1). Each of these rules combines several parameters, such as starting age or a decrease in standard deviation score (SDS) over a certain time period. Table 1 explains the interpretation of each parameter. We used several simulation values for each parameter to see how the diagnostic performance of each rule changes. For example, for the parameter starting age, we used simulation values of 0, ½ and 1 year of age. These simulation values were chosen to investigate if the growth pattern of CD children starts to deviate from the reference population already at birth (0 year), or at the time children commence to eat gluten (½ year) or later (at 1 year).

Table 1 *Growth screening rules with their definitions, interpretation of the used parameters and cut off (simulation) values (see method for details).*

Screening rule	Definition	Parameter	Interpretation	Simulation values
Delta rule*^	For ages e_1 to 2.5 years, refer if	e_1	Age (in years) after which the rule is effective	0, 0.5, 1
	$(SDS_2 - SDS_1) < g_1$	g_1	Change in SDS	-0.5,-1,-1.5,-2,-2.5,-3
Extended delta rule*	For ages e_2 to 2.5 years, refer if	e_2	Age (in years) after which the rule is effective	0, 0.5, 1
	$SDS_2 < f_1$, AND	f_1	SDS cut off level below which the SDS_2 must lie	-1,-1.3, -1.5,-2, -2.5
	$(SDS_2 - SDS_1) < g_2$	g_2	Change in SDS	-0.5,-1,-1.5,-2,-2.5,-3
Slowed growth *	For ages e_3 to 2.5 years, AND	e_3	Age (in years) after which the rule is effective	0, 0.5, 1
	$X_2 - X_1 \geq 3/12$ refer if		Minimal three months interval between ages X_1 and X_2	
	$SDS_2 < f_2$, AND	f_2	SDS cut off level below which the SDS_2 must lie	-1,-1.3, -1.5,-2, -2.5
	$(SDS_2 - SDS_1)/(X_2-X_1) < g_3$	g_3	Change in SDS per year	-0.5,-1,-1.5,-2,-2.5
Conditional weight gain rule	For ages e_4 to 2.5 years, refer if	e_4	Age (in years) after which the rule is effective	0, 0.5, 1
	weight $SDS_2 < f_3$ AND	f_3	SDS cut off level below which SDS_2 must lie	-1,-1.3, -1.5,-2, -2.5
	weight SDSgain= (weight SDS_2 - r weight SDS_1)/ ($\sqrt{1-r^2}$) < g_4	g_4	Change in SDS	-0.5,-1,-1.5,-2,-2.5
Absolute SDS rule*	For ages 0 to e_5 years, refer if	e_5	Age (in years) at which the referral level changes	0, 0.5, 1
	$SDS < f_4$	f_4	SDS cut off level before age e_5	-1, -1.3, -1.5, -2, -2.5, -3, -3.5
	For ages e_5 to 2.5 years, refer if $SDS < f_5$	f_5	SDS cut off level after age e_5	-1, -1.3, -1.5, -2, -2.5, -3
Parental height corrected rule	For ages e_6 to 2.5 years, refer if	e_6	Age (in years) after which the rule is effective	0, 0.5, 1
	length $SDS < f_6$, AND	f_6	Length SDS must lie below this cut off level	-1, -1.3, -1.5, -2, -2.5
	length SDS - TH SDS < g_5	g_5	Difference between length SDS and target height (TH) SDS	-1, -1.3, -1.5, -2, -2.5
Parental height deflection rule	For ages e_7 to 2.5 years, refer if	e_7	Age (in years) after which the rule is effective	0, 0.5, 1
	(length SDS_2 - length SDS_1) < g_6, AND | length SDS_2 - TH SDS | > | length SDS_1 - TH SDS |	g_6	Change in length SDS Length SDS at age X_1 is closer to it's target height than length SDS at age X_2	-0.5,-1,-1.5,-2,-2.5,-3
Combined weight and length deflection rule	For ages e_8 to 2.5 years, AND	e_8	Age (in years) after which the rule is effective	0, 0.5, 1
	(weight SDS_2 - weight SDS_1)< g_7, AND	g_7	Weight change in SDS	-0.25,-0.5,-1,-1.5,-2
	(length SDS_2 - length SDS_1) < g_8, AND $Y_1 > X_1$	g_8	Length change in SDS Starting point length deflection (Y_1) after starting point weight deflection (X_1)	-0.25,-0.5,-1,-1.5,-2

Several screening rules for growth were studied. Each screening rule consists of parameters that we have varied. For more details, see the paragraph screening rules. *Calculated for length (height), weight, and BMI ^For example, if e_1=0.5 year and g_7=-2 weight SDS, then a child is referred if the second weight SDS measurement is -2 below the first weight SDS measurement and both weights were measured after six months of age (or at six months of age for the first measurement).

In total, we formulated eight rules, and each rule is explained in detail in Table 1 and below.

1. The first rule (*delta rule*) refers a child if an absolute change in length SDS, weight SDS or BMI SDS occurs. For example, suppose a child has two weight measurements, one measurement at the age of six months and one at the age of 1.5 years. This child will then be referred according to the delta rule with parameters e_1=0.5 and g_1=-2 (see Table 1) if his or her weight decreases by more than 2 SDS between the first and the second measurement.
2. The second rule (*extended delta rule*) is equal to the first rule with the extension that the second measurement has to have a low SDS (for example less than -1.5 SDS).
3. The third rule (*slowed growth rule*) signals whether an abnormal slowed growth for length, weight or BMI occurs in terms of change in SDS per year in combination with a current low SDS. For example, suppose a child has two length measurements, one measurement at the age of seven months and one measurement six months later. This child will then be referred according to the slowed growth rule with parameters e_3=0.5, g_3=-1 and f_2=-1.5 (see Table 1) if the difference between the second and first length measurement per year exceeds 1 SDS (which corresponds to a decrease of 0.5 SDS within six months) and if the second measurement is less than -1.5 SDS. We prefer the term *slowed* growth over the term *velocity* to indicate the decrease in growth in SDS per year. The term velocity commonly refers to cm or kg/year.
4. The fourth rule (*conditional weight gain rule)* is the conditional weight gain rule that signals whether a child's conditional weight gain SDS is less than a certain value[17,18] with the restriction of having a low weight SDS.
5. The fifth rule (*absolute SDS rule)* refers a child if the length SDS, weight SDS or BMI SDS is low. An example is to refer if a child's length SDS is less than -2 (e_5 =0 and f_5 =-2).
6. We also considered rules that take genetic height potential into account. The sixth rule (*parental height corrected rule*) compares the height SDS of the child to its target height SDS in combination with a low height SDS.
7. The seventh rule (*parental height deflection rule*) signals whether a slowed growth for length SDS of the child moves away from the child's target height. This rule was added because of the assumption that a correction might be needed for parental height in the first years of life: e.g. a baby that is born with a length SDS of -1 and has a target height SDS of +2, would be expected to cross the SD lines in upward direction in the first 2-3 years. A growth disorder could disturb this, and a stable length SDS of this child at -1 over the first 2 years could indicate growth pathology such as CD.

8. Similarly, in the eight rule (*combined weight and length deflection rule)* we combined weight and length, in which a slowed growth for length occurs after a slowed growth for weight.

Several cut off values for age were used as the effectiveness of these rules may increase by examining higher age groups. Slowed growth requires measurements taken at least three months apart. We chose this short time interval to facilitate early detection, taking into consideration that children in the first year of life grow faster than in later years.

It should be noted that some parameters select a subset of the data and assume multiple measurements. The rules were only tested on children that complied with these assumptions. All available pairs of measurements for each infant were used.

Statistical analysis

Each screening rule was implemented using S-Plus version 7.0.3 for Microsoft Windows (2005), and was applied to the longitudinal data of children. We calculated sensitivity, specificity and positive predictive value (PPV) for each rule with several scenarios (simulation values). The rules were ordered according to their sensitivity at high levels of specificity. A higher sensitivity at the same level of specificity, results in a better performance. The results were plotted as a Receiver Operating Characteristic (ROC) curve, but scaled to a different axis than conventionally in order to view the area of most interest (high specificity). Each point in the ROC curve is the false-positive rate against sensitivity of a scenario (combination of simulation values) of a rule. Scenarios of rules with approximately 2% false-positive rates were presented in detail as we assumed that a false-positive rate greater than 2% would result in too many referrals. PPV was calculated assuming that the incidence of CD is 0.54 per 1000 live births in the Caucasian population.[9] Sensitivity analyses were performed to calculate the effect of small variations (0.1-1.0/1000) in the incidence of CD on PPV.

Length, weight and BMI were expressed as SDS, using the Dutch reference growth data.[19,20] In preterm infants (gestational age < 37 weeks) length and weight SDS were corrected for gestational age. The intrauterine growth charts from the Swedish reference population was used to express SDS up to the age corresponding with 40 weeks of gestation.[21] Between 40 and 42 weeks an interpolation between the growth curve of the Swedish reference population and that of the Dutch reference population was used. From 42 weeks of gestation till the age of 2 years, SDS was calculated on ages corrected for gestational age, using the Dutch reference growth data.

We assumed that a child was referred if the growth pattern met the criteria of a given screening rule for the first time. All rules were dealt with separately, meaning that the same child could be referred according to each separate rule.

Results

Table 2 contains general characteristics of the symptomatic CD group and the screened CD group. In the symptomatic group, mean weight SDS was compromised most, followed by mean BMI SDS.

Table 2 *General characteristics of the CD-population.*

Characteristics		Screened (n=26)	Symptomatic (n=96)
Gender (M)		50%	35%
Ethnicity	Dutch	92%	98%
	Others	8%	2%
Median (range) age in years at start diet		3.96 (2.94-6.06)	1.43 (0.41-20.7)
Mean (SD) length SDS *∞		-0.26 (0.98)	-0.89 (1.30)
Mean (SD) weight SDS*∞		-0.06 (0.81)	-1.54 (1.15)
Mean (SD) BMI SDS*∞		0.28 (0.57)	-1.28 (1.15)
Mean (SD) target height SDS		0.41 (0.92)	0.00 (0.75)

*For the children in the screened group figures at diagnosis are given (also when diagnosis is after 2.5 years of age). For the symptomatic CD children figures at the start of the gluten-free diet are given. ∞Based on children with at least one measurement between 6 months before and 3 months after gluten-free diet or diagnosis.

Diagnostic performance of the rules: screened CD children

All screening rules detected less than 5% of the screened CD children at a 2% false-positive rate. Therefore, none of the rules were able to discriminate between the CD children detected by screening and the reference children. This indicates that the screened and the reference children hardly differ in terms of their growth pattern.

Diagnostic performance of the rules: symptomatic CD children

The results are different for the symptomatic CD children. Figure 2 shows the ROC plot for the four best screening rules for the symptomatic CD group. Only scenarios with a false-positive rate of less than 10% are plotted. The line for which sensitivity is equal to 100-specificity is given in the figure. Scenarios on this line are not able to discriminate between the CD and the reference group. The BMI extended delta rule had the highest sensitivities at low false-positive rates. A strict version of this rule is a decrease in BMI SDS of -3 and a BMI SDS less than -1 between birth and 2.5 years of age. This scenario correctly identified 21% (95%-CI 12-30) of the CD children and 99% (95%-CI 98.6-99.4) of the reference children were correctly labeled as disease free. The PPV of this scenario is approximately 1%. For example, suppose a boy has a BMI on the line above the median of the growth chart (SDS=+1) at three months of age. If he crosses three SDS lines (SDS 0, -1 and -2) before the age of 2.5 years, then the boy has a 1% probability of having CD.

A less strict version of the BMI extended delta rule is a decrease in BMI SDS of at least -2 and a BMI SDS of less than -1.5 between birth and 2.5 years of age, with a sensitivity of 38% (95%-CI 27-49), a false-positive rate of 3.4% (95%-CI 2.6-4.2) and a PPV of 0.60%.

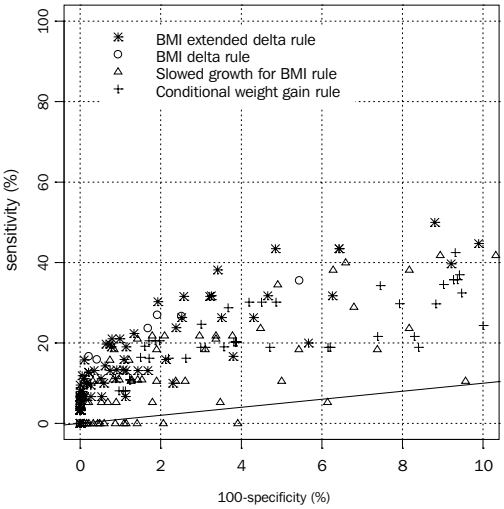


Figure 2 *ROC plots of effective growth screening rules for detecting CD in the symptomatic group. The rules are an absolute change in BMI SDS with or without the restriction of a low BMI SDS, a slowed growth for BMI, and a conditional weight gain in combination with a low weight SDS.*

Table 3 *Properties of the best scenarios with approximately 2% false-positives (=98% specificity).*

(symptomatic) CD	Simulation values			Sensitivity	100-Specificity	PPV
Slowed growth for BMI rule	e_1=0.5	f_1=-1.3	g_1=-2.5	33.9%	2.1%	0.86%
BMI delta rule	e_2=0.5	w_1=-1.5		27.0%	1.9%	0.76%
Conditional weight gain rule	e_3=0.5	f_2=-2.5	w_2=-0.5 to -1.5	20.5%	1.9%	0.58%
Slowed growth for weight rule	e_1=0.5	f_1=-2.5	g_1=-0.5 to -1	19.7%	1.9%	0.56%

The properties of the four best rules for the symptomatic CD group, in terms of sensitivity and PPV at approximately 98% specificity, are presented in table 3. Thirty percent (95%-CI 20-40) of the CD children and 1.9% (95%-CI 1.3-2.5) of the reference children had a decrease in BMI SDS of at least -2.5 and a BMI SDS less than -1.5 between birth and 2.5 years of age. In children with such decrease in BMI SDS, the probability of CD is 0.85%. PPV varied between 0.16% and 1.57% when changing the incidence of CD from 1:10000 to 1:1000 live births. For example, a girl has a BMI on the median of the growth chart at one month of age, and her BMI crosses centiles for a certain time period until she reaches a BMI SDS of less than -2.5. Then this girl will be referred according

to the scenario above. Her probability of actually having CD is 0.85%. Furthermore, 27% (95%-CI 16-38) of the CD children versus 1.9% (95%-CI 1.3-2.5) of the reference children had a decrease in BMI SDS of at least -1.5 when they were older than six months of age. The probability of having CD when a child complies with this rule is 0.76%. Both the slowed growth for BMI rule and the conditional weight gain rule result in a sensitivity of approximately 22% (95%-CI 11-33) at a false-positive rate of 1.9% (95%-CI 1.3-2.5). The PPV is approximately 0.6%. The sensitivity between the first rule (the BMI extended delta rule) and the fourth rule (conditional weight gain rule) differed most. Twenty percent of the CD children complied with the first rule (true-positive) and not with the fourth rule (false-negative), while ten percent of the CD children complied with the fourth rule and not with the first rule. If we combine both rules, sensitivity is 41%. However, the false-positive rate also increased to 3.6%.

The delta rules for length and weight, the slowed growth rule for length and weight, the absolute SDS rule, rules that take genetic height potential into account (parental height corrected rule and parental height deflection rule) and the combined weight and length deflection rule proved less effective (data not shown). At a fixed specificity of 98%, sensitivities for these rules were less than 20%.

Discussion

Our study shows that for detecting or predicting symptomatic CD children by growth, a decrease in BMI is more informative than a decrease in weight or length. The screened CD children grow normally between birth and 2.5 years of age.

The optimal weight rule in this study was the conditional weight gain rule. The conditional weight gain rule corrects for regression to the mean. The amount of regression to the mean depends on the correlation of body weight across age.[17] The correlations that we used in our study were based on children in the UK.[18] The conditional weight gain rule may perform better when using correlations of Dutch children. However, these correlations are presently not available. To validate the UK correlations for the Dutch children, we calculated if the SDSgain has a mean of zero and a SD of 1, and if it is uncorrelated with the first weight SDS. For the reference group of Dutch children, the mean (SD) SDSgain is -0.06 (1.41) and its correlation with the first weight SDS is -0.23. As both SD and correlation are quite high, the conditional weight gain rule may perform better when using Dutch correlations of weights. Furthermore, a rule based on BMI that corrects for regression to the mean may improve discrimination between the symptomatic CD group and the reference group. So far no suitable correlations have been published to calculate this conditional gain.

In this study, PPV of the screening rules may be underestimated for several reasons. Firstly, PPV will be slightly higher as there will be one case of CD in our reference group if we assume that the incidence of CD is 0.54 per 1000 life births. Secondly, PPV was calculated using only the incidence of CD. However, if we keep in mind that children with genetic disorders or diseases other than CD may be detected by some of our rules for failure to thrive, PPV may be higher. For example, if we assume that sensitivity and specificity for the most optimal rule for CD in this study is similar to patients with Cystic Fibrosis (CF), then PPV will be higher if this is based on the incidence of both CD and CF.[22]

As Csizmadia et al. reported earlier, the children with CD detected by screening had a normal weight and length at time of diagnosis.[9] We have confirmed that all children in this group indeed had a normal growth pattern between birth and 2.5 years of age. This corresponds with the asymptomatic character of this silent form of CD. Thus monitoring growth would not seem to be useful for the detection of silent CD at this specific stage of the disease. The prevalence of children with short stature and no gastrointestinal symptoms investigated for CD is 2-8%[7] compared to a prevalence of 1:300 to 1:100 in the general population. Therefore, one may expect that these children would develop abnormal growth after several years.

CD is often atypical or clinically silent, which results in many undiagnosed children. However, since the widespread introduction of serologic testing and the increased awareness of CD in the late 1990s there has been an increase in incidence as well as a change in clinical presentation.[3-6] The classical symptoms, such as malabsorption and poor weight gain no longer dominate the clinical picture. Instead, there is an increase of cases with non-classical symptoms, including unusual intestinal complaints or extra-intestinal symptoms (e.g. short stature) involving older children. In addition, the age of presentation may be changing due to differences in infant feeding practices, duration of breastfeeding and improved recognition of potential CD by general practitioners. As our non-screened population was diagnosed before 1995, we were not able to study the effect of this change in time on the performance of the growth criteria. However, one may assume that for the age group included in our study, the performance of the growth criteria is similar for the present CD-population, as it is mainly the delayed onset variant of the disease (the non-classical form) that has increased during the recent years, suggesting that the growth impairment becomes apparent much later.

Most of the patients in our study were females, as was reported in several other studies.[23] Bardella et al. hypothesized that males escape diagnosis, but that the two sexes are equally affected. This hypothesis is supported by the absence of differences in sex in the screening study (see Table 2).

In conclusion, we support the clinician to consider testing for CD in a diagnostic work-up in young children with failure to thrive. The most sensitive growth parameter is BMI SDS.

We recommend further research with a large sample of children with CD diagnosed in the last few years to study the most valid simulation values for referral rules based on BMI and other diagnostics.

Conclusions

BMI is more efficacious than weight, and much more than length or height, in detecting symptomatic children with CD. Toddlers with CD detected by screening grow normally at this stage of the disease.

Acknowledgements

We acknowledge Jeroen van Rijn for his contribution to the study, Jason Maassen and Lynne Ball for their useful suggestions for improvement of the text and Pfizer for their educational grant.

References

1. Grote FK, Oostdijk W, de Muinck Keizer-Schrama SMPF, Dekker FW, Verkerk PH, Wit JM. Growth monitoring and diagnostic work-up of short stature: an international inventorisation. J Pediatr Endocrinol Metab 2005;18:1031-8.
2. Buuren van S, Dommelen van P, Zandwijken GR, Grote FK, Wit JM, Verkerk PH. Towards evidence based referral criteria for growth monitoring. Arch Dis Child 2004;89:336-41.
3. Garampazzi A, Rapa A, Mura S, Capelli A, Valori A, Boldorini R, Oderda G. Clinical pattern of celiac disease is still changing. J Ped Gastroenterol Nutr 2007;45:611-4.
4. Steens RF, Csizmadia CG, George EK, Ninaber MK, Hira Sing RA, Mearin ML. A national prospective study on childhood celiac disease in the Netherlands 1993-2000: an increasing recognition and a changing clinical picture. J Pediatr 2005;147:239-43.
5. Fasano A. Clinical presentation of celiac disease in the pediatric population. Gastroenterology 2005;128(Suppl 1):S68-73.
6. Bottaro G, Cataldo F, Rotolo N, Spina M, Corazza GR. The clinical pattern of subclinical/silent celiac disease: an analysis on 1026 consecutive cases. Am J Gastroenterol 1999;94:691-6.
7. Rijn van JC, Grote FK, Oostdijk W, Wit JM. Short stature and the probability of coeliac disease, in the absence of gastrointestinal symptoms. Arch Dis Child 2004;89:882-3.
8. George EK, Mearin ML, Franken HC, Houwen RH, Hirasing RA, Vandenbroucke JP. Twenty years of childhood coeliac disease in The Netherlands: a rapidly increasing incidence? Gut 1997;40:61-6.

9. Csizmadia CG, Mearin ML, von Blomberg BM, Brand R, Verloove-Vanhorick SP. An iceberg of childhood coeliac disease in the Netherlands. Lancet 1999;353:813-4.
10. Goddard CJ, Gillett HR. Complications of coeliac disease: are all patients at risk? Postgrad Med J 2006;82:705-12.
11. Muinck Keizer-Schrama SM. [Consensus 'diagnosis of short stature in children.' National Organization for Quality Assurance in Hospitals]. Ned Tijdschr Geneeskd 1998;142:2519-25.
12. Grote FK, van Dommelen P, Oostdijk W, de Muinck Keizer-Schrama SM, Verkerk PH, Wit JM, van Buuren S. Developing evidence-based guidelines for referral for short stature. Arch Dis Child 2008;93:212-7.
13. Damen GM, Boersma B, Wit JM, Heymans HS. Catch-up growth in 60 children with celiac disease. J Pediatr Gastroenterol Nutr 1994;19:394-400.
14. Boersma B, Houwen RH, Blum WF, van Doorn J, Wit JM. Catch-up growth and endocrine changes in childhood celiac disease. Endocrine changes during catch-up growth. Horm Res 2002;58 Suppl 1:57-65.
15. Herngreen WP, Reerink JD, van Noord-Zaadstra B.M., Verloove-Vanhorick SP, Ruys JH. The SMOCC-study: Design of a representative cohort of live-born infants in the Netherlands. Eur J Public Health 1992;2:117-22.
16. Herngreen WP, van Buuren S, van Wieringen JC, Reerink JD, Verloove-Vanhorick SP, Ruys JH. Growth in length and weight from birth to 2 years of a representative sample of Netherlands children (born in 1988-89) related to socioeconomic status and other background characteristics. Ann Hum Biol 1994;21:449-63.
17. Cole TJ. Conditional reference charts to assess weight gain in British infants. Arch Dis Child 1995;73:8-16.
18. Cole TJ. Presenting information on growth distance and conditional velocity in one chart: practical issues of chart design. Stat med 1998;17:2697-2707.
19. Fredriks AM, van Buuren S, Burgmeijer RJ, Meulmeester JF, Beuker RJ, Brugman E Roede MJ, Verloove-Vanhorick SP, Wit JM. Continuing positive secular growth change in The Netherlands 1955-1997. Pediatr Res 2000;47:316-23.
20. Fredriks AM, van Buuren S, Wit JM, Verloove-Vanhorick SP. Body index measurements in 1996-7 compared with 1980. Arch Dis Child 2000;82:107-12.
21. Niklasson A, Ericson A, Fryer JG, Karlberg J, Lawrence C, Karlberg P. An update of the Swedish reference standards for weight, length and head circumference at birth for given gestational age (1977-1981). Acta Paediatr Scand 1991;80:756-62.
22. Bobadilla JL, Macek M, Jr., Fine JP, Farrell PM. Cystic fibrosis: a worldwide analysis of CFTR mutations--correlation with incidence data and application to screening. Hum Mutat 2002;19:575-606.
23. Bardella MT, Fredella C, Saladino V, Trovato C, Cesana BM, Quatrini M, Prampolini L. Gluten intolerance: gender- and age-related differences in symptoms. Scand J Gastroenterol 2005;40:15-9.

Growth monitoring to detect cystic fibrosis

8

Paula van Dommelen[1#], Floor K. Grote[2#], Wilma Oostdijk[2],
Sabine M.P.F. de Muinck Keizer- Schrama[3], Jan Bouquet[3], Johannes J.E. Hendriks[4],
Jan Kouwenberg[5], Paul H. Verkerk[1], Stef van Buuren[1,6] and Jan Maarten Wit[2]

[1]TNO Quality of life, Leiden, The Netherlands
[2]Leiden University Medical Center, Leiden, The Netherlands
[3]Erasmus MC - Sophia Children's Hospital, Rotterdam, The Netherlands
[4]University Hospital Maastricht, Maastricht, The Netherlands
[5]Hagahospital - Juliana Children's Hospital, The Hague, The Netherlands
[6]University of Utrecht, Utrecht, The Netherlands
[#]These authors contributed equally to this work

Submitted to Hormone Research

Abstract

Background/Aims: Cystic fibrosis (CF) in infancy and childhood is often associated with failure to thrive (FTT). This would suggest that in countries without a newborn screening program for CF, FTT could be used as a clinical screening tool. The aim of this study is to assess the diagnostic performance of FTT for identifying children with CF.
Methods: Longitudinal length and weight measurements up to 2.5 years of age were used from CF patients (n=123) and a reference group (n=2,151) in the Netherlands. Growth measurements after diagnosis were excluded. We developed five potential screening rules based upon length, weight and body mass index (BMI) standardized by age and sex (SDS). Outcome measures were sensitivity, specificity and positive predictive value (PPV).
Results: BMI SDS had the highest sensitivity at low false-positive rates. An efficient scenario is a BMI SDS below -2.5 SD in combination with a decrease in BMI SDS of at least 0.5 SD. This scenario had a sensitivity of 32%, a specificity of 98.3% and a PPV of 0.75%.
Conclusion: In the absence of a newborn screening program, young children with FTT for BMI are candidates to consider testing for CF.

Introduction

Cystic fibrosis (CF) is the most common life-threatening autosomal recessive disease in the Caucasian population.[1] In Caucasian European newborns the incidence is about 1:2,500 whereas in Caucasian North American newborns the incidence is approximately 1:3,500.[1,2] The symptoms of CF usually start at an early age and include meconium ileus, recurrent respiratory symptoms (cough, wheeze, pneumonia), steatorrhoea, diarrhea, abdominal distension and failure to thrive (FTT) (slowed growth).[3,4] In several countries newborn screening programs for CF have been introduced. In areas where there is no such program, CF is often diagnosed late, because the presentation of the symptoms is variable.[5-7] Diagnostic delay can lead to malnutrition[8], deterioration in lung function, an increase in immunoglobulin levels and a reduced life expectancy.[9-12] The standard diagnostic strategy is a sweat test after recognition of symptoms or a positive family history, followed by further laboratory testing and DNA analysis.[13]

Several studies have compared the growth pattern of CF-patients with that of healthy children.[14-20] Many cases show FTT for weight, length and body mass index (BMI). At the age of 1 year, mean weight and length standardized by age and sex (SDS) generally do not exceed -1.3 SD and -1 SD respectively.[14-18] It is suggested that FTT for weight[15], length[15] and BMI[17] is more severe in girls than in boys. Mean BMI SDS for girls was approximately -1.2 SD at the age of 1 year, while this was -0.8 SD for boys.[17] A decrease

in weight corrected for height was most pronounced in children with predominantly pulmonary symptoms[19] and height and weight of CF-patients who were not colonized with Pseudomonas aeruginos were within normal limits.[20] Most children experienced catch-up growth after diagnosis.

The main goal of growth monitoring in developed countries is the detection of undiagnosed illnesses. Although most patients with CF have FTT for weight, length and BMI, it is unknown how effective FTT is as a screening tool for CF. The aim of this study is to assess the diagnostic performance of growth-based criteria for detecting CF.

Materials and methods

Patients

Longitudinal length and weight data of the patients with CF were collected retrospectively in the year 2005 from three major CF clinics in The Netherlands: Erasmus MC - Sophia Children's Hospital in Rotterdam, University Hospital Maastricht and Haga/Juliana Children's Hospital in The Hague. Additional growth information of these children was obtained from physicians in the Regional Child Health Care Centres with permission from the patient or his or her parents. The following information was obtained from the patient files: date of birth, date of referral, sex, ethnicity, perinatal information (birth weight, length, gestational age), date of diagnosis of CF and DNA-mutation. If ethnicity was not recorded, it was assessed based on the patient's first and family name according to an algorithm reported earlier.[21] We included only growth data before or at diagnosis, with a maximum at the age of 2.5 years. In total, 123 children were available.

Reference sample

A reference sample was obtained from the Social Medical Survey of Children Attending Child Health Clinics (SMOCC) cohort, a nationally representative cohort of 2,151 children born in the Netherlands during 1988-1989.[22] Of this cohort longitudinal data of length and weight of children from birth to 2.5 years of age were available. The length and weight distributions from birth to two years of these children were previously described by Herngreen et al.[23]

Screening rules

We developed several screening rules based upon growth (Table 1). The rules are meant to serve as criteria for referral to specialist care. The same rules have also been examined for the detection of children with celiac disease.[24] Each of these rules combines several parameters, such as a certain amount of a decrease in SDS over some time period.
Table 1 provides the explanation of the individual parameter in the rules studied.

Table 1 *Growth screening rules with their definitions, interpretation of the used parameters and cut off (simulation) values (see methods section for details).*

Screening rule	Definition	Parameter	Interpretation	Simulation values
Delta rule*	For ages 0 to 2.5 years, refer if			
	$(SDS_2 - SDS_1) < g_1$	g_1	Change in SDS	-0.5,-1,-1.5,-2,-2.5,-3
Extended delta rule*	For ages 0 to 2.5 years, refer if			
	$SDS_2 < f_1$, AND	f_1	SDS cut off level below which the SDS_2 must lie	-1,-1.3, -1.5,-2, -2.5
	$(SDS_2 - SDS_1) < g_2$	g_2	Change in SDS	-0.5,-1,-1.5,-2,-2.5,-3
Slowed growth *	For ages 0 to 2.5 years, AND			
	$X_2 - X_1 \geq 3/12$ refer if		Minimal three months interval between ages X_1 and X_2	
	$SDS_2 < f_2$, AND	f_2	SDS cut off level below which the SDS_2 must lie	-1,-1.3, -1.5,-2, -2.5
	$(SDS_2 - SDS_1)/(X_2-X_1) < g_3$	g_3	Change in SDS per year	-0.5,-1,-1.5,-2,-2.5
Conditional	For ages 0 to 2.5 years, refer if			
weight gain rule	Weight $SDS_2 < f_3$ AND	f_3	SDS cut off level below which the SDS_2 must lie	-1,-1.3, -1.5,-2, -2.5
	Weight SDSgain=	g_4	Change in SDS	-0.5,-1,-1.5,-2,-2.5
	$(\text{weight } SDS_2 - r \text{ weight } SDS_1)/(\sqrt{1-r^2}) < g_4$			
Absolute SDS rule*	For ages 0 to e_1 years, refer if	e_1	Age (in years) at which the referral level changes	0, 0.5, 1
	$SDS < f_4$	f_4	SDS cut off level before age e_1	-1, -1.3, -1.5, -2, -2.5, -3, 3.5
	For ages e_1 to 2.5 years, refer if			
	$SDS < f_5$	f_5	SDS cut off level after age e_1	-1, -1.3, -1.5, -2, -2.5,-3

Several screening rules based upon growth were studied. Each screening rule consists of parameters that we have varied. For more details, see the paragraph screening rules.

*Calculated for length, weight, and BMI.

In a simulation analysis, we varied each parameter to see how the diagnostic performance of each rule would change. As the effectiveness of a rule may depend on age, we also studied the effect of an age cut-off. For example, the effectiveness of a rule may be higher for children aged 1 onwards compared to younger children.

We used the following five rules:

1. The first rule (*delta rule*) refers a child if an absolute change in length SDS, weight SDS or BMI SDS occurs. For example, suppose a child has two weight measurements, one measurement at the age of six months and one at the age of 1.5 years. This child will be referred according to the delta rule with parameter g_1=-2 (see Table 1) if his or her weight decreases by more than 2 SD during this period.
2. The second rule (*extended delta rule*) is equal to the first rule with the extra condition that the second measurement must have a low SDS (for example below -1.5 SDS).
3. The third rule (*slowed growth rule*) signals whether an abnormal slowed growth for length, weight or BMI occurs in terms of change in SDS per year in combination with a current low SDS. Slowed growth requires measurements taken at least three months apart. For example, suppose a child has two length measurements, one measurement at the age of seven months and one measurement six months later. This child will be referred according to the slowed growth rule with parameters g_2=-1 and f_1=-1.5 (see Table 1) if the difference between the second and first length measurement per year exceeds 1 SD (which corresponds to a decrease of 0.5 SD within six months) and if the second measurement is below -1.5 SDS.
4. The fourth rule (*conditional weight gain rule)* signals whether a child's conditional weight gain SDS is below a certain value, in combination with the extra condition of a low weight SDS.[25,26] The conditional weight gain rule accounts for regression to the mean.
5. The fifth rule (*absolute SDS rule)* refers a child if the length SDS, weight SDS or BMI SDS is low. An example is to refer if a child's length SDS is below -2 (e_1 =0 and f_5 =-2).

Some parameter settings effectively select a subset of data. Rules 1 to 4 need the availability of multiple measurements or a measurement after a certain age. The rules were tested only on children for whom appropriate data were available. In the case of three or more measurements, all possible pairs of measurements were calculated. For example, if weight is measured at age A, B and C, the method calculates the weight gain for the intervals AB, BC and AC.

Statistical analysis

Each screening rule was implemented using S-Plus version 7.0.3 for Microsoft Windows (2005), and was applied to both sets of longitudinal data. We calculated sensitivity,

specificity and positive predictive value (PPV) for each rule under several scenarios. A scenario is a unique combination of parameter values. The rules were ordered according to their sensitivity at high levels of specificity. A higher sensitivity at the same level of specificity, results in a better performance. The results were plotted as a Receiver Operating Characteristic (ROC) plot, but scaled to a different axis than conventionally in order to view the area of most interest (high specificity). Scenarios of rules up to 2% false-positive rates were presented in detail, because low false-positive rates are desirable from a societal perspective. PPV was calculated assuming that the mean incidence of CF is 1 per 2,500 live births in the Caucasian population.[2] Tables of agreement and differences between rules in both the CF-group and reference group were calculated, because such tables provide insight into the diagnostic performance for different subsets of the data.

Length, weight and BMI measurements were expressed as SDS using the Dutch reference growth data.[27,28] In preterm infants (gestational age < 37 weeks) length and weight SDS were corrected for gestational age. The intrauterine growth charts from the Swedish reference population was used to express SDS up to the age corresponding with 40 weeks of gestation.[29] Between 40 and 42 weeks an interpolation between the growth curve of the Swedish reference population and that of the Dutch reference population was used. From 42 weeks of gestation till the age of 2 years, SDS was calculated on ages corrected for gestational age, using the Dutch reference growth data.

We assumed that a child would be referred if his or her growth pattern met the criteria of a given screening rule at the earliest age possible. All rules were dealt with separately, meaning that the same child could be referred according to each separate rule.

Results

Table 2 contains the general characteristics of the 123 CF-patients. Mean weight SDS at time of diagnosis was -1.7 SD for girls and -1.5 SD for boys (not statistically significant, data not shown).

Table 2 *General characteristics of the CF patients (n=123).*

Characteristics		% or mean/median (sd/range)
Gender (M)		51%
Ethnicity	Dutch/European	91%
	Turkish	2%
	Moroccan	1%
	Others	4%
	Unknown	2%
Median (range) age in years at time of diagnosis		0.59 (0-15)
Children with diagnosis at birth#		3%
Children diagnosed <1 year		62%
Children with ≥ 2 measurements between birth and diagnosis		64%
Mean (SD) length SDS at time of diagnosis∞		-1.08 (1.13)
Mean (SD) weight SDS at time of diagnosis∞		-1.60 (1.35)
Mean (SD) BMI SDS at time of diagnosis∞		-1.13 (1.79)
DNA	Homozygous for dF508	47.2%
	Heterozygous for dF508*	20.3%
	Others*	3.3 %
	Unknown	29.3%

One of their siblings is known with CF or the neonate presents with meconeum ileus
* Mutations other than dF508 were: 'A455E', 'G542X', 'N1303K', 'R1162X', 'R553X', '1717-1G>A' 'IVS17bTA', 'Q552P', 'R1066C', 'S1251N', 'G542x', '1677d', 'G178R', 'Q493X' and '3659delC'.
∞ Based on children with at least one measurement between 6 months before or 3 months after diagnosis.

Receiver Operating Characteristic curves

Figure 1 shows the ROC plot of the five best screening rules. The diagonal line indicates where sensitivity is equal to 100-specificity. The BMI extended delta was most successful in terms of high sensitivity at a low false-positive rate. Of the rules that consider only length, the length extended delta rule had the best diagnostic performance. All other rules that considered length and weight separately had sensitivities below 20% at a 2% false-positive rate.

Scenarios of the best screening rule

A very strict version of the BMI extended delta rule is a BMI SDS below -2 SD combined with a decrease in BMI of >3.0 SD between birth and 2.5 years of age. This scenario correctly identified 17% (95%-CI 8-26) of the CF children while 99.9% (95%-CI 99.8-100) of the reference children were correctly labelled as disease free. The PPV of this scenario is approximately 6%. For example, suppose a boy has a BMI of +1 SD at three months of age. If he crosses three SDS lines (SDS 0, -1 and -2) before the age of 2.5 years, then the boy has a 6% probability of having CF. Less strict rules identify more children, but the probability of having CF rapidly decreases. For example, in children with a BMI SDS below -2 SD and a decrease in BMI SDS of >1 SD between birth and 2.5 years of age, the probability of having CF is only 0.47%.

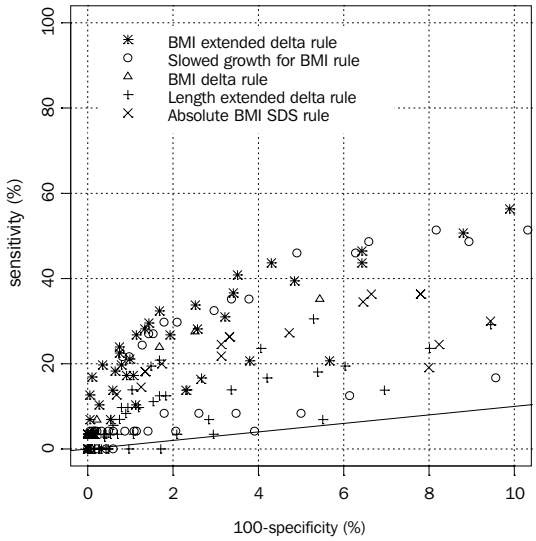


Figure 1 *ROC plot of effective screening rules to detect CF.*

Diagnostic performance of the rules

Table 3 presents the properties of the five best rules, in terms of sensitivity and PPV at approximately 2% false-positives. About 32% of all CF children and 1.7% of the reference children had a BMI SDS below -2.5 SD and a decrease in BMI SDS of more than 0.5 SD (PPV=0.75%). If this rule would have been used, median referral age would be almost three months earlier than the observed age at diagnosis. For example, suppose a girl has a BMI SDS on the -2 SD line at six month of age, and her BMI crosses the -2.5 SD line three months later. Then this girl will be referred according to the above scenario. Her probability of actually having CF is 0.75%.

The best rule using length as a parameter was the extended delta rule. Referral is warranted if length SDS is below -2.5 SD and if a decrease in length SDS of >0.5 SD occurs between birth and 2.5 years of age.

Table 3 *Simulation values and the percentage of detected CF children (sensitivity) with approximately 2% false-positives (=98% specificity).*

Rule	Growth	Simulation values			Sensitivity (95%-CI)	False-positives (95%-CI)	PPV
Extended delta rule	BMI	f_1=-2.5	g_2= -0.5		32 (21-43)	1.7 (1.1-2.3)	0.75%
Slowed growth	BMI	f_2=-2.5	g_3= -0.5		27 (13-41)	1.5 (1.0-2.0)	0.71%
Delta rule	BMI		g_1= -3		24 (14-34)	1.7 (1.1-2.3)	0.56%
Extended delta rule	Length	f_1=-2.5	g_2= -0.5		21 (11-31)	1.7 (1.1-2.3)	0.49%
Absolute SDS rule	BMI	e_1=1	f_4= -3	f_5= -2.5	20 (13-27)	1.7 (1.1-2.3)	0.47%

Agreement

Contingency tables of agreement between rules in the CF-group and in the reference group revealed that, except for the absolute BMI SDS rule, the diagnostic performance changed only slightly when looking at different subsets in the data. Sensitivity of the absolute BMI SDS rule was 5% higher (from 20% to 25%) for the children with at least two measurements. Therefore, this rule should actually be at the third position instead of the fifth in Table 3.

Discussion

Our study shows that a combination of a low BMI SDS (<-2.5 SDS) and a decrease of BMI SDS over the previous period (>0.5 SD) is the most sensitive rule to detect CF at an acceptable false-positive rate (1.7%). However, even for this rule, the PPV is low (0.75%), and for all scenarios PPV ranged from 0.5% to 6%. Thus, at best only 1 in 16 children that are referred according to BMI SDS actually have CF. While such a yield is low in absolute terms, one should realize that the prevalence of CF in the open population is about 1:2,500. Thus one could also argue that screening on BMI SDS is useful since it will increase the probability of identifying CF from 1:2,500 to 1:16.

With respect to the possible generalizability of our results, future studies on other CF patients must be awaited. However, the decrease of weight, length and BMI over time until diagnosis observed in our study is similar to the findings of earlier studies,[14-20] suggesting that the proposed screening rules may apply to other populations as well.

Screening rules for growth monitoring can be divided into rules with respect to a single measurement, with respect to velocity (e.g. a decrease in SDS or kg/year), and combinations. Traditionally, rules based on velocity have been considered as a better screening tool. However, such rules are more sensitive to measurement error than rules based on single growth measurements. Voss et al. reported that height velocity lacks the precision to provide a reliable index of growth in short children.[30] In our study, measurement errors may have led to more variation in velocity in the reference group and in the CF group. Therefore, one may need stricter cut-off values. Despite this phenomenon, it appears that velocity is a more informative predictor of CF than a single measurement. We found similar results for the detection of children with celiac disease.[24]

The term “failure to thrive”, though used for a long time, suffers from a lack of consensus on its definition.[31] Some authors define FTT as weight or height falling below the third or fifth centile, or falling two major centiles of the standard National Center for Health Statistics growth chart. Others state that malnutrition (weight <80% of ideal body weight for age) should be present to state that a child is failing to thrive.[32,33] We believe that

our approach, including studying the diagnostic performance of various definitions of FTT, is a fruitful basis for a discussion of the concept of FTT. While we have shown that BMI SDS and its decrease is useful for detecting CF and celiac disease, body length is superior to BMI in detecting girls with Turner's syndrome.[34] The optimal definition depends on the pathological causes of growth failure in infants.

In countries where newborn screening for CF is not available, growth monitoring (including assessment of BMI and its change over time), combined with a thorough medical history and physical examination, may be considered as the best screening procedure for CF. In infants and toddlers with an increased probability of CF, a sweat test is the next step. The diagnostic performance of the sweat test is known to vary, and widely different practices and standards in sweat testing are used. As a follow-up of this study, it would be interesting to investigate the diagnostic properties of sweat testing in combination with referral for FTT. Note that such a study would require that sweat testing is performed according to an evidence-based guideline. [35]

Newborn screening has shown beneficial effects on the prognosis of CF and has been implemented in several countries.[8-12] In countries where such program is absent, we suggest the clinician to consider testing for CF (i.e. a sweat test) in the diagnostic work-up in infants and young children with FTT. The most sensitive growth parameter is a combination of a low BMI SDS at examination and a decrease of BMI SDS over the previous period. Referral criteria implemented in a computer system in Community Child Health Care Centers can be helpful to perform the calculations. We recommend that future research with larger samples of children with CF should be performed to further optimize referral criteria.

Acknowledgements

We acknowledge Wendelien Planten for her great contribution to the study, all the members of the CF team of the Erasmus MC - Sophia Children's Hospital, with Dr. H.A.W.M. Tiddens as coordinator, for gathering the information, Dorien Holtslag for her help gathering all information in Maastricht and Pfizer for their financial support.

References

1. Ratjen F, Döring G. Cystic fibrosis. Lancet 2003;361:681-9.
2. Bobadilla JL, Macek M, Jr., Fine JP, Farrell PM. Cystic fibrosis: a worldwide analysis of CFTR mutations--correlation with incidence data and application to screening. Hum Mutat 2002;19:575-606.
3. Davies JC, Alton EW, Bush A. Cystic fibrosis. BMJ 2007;335:1255-9.

4. Giglio L, Candusso M, D'Orazio C, Mastella G, Faraguna D. Failure to thrive: the earliest feature of cystic fibrosis in infants diagnosed by neonatal screening. Acta Paediatr 1997;86:1162-5.
5. Kharrazi M, Kharrazi LD. Delayed diagnosis of cystic fibrosis and the family perspective. J Pediatr 2005;147:S21-S25.
6. Farrell PM. Improving the health of patients with cystic fibrosis through newborn screening. Adv Pediatr 2000;47:79-115.
7. Dodge JA, Morison S, Lewis PA. Incidence, population, and survival of cystic fibrosis in the UK, 1968-1995. Arch Dis Child 1997;77:493-6.
8. Farrell PM, Kosorok MR, Rock MJ, Laxova A, Zeng L, Lai HC, Hoffman G, Laessig RH, Splaingard ML. Early diagnosis of cystic fibrosis through neonatal screening prevents severe malnutrition and improves long-term growth. Wisconsin Cystic Fibrosis Neonatal Screening Study Group. Pediatrics 2001;107:1-13.
9. Dankert-Roelse JE, te Meerman GJ. Long term prognosis of patients with cystic fibrosis in relation to early detection by neonatal screening in a cystic fibrosis centre. Thorax 1995;50:712-8.
10. Lai HJ, Cheng Y, Farrell PM. The survival advantage of patients with cystic fibrosis diagnosed through neonatal screening: evidence from the United States Cystic Fibrosis Foundation registry data. J Pediatr 2005;147:S57-S63.
11. Sims EJ, McCormick J, Mehta G, Mehta A. Neonatal screening for cystic fibrosis is beneficial even in the context of modern treatment. J Pediatr 2005;147:S42-S46.
12. Dankert-Roelse JE,.Merelle ME. Review of outcomes of neonatal screening for cystic fibrosis versus non-screening in Europe. J Pediatr 2005;147:S15-S20.
13. Farrell PM, Rosenstein BJ, White TB, Accurso FJ, Castellani C, Cutting GR, Durie PR, Legrys VA, Massie J, Parad RB, Rock MJ, Campbell PW 3rd; Cystic Fibrosis Foundation. Guidelines for diagnosis of cystic fibrosis in newborns through older adults: Cystic Fibrosis Foundation consensus report. J Pediatr 2008;153:S4-S14.
14. Haeusler G, Frisch H, Waldhor T, Gotz M. Perspectives of longitudinal growth in cystic fibrosis from birth to adult age. Eur J Pediatr 1994;153:158-63.
15. Sproul A, Huang N. Growth patterns in children with cystic fibrosis. J Pediatr 1964;65:664-76.
16. Karlberg J, Kjellmer I, Kristiansson B. Linear growth in children with cystic fibrosis: I. Birth to 8 years of age. Acta Paediatr Scand 1991;80:508-14.
17. Morison S, Dodge JA, Cole TJ, Lewis PA, Coles EC, Geddes D, Russell G, Littlewood JM, Scott MT. Height and weight in cystic fibrosis: a cross sectional study. Arch Dis Child 1997;77:497-500.
18. Lai HC, Kosorok MR, Sondel SA, Chen ST, FitzSimmons SC, Green CG, Shen G, Walker S, Farrell PM. Growth status in children with cystic fibrosis based on the National Cystic Fibrosis Patient Registry data: evaluation of various criteria used to identify malnutrition. J Pediatr 1998;132:478-85.

19. Kraemer R, Rudeberg A, Hadorn B, Rossi E. Relative underweight in cystic fibrosis and its prognostic value. Acta Paediatr Scand 1978;67:33-7.
20. Pamakcu A, Bush A, Buchdahl R. Effects of Pseudomonas aeruginosa colonization on lung function and anthropometric variables in children with cystic fibrosis. Pediatr Pulmonol 1995;19:10-5.
21. Landsmeer-Beker EA, Bouwmeester-Landweer MBR, Korbee-Haverhoek HD, Kousemaker NPJ, Baartman HEM, Wit JM, Dekker FW. Non-response on a postal questionnaire adressing risk factors for child maltreatment. In: Bouwmeester-Landweer MBR. Early home visitation in families at risk for maltreatment. Thesis, Leiden University Medical Center, 2006.
22. Herngreen WP, Reerink JD, van Noord-Zaadstra B.M., Verloove-Vanhorick SP, Ruys JH. The SMOCC-study: Design of a representative cohort of live-born infants in the Netherlands. Eur J Public Health 1992;2:117-22.
23. Herngreen WP, van Buuren S, van Wieringen JC, Reerink JD, Verloove-Vanhorick SP, Ruys JH. Growth in length and weight from birth to 2 years of a representative sample of Netherlands children (born in 1988-89) related to socioeconomic status and other background characteristics. Ann Hum Biol 1994;21:449-63.
24. van Dommelen P, Grote FK, Oostdijk W, Keizer-Schrama SM, Boersma B, Damen GM, Csizmadia CG, Verkerk PH, Wit JM, van Buuren S. Screening rules for growth to detect celiac disease: a case-control simulation study. BMC Pediatr 2008;8:35.
25. Cole TJ. Conditional reference charts to assess weight gain in British infants. Arch Dis Child 1995;73:8-16.
26. Cole TJ. Presenting information on growth distance and conditional velocity in one chart: practical issues of chart design. Stat med 1998;17:2697-707.
27. Fredriks AM, van Buuren S, Burgmeijer RJ, Meulmeester JF, Beuker RJ, Brugman E, Roede MJ, Verloove-Vanhorick SP, Wit JM. Continuing positive secular growth change in The Netherlands 1955-1997. Pediatr Res 2000;47:316-23.
28. Fredriks AM, van Buuren S, Wit JM, Verloove-Vanhorick SP. Body index measurements in 1996-7 compared with 1980. Arch Dis Child 2000;82:107-12.
29. Niklasson A, Ericson A, Fryer JG, Karlberg J, Lawrence C, Karlberg P. An update of the Swedish reference standards for weight, length and head circumference at birth for given gestational age (1977-1981). Acta Paediatr Scand 1991;80:756-62.
30. Voss LD, Wilkin TJ, Bailey BJ, Betts PR. The reliability of height and height velocity in the assessment of growth (the Wessex Growth Study). Arch Dis Child 1991;66:833-7.
31. Steward DK, Ryan-Wenger NA, Boyne LJ. Selection of growth parameters to define failure to thrive. J Pediatr Nurs 2003;18:52-9.
32. Bithoney WG, Dubowitz H, Egan H. Failure to thrive/growth deficiency. Pediatr Rev 1992;13:453-60.

33. Bassali RW, Benjamin J. Failure to Thrive. Emedicine, August 11, 2004.
34. Grote FK, van Dommelen P, Oostdijk W, de Muinck Keizer-Schrama SM, Verkerk PH, Wit JM, van Buuren S. Developing evidence-based guidelines for referral for short stature. Arch Dis Child 2008;93:212-7.
35. Baumer JH. Evidence based guidelines for the performance of the sweat test for the investigation of cystic fibrosis in the UK. Arch Dis Child 2003;88:1126-7.

Reference chart for relative weight change to detect hypernatraemic dehydration

9

Paula van Dommelen[1], Jacobus P. van Wouwe[1], Jacqueline M. Breuning-Boers[1], Stef van Buuren[1] and Paul H. Verkerk[1]

[1]TNO Quality of Life, Leiden, The Netherlands

Arch Dis Child 2007;92:490-4.

Abstract

Objective: The validity of the rule of thumb that infants may have a weight loss of 10% in the first days after birth is unknown. We assessed the validity of this and other rules to detect breast-fed infants with hypernatraemic dehydration.
Design: A reference chart for relative weight change was constructed by the LMS method. The reference group was obtained by a retrospective cohort study. Participants: 1,544 healthy, exclusively breast-fed infants with 3,075 weight measurements born in the Netherlands and 83 cases of breast fed infants with hypernatraemic dehydration obtained from literature.
Results: The rule of thumb had a sensitivity of 90.4%, a specificity of 98.3% and a positive predictive value of 3.7%. Referring infants if their weight change is below -2.5 SDS (0.6th centile) in the reference chart in the first week of life and using the rule of thumb in the second week had a sensitivity of 85.5%, a specificity of 99.4% and a positive predictive value of 9.2%.
Conclusions: The rule of thumb is likely to produce too many false-positive results, assuming that for screening purposes the specificity needs to be high. A chart for relative weight change can be helpful to detect infants.

Introduction

Exclusive breast feeding up to the sixth month of life is important for optimum infant development and growth as breast milk contains all the necessary nutrients in ideal proportions.[1] Breast feeding protects against infections and allergies, and plays a major role in mother-infant bonding.[2]

In the Netherlands, 78% of mothers initiated breast feeding in the period 2001-2003. After 1 month 51% and after 4 months 25% of infants were fed primarily on human milk.[3] The WHO and UNICEF started the "Baby Friendly Hospital Initiative" to promote breast feeding.[4] In the Netherlands, this program is mainly focused on improvement of support and encouragement of breast feeding in general health care.

Almost all mothers are capable of breast feeding their infant successfully. However, in some cases initial milk supply is insufficient because of a poor start to milk production or transfer. If the infant's needs are not met for several days, dramatic weight loss and an increase in serum sodium concentration occur and the infant develops hypernatraemic dehydration.[5-7] Hypernatraemic dehydration may cause serious complications, such as fits, disseminated intravascular coagulation and multiple cerebrovascular accidents, and may even result in death.[8-12]

A retrospective, population-based study reported an incidence rate of hypernatraemic dehydration of 7.1 per 10,000 breast-fed infants.[6] This is probably the minimum incidence as cases may have been missed because they occurred in infants before initial discharge from hospital[13,14] or because appropriate investigations were not performed. The clinical day of presentation of hypernatraemic dehydration is usually around 10 days of age.[6]

In clinical practice, weighing is an essential part of the assessment of an infant's growth and hydration status. However, there is no evidence-based consensus for "normal" and "abnormal" early relative weight change (RWC). Several studies reported the normal (50th centile) or extreme (1st, 2.5th or 5th) RWC centile for exclusively breast-fed infants.[6,15] However, these centiles are not precisely described with respect to day of measurement nor shown on standard growth charts. Several authors propose different rules of thumbs for identifying "abnormal" RWC.[16-20] It is suggested that many midwives use the rule of thumb that infants may have a weight loss of 10% (= -10% RWC) and should regain birth weight by 10-14 days of life.[5] To our knowledge, no evidence-based referral rule is available to detect infants with hypernatraemic dehydration.

This study describes a reference chart for breast-fed infants between postnatal days 2 and 11. This chart, together with reports of hypernatraemic dehydration obtained from the literature, will be used to define an evidence-based referral rule. The centiles of the chart can be used as a test to detect infants with hypernatraemic dehydration. The test is considered positive if a breast-fed infant's relative weight decreases below a chosen centile and negative if it stays above. Sensitivity, specificity and positive predictive value (PPV) will be used to optimize this rule. This test will be compared to the rule of thumb that infants may have a maximal weight loss of 10%.

Methods

Population

We selected a representative reference group of healthy, exclusively breast-fed infants and a group of breast-fed infants diagnosed with hypernatraemic dehydration. The reference group was obtained from a retrospective cohort study initiated in three primary care midwife practices in the Netherlands (metropolitan Amsterdam South-East, rural Heerhugowaard and the country town of Veenendaal). In the Netherlands, a midwife either assists the delivery at home or in an outpatient clinic, or is involved in follow-up care after hospital delivery by a gynecologist. We selected 1,544 infants born in 2002 with a weight measurement (in grams) at birth and at least one weight measurement between postnatal days 2 and 11. The infants were weighed at home by a midwife with a calibrated electronic scale.

Infants hospitalized with hypernatraemic dehydration were identified by a literature search. Articles written in Dutch, English, French or German published between 1970 and 2005 that describe infants with hypernatraemic dehydration were obtained using the search program PubMed with the MESH terms "dehydration" and "breastfeeding". References in these articles were used to increase the number of articles describing infants with hypernatraemic dehydration. We assumed that an infant had hypernatraemic dehydration when the author(s) of the article diagnosed the infant as such. In 47 articles we identified 129 cases of breast-fed infants with hypernatraemic dehydration with a weight measurement at birth and day of presentation or a calculated RWC at day of presentation. A total of 83 literature cases had a day of presentation between 2 and 11 days of life and these were used in this study.[6,9,11,14,19,21-42] Serum sodium concentration was known for 80 literature cases. All cases were born at term.

Statistical analysis

RWC was calculated as the difference in weight at day of presentation (w(t)) and birth weight (w(t0)) divided by birth weight in percentage, or in formula: 100%*(w(t)-w(t0))/ w(t0)). Day of birth was represented by day 0. A reference chart for relative weight was obtained by the LMS method.[43] The LMS method summarizes the distribution of relative weight as it changes according to age by three curves representing the Box-Cox power (L-curve), the median (M-curve) and the coefficient of variation (S-curve). The L-, M-, and S-curves were used to convert data into standard normally distributed data. Such a data point is called a z score or standard deviation score (SDS). Normality of SDS was tested by so called "worm plots" for different age groups.[44] A log power-transformation was applied to age in the LMS method. Since the LMS method works only with positive values, an amount of 25% was added to relative weight and afterwards subtracted from the centiles. Each infant had multiple weight measurements. All weights were included in the analysis and were treated as independent as we did not find an association between the number of measurements and birth weight (t-test, t=1.14, p=0.26).

The centiles of the curve and the 10% weight loss were used as a test. The specificity of the 10% rule was calculated as the mean of the percentiles of the reference chart that have a 10% weight loss for each day. To calculate PPV, we assumed that the incidence of hypernatraemic dehydration is 7.1 per 10,000 breast-fed infants.[6]

Calculations for the LMS method were performed with LMS Light version 1.16 (Institute of Child Health, London, UK) compiled on 15 April 2002. All other analyses were performed with S-plus version 6.2 (Insightful, Seattle, WA, USA).

Results

The characteristics of the reference infants are given in Table 1. The number of measurements in reference infants and in those with hypernatraemic dehydration are shown in Table 2.

Table 1 *Characteristics of healthy, breast-fed infants (n=1,544).*

Characteristics		Means (SD) or %
Maternal age in years		30 (4.7)
Girls (%)		49
Gestation in wks#		39.5 (1.4)
Preterm <37 wks (%)		2.0
Parity (%)	First	45
	Second	36
	Third or more	19
Delivery (%)	Spontaneous	80
	Caesarean section	10
	By vacuum extraction or forceps	10
Birth weight in kg		3.44 (0.46)

#N=1,543

Table 2 *Number of measurements between 2-11 days of life in healthy, breast-fed infants and infants with hypernatraemic dehydration.*

Characteristics		healthy, breast-fed infants	infants with hypernatraemic dehydration
Number of infants		1,544	83
Number of measurements on	day 2	9	0
	day 3	505	9
	day 4	263	9
	day 5	618	4
	day 6	128	15
	day 7	287	10
	day 8	272	11
	day 9	864	6
	day 10	93	16
	day 11	36	3

RWC was not normally distributed (Shapiro-Wilk normality test: $W=0.975$, $p<0.01$). To obtain normally distributed SDS for RWC, we used the LMS method with a Box-Cox power transformation of approximately 0.5. Normality of SDS was tested by worm plots of different age groups. The shape of the worm plots was reasonably flat, indicating that the data follow the assumed distribution in this age period.

Figure 1 shows a reference chart with standard deviation lines of the RWC of healthy, breast-fed infants as well as the RWC of 83 infants with hypernatraemic dehydration on the day of presentation. The rule of thumb of 10% weight loss is also indicated on the chart. The standard deviation lines or percentiles on this chart show which percentages of infants have the same RWC. For example, if a 5 day old infant weighs 3315 g and has a birth weight of 3750 g, then the calculated RWC is 100%*(3315-3750)/3750=-11.6%. Notice that -11.6% RWC at day 5 on the chart corresponds to -2.6 SDS or the 0.5th percentile. This means that only 0.5% of 5 day old infants have a RWC less than this infant. To avoid the user calculating weight as a percentage, we converted the -2.5 SDS RWC centile to weights by age for a given birth weight. This converted -2.5 SDS centile is shown on fig 2 for different birth weights. The infant in the previous example has a birth weight of 3750 g. The -2.5 SDS centile for this infant is shown by the fourth line from the top, starting at 3750 at day 0. Follow this line until you reach day 5 and notice that 3315 g at day 5 is just below the line.

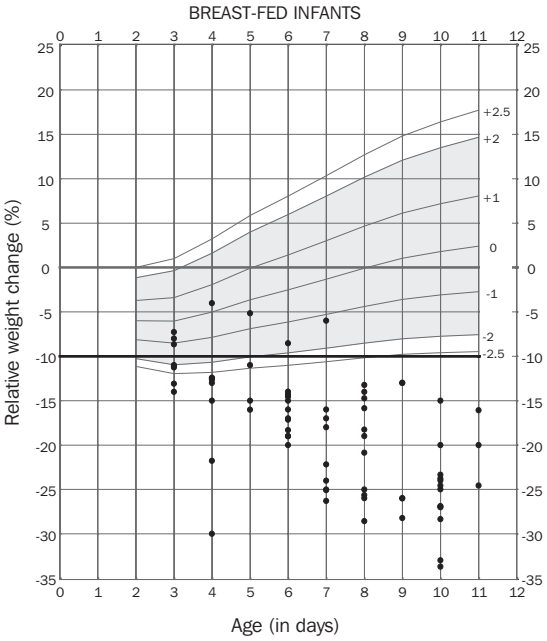


Figure 1 *Reference chart with standard deviation lines of relative weight change (=100 %* (weight - birth weight) / birth weight) for healthy, breast-fed infants as well as relative weight change in 83 cases of hypernatraemic dehydration at their day of presentation (day of birth is day 0) and the rule of thumb of 10% weight loss.*

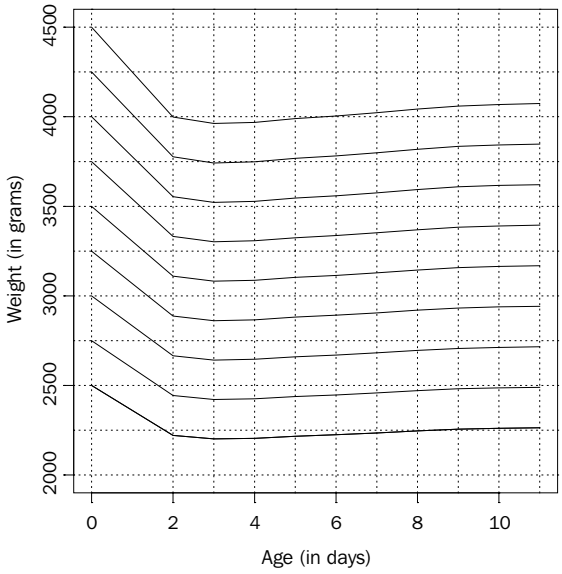


Figure 2 *The -2.5 SDS relative weight change centile converted to weights by age for a given birth weight.*

Maximal negative RWC for healthy, breast-fed infants is at 3 days after birth, with a mean RWC of -6.0% (95% CI: -5.7% to -6.2%). The mean increases by approximately 1% per day from -6% at day 3 to 0% at day 8. However, even after 11 days about a third of these infants have not yet regained their birth weight; in contrast with healthy, breast-fed infants, the mean of these patients is consistently declining. The mean RWC for the infants with hypernatraemic dehydration is -18.5% (95% CI: -17.0% to -19.9%). The mean decreases by approximately 2% per day from -10% at day 3 to -25% at day 10.

Notice that there were no cases of hypernatraemic dehydration before day 3, probably due to the fact that it takes some time before insufficient breast feeding leads to weight loss. We therefore applied the rules from 3 days up until 11 days after birth.

Table 3 shows sensitivity, specificity and PPV for several referral rules: the rule of thumb (10% test), the SDS rules and a combination of the -2.5 SDS test in the first week (3-6 days after birth) and the 10% test after the first week. All sensitivities for these tests were above 85% with less than 3% false positives. The sensitivity of the 10% test was similar to that of the -2 SDS rule and specificity was slightly higher in the first week, although not significantly so ($p>0.05$). Combining the -2.5 SDS test in the first week with the 10% rule after the first week results in a sensitivity of 85.5% and a specificity of 99.4%; this is similar to the -2.5 SDS test for the first 2 weeks. This specificity is significantly higher ($p<0.05$) than that for the -2 SDS rule.

Table 3 *Sensitivity, specificity and positive predictive value (PPV) for several referral rules in the period from 3 days up until 11 days after birth.*

Test	Sensitivity (%)	Specificity (%)	PPV (%)
10%	90.4	98.3	3.7
-2.5 SDS	85.5	99.4	9.2
-2 SDS	90.4	97.7	2.7
-2.5 SDS 3-6 days and 10% 7-11 days	85.5	99.4	9.2

Cases with a positive -2 or -2.5 SDS test had a significantly higher mean serum sodium concentration (163 mM) compared to cases with a negative -2 SDS test (149 mM) (t=2.6, df=78, p=0.01) and with a negative -2.5 SDS test (151 mM) (t=3.0, df=78, p=0.004). Of the cases with a positive -2.5 SDS test, 89% had a concentration of >149 mM, so the test detects the more severe cases of dehydration. Of the cases with a concentration of >149 mM, 91% had a positive -2.5 SDS test and 97% a positive -2 SDS test, and of the cases with a concentration of >159 mM, all cases had a positive -2.5 SDS test (and therefore a positive -2 SDS test).

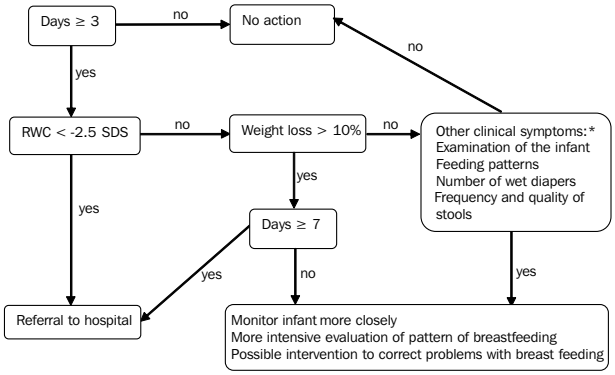
Eight cases of hypernatraemic dehydration had a very small RWC. Three cases had a RWC between -2 SDS and -1 SDS and five cases had a weight change above -1 SDS.[19,35,36,42] Clinical information was given in some studies; only mild and transient symptoms in these infants were reported. Serum sodium concentration was reported for six cases: four cases had a concentration below 149 mM and two above 149 mM (both 157 mM).

Discussion

We developed a reference chart for breast-fed infants between postnatal days 2 and 11. This chart, together with cases of hypernatraemic dehydration obtained from the literature, was used to define an evidence-based referral rule. As far as we know, this is the first reference chart for RWC and the first evidence-based investigation of referral rules to detect infants with hypernatraemic dehydration. Our results show that a reference chart for RWC can be helpful to detect infants with hypernatraemic dehydration.

The RWC chart shows that the mean maximal weight loss occurs 3 days after birth and is 6% for a healthy, breast-fed infant. This is in agreement with several other studies which reported that breast-fed infants may lose up to 6%[45-47] or 7%[15,19,48,49] of their birth weight during the first week of life. The American Academy of Pediatrics and others also reported that normal weight loss reaches its peak at 3-5 days after birth.[50] Livingstone and the American Academy of Pediatrics Work Group on Breastfeeding suggested that a weight loss of greater than 7% of birth weight indicates possible breast feeding problems.[19,50] Others suggested that a weight loss of 8% or more warrants further investigation.[16-18]

Most authors reported that many midwives use the rule of thumb that infants may lose up to 10% of birth weight. Our results show that most infants with hypernatraemic dehydration have a weight loss of >10%. However, referral to a hospital of all infants with a weight loss of >10% would probably lead to many false positive results in the first week of life, assuming that for screening purposes the specificity needs to be sufficiently high. Therefore, we suggest applying the 0.6th centile (-2.5 SDS) as a criterion for referral to a hospital in the first week of life or using a weight loss of >10% after the first week of life. At the hospital, further diagnostic biochemical testing should be carried out. Clinical differentiation between normal infants and those with hypernatraemic dehydration is not really possible in the first 2 days after birth. Infants with a weight loss of >10% (or -2 SDS) in the first week after day 2, should be monitored closely and require more intensive evaluation of breast feeding and possible intervention to correct problems with breast feeding. Furthermore, referral may also be warranted in infants with other clinical symptoms of dehydration even if weight loss is not particularly high. Clinicians should combine RWC values with examination of the infant, knowledge of feeding patterns, and number of wet diapers and frequency and quality of stools. We suggest using the flowchart in fig 3.



* Monitor the infant when mild clinical symptoms are present and refer the infant to the hospital when severe clinical symptoms are present

Figure 3 *Flowchart to detect dehydrated infants or infants that are at risk of dehydration.*

In addition to the 10% weight loss, another rule of thumb among midwives is that infants regain their birth weight by 10-14 days. The chart in this study shows that 50% of infants have regained their birth weight 8 days after birth, which is also consistent with other reports.[6,15] This study also shows that even after 11 days, about a third of infants have not yet regained their birth weight. We also expect that at day 14 a high percentage of infants will not have regained their birth weight. Therefore, we assume that this rule will

lead to many false positive results. Macdonald et al[15] suggested a revised intervention criterion: offer additional breast feeding support to those losing 10% of their birth weight but still consider this as normal and only consider weight loss above 12.5% or failure to regain birth weight by 21 days as being abnormal and requiring medical assessment. We applied the 12.5% weight loss rule to our data with infants from birth to 11 days old and found a sensitivity of 83.1% and a specificity of 99.9%. This rule has a better specificity (+0.5%) at the cost of a lower sensitivity (-2.4%) compared to the -2.5 SDS rule. With the 12.5% weight loss rule, 2.4% of the cases are missed. We think that a decrease in sensitivity of 2.4% is high and we therefore recommend using the proposed flow chart. However, one could consider using the 12.5% weight loss rule at day 3 as the -2.5 SDS line almost reaches 12.5% at day 3.

In our study we used information from cases with hypernatraemic dehydration reported in the literature. We expected that this information is biased towards the more severe cases of hypernatraemic dehydration, since severe cases are more likely to be reported than mild cases. Recently Moritz et al[51] found that only 17% of cases of hypernatraemic dehydration had non-metabolic complications. Therefore, the sensitivity and PPV in this study are likely to be lower for all infants with hypernatraemic dehydration. On the other hand, PPV may also be an underestimate as this value was based on a minimum incidence rate of hypernatraemic dehydration. It would be very interesting in the future to test and possibly optimize our proposed referral rules using new cases with dehydration.

There is evidence that the degree of weight loss in babies born in a particular environment may be associated with the way that environment is managed.[52-53] In populations with ”baby friendly” care, the prevalence of hypernatraemic dehydration may be lower than in populations with care that is less baby friendly. We assumed that the prevalence of hypernatraemic dehydration is 7.1 per 10,000 breast-fed infants. Based on this prevalence we calculated the PPV of several referral criteria. Since PPV is dependent on prevalence, in populations with a lower prevalence (perhaps due to baby friendly care) the PPV may be lower, whereas in populations with a higher prevalence the PPV of the same referral criteria will be higher.

We assumed that RWC expressed as a percentage is uncorrelated with birth weight. This means that a heavy child and a light child have the same distribution of RWC. However, this may not be true, as the degree, timing and variability of RWC may be quite different in small infants compared to large infants. We therefore tested the relationship between birth weight and RWC corrected for age using a linear mixed-effects model (residual variance=1.53, AIC=15 864). We found that an infant with a birth weight of 2.5 kg has on average a 1% greater RWC than an infant with a birth weight of 4.5 kg. As this is a relatively small difference for a large difference in birth weight, we decided to use the

methodology unconditional on birth weight. The latter approach is also more convenient in practice than, for instance, various RWC curves for different categories of birth weight.

In this study, the infants were weighed routinely. This means that the number of measurements should not depend on the status of the infant. To determine if this is indeed the case, using standard two-sample t-tests we tested the dependence of the number of measurements and the status of the infants by testing the difference in RWC each day between the infants whose weight was being measured for the first time (besides their birth weight) and those who were being reweighed. We refitted the LMS method without the cases which were possibly reweighed because of a high RWC, and found that the difference between the median RWC in the newly constructed growth chart and the reference chart based on all infants was negligible ($\leq$-0.2%).

We conclude that the rule of thumb that infants may have a relative weight loss of 10% is excellent after the first week of life. However, in the first week of life this rule will produce too many false positive results. A chart for RWC can be helpful to detect infants with hypernatraemic dehydration.

Acknowledgements

We are grateful to Mieke Huisenga, Wendelien Roepke, Marleen Snabilie and the midwife practices for obtaining the reference group data. We thank Caren Lanting and Jan Maarten Wit for their suggestions for improvement.

References

1. World Health Organization. The optimal duration of exclusive breastfeeding. Report of an expert consultation. Geneva: World Health Organization, 2001.
2. Lawrence RA, Howard CR. Given the benefits of breastfeeding, are there any contraindications? Clin Perinatol 1999;26:479-90,viii.
3. Lanting CI, Van Wouwe JP, Reijneveld SA. Infant milk feeding practices in the Netherlands and associated factors. Acta Paediatr 2005;94:935-42.
4. Naylor AJ. Baby-Friendly Hospital Initiative. Protecting, promoting, and supporting breastfeeding in the twenty-first century. Pediatr Clin North Am 2001;48:475-83.
5. Laing IA, Wong CM. Hypernatraemia in the first few days: is the incidence rising? Arch Dis Child Fetal Neonatal Ed 2002;87:F158-62.
6. Oddie S, Richmond S, Coulthard M. Hypernatraemic dehydration and breast feeding: a population study. Arch Dis Child 2001;85:318-20.

7. Clarke TA, Markarian M, Griswold W, et al. Hypernatremic dehydration resulting from inadequate breast-feeding. Pediatrics 1979;63:931-2.
8. Arboit JM, Gildengers E. Breast-feeding and hypernatremia. J Pediatr 1980;97:335-6.
9. Rowland TW, Zori RT, Lafleur WR, et al. Malnutrition and hypernatremic dehydration in breast-fed infants. JAMA 1982;247:1016-7.
10. Clarke AJ, Sibert JR. Hypernatraemic dehydration and necrotizing enterocolitis. Postgrad Med J 1985;61:65-6.
11. Cooper WO, Atherton HD, Kahana M, et al. Increased incidence of severe breastfeeding malnutrition and hypernatremia in a metropolitan area. Pediatrics 1995;96:957-60.
12. Kaplan JA, Siegler RW, Schmunk GA. Fatal hypernatremic dehydration in exclusively breast-fed newborn infants due to maternal lactation failure. Am J Forensic Med Pathol 1998;19:19-22.
13. Chaudhary R, Twaddle S, Levi R, et al. Hypernatremic dehydration in breast-fed infants. Paediatrics Today 1998;6:85-7.
14. Ng PC, Chan HB, Fok TF, et al. Early onset of hypernatraemic dehydration and fever in exclusively breast-fed infants. J Paediatr Child Health 1999;25:585-7.
15. Macdonald PD, Ross SR, Grant L, et al. Neonatal weight loss in breast and formula fed infants. Arch Dis Child Fetal Neonatal Ed 2003;88:F472-6.
16. Neifert MR. Clinical aspects of lactation. Promoting breastfeeding success. Clin Perinatol 1999;26:281-306,v-vi.
17. Neifert MR. Prevention of breastfeeding tragedies. Pediatr Clin North Am 2001;48:273-97.
18. Simoes EA, Pereira SM. The growth of exclusively breastfed infants. Ann Trop Paediatr 1986;6:17-21.
19. Livingstone VH, Willis CE, Abdel-Wareth LO, et al. Neonatal hypernatremic dehydration associated with breast-feeding malnutrition: a retrospective survey. CMAJ 2000;162:647-52.
20. DeMarzo S, Seacat J, Neifert M. Initial weight loss and return to birth weightcriteria for breast-fed infants: challenging the 'rule of thumb'. Am J Dis Children 1991;145:402.
21. Bajpai A. Hypernatremic dehydration in a neonate. Indian Pediatr 2002;39:599-600.
22. Becker PG, Conard JA. Breast-feeding failure. Indiana Med 1990;83:648-50.
23. Boumahni B, Pyaraly S, Randrianaly H, et al. Hypernatremic dehydration and breastfeeding. Arch Pediatr 2001;8:731-3.
24. Canet E, Canet J, Perrain MF. Quel est votre diagnostic? Concours Med 1984;106:1689-91.
25. Chilton LA. Prevention and management of hypernatremic dehydration in breastfed infants. West J Med 1995;163:74-6.

26. Gebara BM, Everett KO. Dural sinus thrombosis complicating hypernatremic dehydration in a breastfed neonate. Clin Pediatr (Phila) 2001;40:45-8.
27. Harding D, Moxham J, Cairns P. Weighing alone will not prevent hypernatraemic dehydration. Arch Dis Child Fetal Neonatal Ed 2003;88:F349.
28. Hilliard TN, Marsh MJ, Malcolm P, et al. Radiological case of the month. Sagittal sinus thrombosis in hypernatremic dehydration. Arch Pediatr Adolesc Med 1998;152:1147.
29. Jaramillo I, Lopez G, Hernandez H. Hypernatremic dehydration and death in an infant. Pediatr Emerg Care 2003;19:62-3.
30. Kini N, Zahn S, Werlin SL. Hypernatremic dehydration in breast-fed infants. Wis Med J 1995;94:143-5.
31. Macdonald PD, Grant L, Ross SR. Hypernatraemia in the first few days: a tragic case. Arch Dis Child Fetal Neonatal Ed 2003;88:F350.
32. Marino R, Gourji S, Rosenfeld W. Neonatal metabolic casebook. Hypernatremia and breast feeding. J Perinatol 1989;9:451-3.
33. Mercier JC, Outin S, Paradis K, et al. [Breast feeding and hypernatremic dehydration. 3 case studies]. Arch Fr Pediatr 1986;43:465-70.
34. Molteni KH. Initial management of hypernatremic dehydration in the breastfed infant. Clin Pediatr (Phila) 1994;33:731-40.
35. Niestijl AL, Sauer PJ. Breast feeding during the first few days after birth: sometimes insufficient. Ned Tijdschr Geneeskd 2003;147:2405-7.
36. Pascale JA, Brittian L, Lenfestey CC, et al. Breastfeeding, dehydration, and shorter maternity stays. Neonatal Netw 1996;15:37-43.
37. Rand SE, Kolberg A. Neonatal hypernatremic dehydration secondary to lactation failure. J Am Board Fam Pract 2001;14:155-8.
38. Roddey OF, Martin ES, Swetenburg RL. Critical weight loss and malnutrition in breast-fed infants. Am J Dis Child 1981;135:597-9.
39. Rosenfeld W, deRomana GL, Kleinman R, et al. Improving the clinical management of hypernatremic dehydration. Observations from a study of 67 infants with this disorder. Clin Pediatr (Phila) 1977;16:411-7.
40. Sofer S, Ben Ezer D, Dagan R. Early severe dehydration in young breast-fed newborn infants. Isr J Med Sci 1993;29:85-9.
41. van der Heide PA, Toet MC, van Diemen-Steenvoorde JA, et al. Hypertonic dehydration in ‘‘silent’’ malnutrition of breast-fed infants. Ned Tijdschr Geneeskd 1998;142:993-5.
42. Willis CE, Livingstone V. Infant insufficient milk syndrome associated with maternal postpartum hemorrhage. J Hum Lact 1995;11:123-6.
43. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med 1992;11:1305-19.

44. van Buuren S, Fredriks M. Worm plot: a simple diagnostic device for modelling growth reference curves. Stat Med 2001;20:1259-77.
45. Maisels MJ, Gifford K. Breast-feeding, weight loss and jaundice. J Pediat 1983;102:117-18.
46. Marchini G, Fried G, Ostlund E, et al. Plasma leptin in infants: relations to birth weight and weight loss. Pediatrics 1998;101:429-32.
47. Marchini G, Stock S. Thirst and vasopressin secretion counteract dehydration in newborn infants. J Pediatr 1997;130:736-9.
48. Maisels MJ, Gifford K, Antle CE, et al. Jaundice in the healthy newborn infant: a new approach to an old problem. Pediatrics 1988;81:505-11.
49. Thureen PJ, Hay WW Jr. Advice for selected breast-feeding issues. Compr Ther 1996;22:802-5.
50. American Academy of Pediatrics Section on Breastfeeding. Breastfeeding and the use of human milk. Pediatrics 2005;115:496-506.
51. Moritz ML, Manole MD, Bogen DL, et al. Breastfeeding-associated hypernatremia: are we missing the diagnosis? Pediatrics 2005;116:343-7.
52. Mikiel-Kostyra K, Mazur J. Hospital policies and their influence on newborn body weight. Acta Paediatr 1999;88:72-5.
53. Avoa A, Fischer PR. The influence of perinatal instruction about breast-feeding on neonatal weight loss. Pediatrics 1990;86:313-5.

Prevalence of overweight and obesity in the Netherlands in 2003 compared to 1980 and 1997

10

Katja van den Hurk[1,2], Paula van Dommelen[1], Stef van Buuren[1,3], Paul H. Verkerk[1] and Remy A. HiraSing[1,2]

[1]TNO Quality of Life, Leiden, The Netherlands
[2]EMGO Institute, VU Medical Center, Amsterdam, The Netherlands
[3]University of Utrecht, Utrecht, The Netherlands

Arch Dis Child 2007;92:992-5.

Abstract

Objective: To assess the prevalence of overweight and obesity in children living in the Netherlands and compare the findings with the Third and Fourth National Growth Studies carried out in 1980 and 1997, respectively.
Design and methods: Data were obtained from the child health care system. International cut-off points for body mass index (BMI) were used to determine overweight and obesity. Cases were weighted for ethnicity and municipality size in such a way that the sample matched the distribution in the general population. The LMS method was used to calculate the age-related distribution of BMI, and the prevalence was calculated from the fitted distribution.
Patients: Data on 90,071 children aged 4-16 years were routinely collected by 11 community health services during 2002-2004.
Results: On average, 14.5% of the boys and 17.5% of the girls were overweight (including obesity), which is a substantial increase since 1980 (boys 3.9%, girls 6.9%) and 1997 (boys 9.7%, girls 13.0%). Similarly, 2.6% of the boys and 3.3% of the girls aged 4-16 years were obese, which is much higher than in 1980 (boys 0.2%, girls 0.5%) and 1997 (boys 1.2%, girls 2.0%). At the age of 4, 12.3% of the boys and 16.2% of the girls were already overweight.
Conclusions: The prevalence of overweight and obesity in the Netherlands is still rising, and at an even faster rate than before. Evidence-based interventions are needed to counter the obesity epidemic, and there is an urgent need for pre-school intervention programs.

Introduction

Overweight is a rapidly growing global public health problem. Overweight and obesity increase the risk of early mortality and severe illnesses, such as heart and vascular diseases, diabetes and psychosocial problems.[1-5] Awareness of obesity in the Netherlands has increased, especially after the results of the Fourth National Growth Study in 1997 showed a substantial increase in the prevalence of overweight and obesity since 1980.[6,7] It is not known whether, and if so, how the situation has changed since 1997. The aim of this study was to determine the prevalence of overweight and obesity in 2003, and to compare the results to those of earlier studies carried out in 1980 and 1997.

Methods

The child health care system in the Netherlands routinely monitors the health of approximately 95% of all 0-19-year-old children living in the Netherlands.[8] Local community health services examine children aged 4-19 years, and a number of these

organizations maintain an electronic record for each child. We obtained from these records data on height, weight, age, sex, and postal code or municipality of residence. Since the child health care system uses standardized methods to measure children, we did not include data on children who were examined on indication.

The body mass index (BMI) of the children was calculated, and they were classified as normal (including underweight), overweight (including obesity) or obese, based on internationally accepted cut-off points.[9] The results are presented according to sex and age. Because the sample was not random, it contained a relatively high number of children from large cities and from Turkish and Moroccan ethnic minorities, populations which are known to have a higher prevalence of overweight.[10] The sample was therefore reweighed in such a way that the proportion of cases per combination of city size and ethnicity equaled that in the population of all children living in the Netherlands on January 1, 2003.[11] The LMS model of Cole and Green was used to fit the age-conditional distribution of BMI for all children living in the Netherlands separately for boys and girls.[12] P-splines were used to smooth the distribution over age, the calculations were carried out using the R-function GAMLSS[13] and the worm plot was used to assess the quality of the solution.[14] The prevalence of overweight and obesity according to age and sex was calculated from the fitted L, M and S curves.

The prevalence of overweight and obesity in 1997 for all children living in the Netherlands was calculated as a weighted average of the published Dutch, Moroccan and Turkish prevalences.[10] The weights used were 0.933, 0.031 and 0.036, respectively, which correspond to the percentage of Dutch, Moroccan and Turkish children aged 5-15 years living in the Netherlands on January 1, 2003.[11] The prevalence of overweight and obesity in 1980 was calculated from the L, M and S curves of BMI data from the Third Dutch Growth Study in 1980, as published by Cole and Roede.[15,16]

Results

Eleven community health services (31% of all the community health services in the Netherlands) provided routinely collected electronic data on height and weight. The total sample consisted of 90,071 children (approximately 3.8% of the child population) measured in the period 2002-2004. Table 1 lists the number of children per service, and their age.

Table 1 *Number of children measured during the period 2002-2004, according to Community Health Service and age (complete years lived).*

Health service	Age (year)													
	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
Den Haag	1418	4089	126	1288	1857	12	1	2	2	99	146	36	1	9077
Eemland	30	3074	2968	207	0	0	0	0	271	3084	2129	190	0	11953
Eindhoven	10	1108	604	0	0	6	906	746	1195	1212	147	8	0	5942
Fryslan	0	0	0	0	0	102	5269	4125	414	2	0	0	0	9912
Groningen	31	625	94	2	14	422	140	2	0	0	0	0	0	1330
Kennemerland	68	2740	1455	177	15	59	1992	1520	180	882	1845	479	59	11471
Nijmegen	62	1671	394	9	1	22	801	969	1169	966	96	3	0	6163
OZ Limburg	71	1476	680	27	176	2289	710	55	148	895	1192	227	26	7972
Utrecht	33	1950	1977	202	36	1266	1484	331	948	1549	389	36	2	10203
Zuid-Holland Noord	51	1663	627	40	31	100	595	353	56	797	1008	162	2	5485
Zuid-Holland Zuid	7	1024	1182	1365	1607	184	712	1057	167	1082	1781	380	15	10563
Total	1781	19420	10107	3317	3737	4462	12610	9160	4550	10568	8733	1521	105	90071

Figures 1 and 2 (boys and girls, respectively) show the prevalence of overweight and obesity in 2003. For comparison, the prevalence rates in 1980 and 1997 have also been plotted. More girls than boys were overweight and obese at nearly all ages and during all periods. In 1997, a peak occurred around the age of 6, and the prevalence was lower for older children. In 2003, the prevalence at the age of 6 was similar to the peak in 1997, but the peak in 2003 shifted towards the age of 8 (boys 18.7%, girls 24.4%). The differences in prevalence are fairly large during puberty. Children aged 6 in 1997 were approximately 12 years of age in 2003. By shifting the entire 1997 prevalence curves towards the right by 6 years, we can compare the prevalence within the same birth cohort at different time points. It appears that for nearly all birth cohorts the prevalence in 2003 is equal to or higher than that in 1997. The generation of children born around the year 1995 seems to be particularly at risk of developing overweight and obesity. Note that in 1997, this generation had about the same prevalence as the 2-year-olds in 1980. The increase must thus have occurred between the ages of 2 and 8.

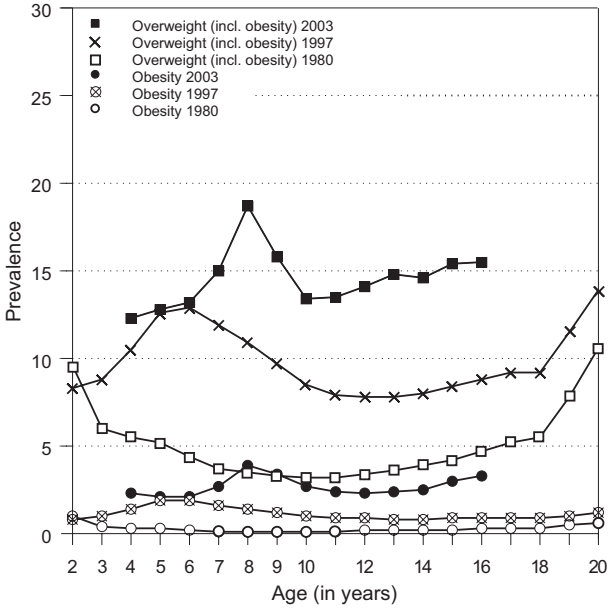


Figure 1 *Prevalence of overweight (incl. obesity) and obesity in boys living in the Netherlands, according to age (in 1980, 1997, and 2003).*

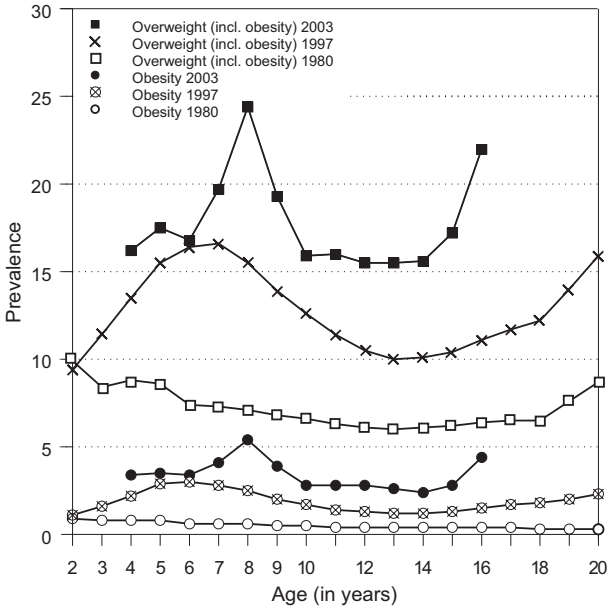


Figure 2 *Prevalence of overweight (incl. obesity) and obesity in girls living in the Netherlands, according to age (in 1980, 1997, and 2003).*

Table 2 shows the estimated prevalence rates in 1980, 1997 and 2003, with all ages combined. The prevalence of overweight and obesity rose between 1980 and 1997, and rose even faster between 1997 and 1980 (fig 3).

Table 2 *Prevalence of overweight (incl. obesity) and obesity in 4-15 year-old children living in the Netherlands in 1980, 1997, and 2003.*

	Boys						Girls					
	Overweight (incl obesity)			Obesity			Overweight (incl obesity)			Obesity		
Age	1980	1997	2003	1980	1997	2003	1980	1997	2003	1980	1997	2003
4	5.5	10.5	12.3	0.3	1.4	2.3	8.8	13.5	16.2	0.8	2.2	3.4
5	5.2	12.6	12.8	0.3	1.9	2.1	8.6	15.5	17.5	0.8	2.9	3.5
6	4.4	12.9	13.2	0.2	1.9	2.1	7.4	16.4	16.8	0.6	3.0	3.4
7	3.7	11.9	15.0	0.1	1.6	2.7	7.3	16.6	19.7	0.6	2.8	4.1
8	3.5	10.9	18.7	0.1	1.4	3.9	7.1	15.5	24.4	0.6	2.5	5.4
9	3.3	9.7	15.8	0.1	1.2	3.4	6.8	13.9	19.3	0.5	2.0	3.9
10	3.2	8.5	13.4	0.1	1.0	2.7	6.6	12.6	15.9	0.5	1.7	2.8
11	3.2	7.9	13.5	0.1	0.9	2.4	6.3	11.4	16.0	0.4	1.4	2.8
12	3.4	7.8	14.1	0.2	0.9	2.3	6.1	10.5	15.5	0.4	1.3	2.8
13	3.6	7.8	14.8	0.2	0.8	2.4	6.0	10.0	15.5	0.4	1.2	2.6
14	3.9	8.0	14.6	0.2	0.8	2.5	6.1	10.1	15.6	0.4	1.2	2.4
15	4.2	8.4	15.4	0.2	0.9	3.0	6.2	10.4	17.2	0.4	1.3	2.8

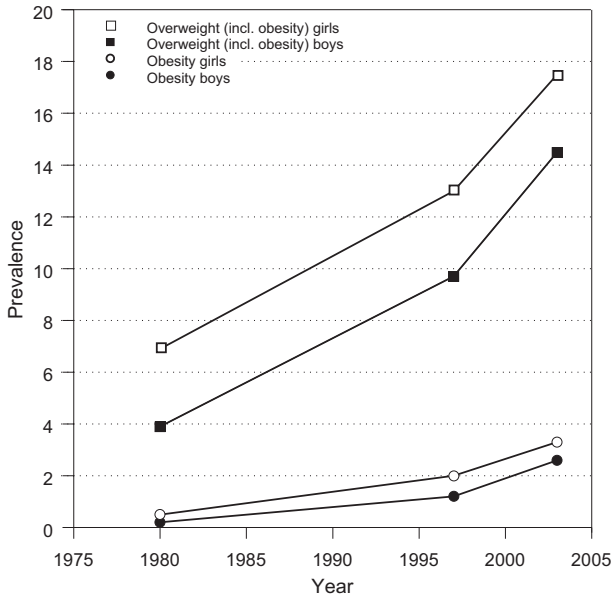


Figure 3 *Secular trend 1980-2003 in the prevalence of overweight (incl. obesity) and obesity in 4-15 year-old children living in the Netherlands.*

For boys, the rate of increase in the prevalence of overweight between 1980 and 1997 was approximately +0.34% per year, whereas between 1997 and 2003 it was +0.80% per year. For girls the rate of increase was +0.36% and +0.75% per year, respectively. With regard to obesity, the rate of increase for boys was +0.06% and +0.23% per year, and for girls it was +0.09% and +0.22% per year. The rate of increase generally doubled or tripled between 1997 and 2003.

Discussion

The 1997 study reported a large secular increase in the prevalence of overweight since 1980. The 2003 study indicates that this trend is continuing at an even faster rate. This finding is in line with results in other European countries.[17] In 2003, the prevalence of overweight in boys aged 4-16 years varied from 12.3% to 18.7% (average 14.5%), and in girls varied from 15.5 to 24.4% (average 17.5%). The prevalence of obesity was also higher among girls: 2.4% to 5.4% (average 3.3%) for girls compared to 2.1% to 3.9% (average 2.6%) for boys.

A particularly worrying aspect is that the dip in the prevalence after the age of 6 that was found in 1997 seems to have vanished in 2003. In 1997, two possible reasons were put forward to explain this: either the prevalence for 5-7-year-olds was higher because the cut-off points for these age-groups were somehow too low (methodological effect) or the cohort of children born since 1990 is structurally different (i.e., heavier) than previous cohorts (cohort effect).[7] The 2003 results point strongly towards the second explanation of a structural cohort effect. Given the present analysis, a methodological effect seems highly unlikely.

It is not known what causes the difference in prevalence between girls and boys, which already exists before they go to school. In order to prevent young girls and boys (up to the age of 4) from becoming overweight, there is an urgent need for pre-school intervention programs.

Two methods can be used to reliably compare the prevalence in regions of the Netherlands, or in other countries, with the prevalence curves presented here. The first option is to apply the methods used in the present study, including the LMS method, and to compare the resulting age-smoothed prevalence curves directly. The second option is to classify children as normal, overweight or obese, based on the internationally accepted cutoff points for BMI, and then calculate the prevalence and the 95% confidence intervals per age group. An overlap of the confidence interval with the Dutch prevalence at the relevant ages would indicate that the difference between the observed prevalence and the expected prevalence could be due to chance.

Conclusion

The global obesity epidemic is also occurring in the Netherlands, and evidence-based interventions are needed to halt the increase. The Dutch child health care system plays a vital role, because it examines all children living in the Netherlands and therefore provides good opportunities for intervention. Preschool intervention programs may be useful.

Acknowledgements

We would like to thank GGD Den Haag, GGD Eemland, GGD Eindhoven, GGD Fryslan, GGD Groningen, GGD Kennemerland, GGD Oostelijk Zuid-Limburg, GGD Regio Nijmegen, GG&GD Utrecht, GGD Zuid-Holland Noord and GGD Zuid-Holland Zuid for providing the data.

References

1. Fontaine KR, Redden DT, Wang C, et al. Years of life lost due to obesity. JAMA 2003;289:187-93.
2. Dietz WH. Health consequences of obesity in youth: childhood predictors of adult disease. Pediatrics 1998;101(3 Pt 2):518-25.
3. Wabitsch M. Overweight and obesity in European children: definition and diagnostic procedures, risk factors and consequences for later health outcome. Eur J Pediatr 2000;159(Suppl 1):S8-13.
4. International Obesity Task Force. Obesity in Europe: the case for action. London: IOTF, 2002.
5. Renders CM, Delemarre-van de Waal HA, Dekker JM, et al. Insulin resistance and diabetes type 2 in overweight children. Ned Tijdschr Geneeskd 2002;147:2060-3.
6. Fredriks AM, van Buuren S, Wit JM, et al. Body index measurements in 1996-7 compared with 1980. Arch Dis Child 2000;82:107-12.
7. Hirasing RA, Fredriks AM, van Buuren S, et al. Toegenomen prevalentie van overgewicht en obesitas bij Nederlandse kinderen en signalering daarvan aan de hand van internationale normen en nieuwe referentiediagrammen. Ned Tijdschr Geneeskd 2001;145:1303-8.
8. Burgmeijer RJF, van Geenhuizen YM, Filedt Kok-Weimar T, et al. Op weg naar volwassenheid. Evaluatie jeugdgezondheidszorg 1996. Leiden: TNO Preventie en Gezondheid/KPMG, 1997.
9. Cole TJ, Bellizzi MC, Flegal KM, et al. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ 2000;320:1240-3.

10. Fredriks AM, van Buuren S, HiraSing RA, et al. Alarming prevalences of overweight and obesity for children of Turkish and Moroccan origin, according to the international standard. Acta Paediatr 2005;94:496-8.
11. Netherlands Bureau of Statistics. Statline. Available from www.cbs.nl (accessed 8 August 2007).
12. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med 1992;11:1305-19.
13. Rigby RA, Stasinopolous DM. Generalized additive models for location, scale and shape (with discussion). Appl Stat 2005;54:507-54.
14. Van Buuren S, Fredriks AM.Worm plot: a simple diagnostic device for modeling growth reference curves. Stat Med 2001;20:1259-77.
15. Cole TJ, Roede MJ. 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 1999;26:303-8.
16. Roede MJ, van Wieringen JC. Growth diagrams 1980: Netherlands third nationwide survey. Tijdschr Soc Gezondheidsz 1985;63S:1-34.
17. Jackson-Leach R, Lobstein T. Estimated burden of paediatric obesity and comorbidities in Europe. Part 1. The increase in the prevalence of child obesity in Europe is itself increasing. Int J Pediatr Obes 2006;1:26-32.

Impact of height bias on overweight prevalence in childhood

11

Paula van Dommelen[1] and Stef van Buuren[1,2]

[1]TNO Quality of Life, Leiden, The Netherlands
[2]University of Utrecht, Utrecht, The Netherlands

Submitted to Obesity

Abstract

There are several definitions for overweight based on Body Mass Index (BMI). It is known that, according to BMI, tall children are proportionally more overweight than short children. We studied the impact of height bias on (inter)national comparisons of overweight prevalence in childhood, and evaluate an alternative overweight definition. Two overweight measures were studied: (a) BMI in kg/m^2 ($\geq$ IOTF cut-off points) and (b) $(WSDS-\rho(HSDS))/\sqrt{(1-\rho^2)}$ (≥ 1.4) where WSDS and HSDS are the standard deviation score for weight (W) and height (H) respectively, and where ρ is their correlation. Data from three Dutch nation-wide surveys were used. For measure (a) tall boys aged 4-14 years had a significantly higher prevalence of overweight than short boys (OR=4.6). Likewise, tall girls aged 4-12 years had a significantly higher prevalence of overweight than short girls (OR=3.8). No such differences were found for measure (b). If children from the United States (US) were as tall as the Dutch children, the US prevalence of overweight according to measure (a) would increase by 6.9 percent points. Reversely, if the Dutch children were as short as the US children, the Dutch prevalence of overweight would decrease by 2.9 percent points. We conclude that overweight definitions based on BMI exhibit a strong height bias in childhood overweight prevalence. BMI does not properly adjust weight for height in children. The use of BMI to compare childhood prevalence of overweight is complicated if the populations differ in height. We suggest using an alternative definition of overweight that is insensitive to height bias.

Introduction

Overweight is a rapidly growing global public health problem. Overweight and obesity increase the risk of morbidity and early mortality.[1,2] Assessment of overweight should be based on precise measurements of lean mass (i.e. fat-free body mass) and fat mass. Techniques that can be applied to predict body fatness are hydrostatic weighing, x-ray, total body electrical conductivity or bioimpedance (BIA), near-infrared interactance (NIR), ultrasound, computer tomography (CT), air displacement method (BOD POD), magnetic resonance imaging (MRI) and skinfold thickness measurements.[2] However, these methods are not always readily available, and they are either expensive or need highly trained personnel.

Indices based on body weight and height are often used. A good index should be highly correlated both to weight and body fatness, and it should be independent of height. A popular index is body mass index (BMI), calculated as weight in kilograms divided by height in meters squared. There are two widely accepted international definitions for overweight and obesity in childhood for BMI. The first definition uses BMI centiles linked to adult cut-off points. A cut-off point of 25 kg/m^2 is accepted as a definition

of adult overweight, and 30 kg/m^2 as adult obesity. The extension into childhood was proposed by the International Obesity Task Force (IOTF).[3] The second definition is based on the 85th (overweight) and 95th (obesity) centiles of BMI on a nationally representative survey in the United States (US). This definition was developed by the Centers of Disease Control and prevention (CDC).[4] Both definitions are recommended to be used in international comparisons of prevalence of overweight and obesity.

It is known that, according to BMI, tall children are proportionally more overweight than short children.[5,6] Mulligan and Voss compared BMI, weight and height centiles in hypothetical fat and thin children aged 2-9 years.[5] They found that, in fat children, weight exceeded height by four to five centiles (each 0.67 SD) if they were short, but by only one centile if tall. Conversely in thin children, weight was about two centiles below height in short children, but four centiles below in tall children.

Cole attempted to find values of p in a weight/height^p index not correlated with height.[7] He modeled the changing relationship between BMI and age by fitting a Benn-type index where the power of height (p) varied systematically with age, starting at 2 soon after birth then rising steadily to 3 in mid-puberty and then falling to 2 or less in adulthood. The patterns varied slightly by sex. Part of the rise could be attributed to differences in the timing of puberty. Using an appropriate age-specific value for p strongly reduces the correlation between weight/height^p and height.[8]

An alternative measure uses the standard deviation score (SDS).[9,10] The SDS expresses the measurement relative to a reference population in units of standard deviations above or below the median.[11] A measure that predicts the BMI centile from his weight (W) and height (H) centiles was defined as BMI SDS = 1.434(WSDS)-0.794(HSDS) for children from early life to 18 years.[10] As the BMI-weight-height relationship depends on age and sex, the equation can be improved. A measure, which we call the "Weight for Height Estimate" (WHE), that is more flexible and sensitive to age and sex changes in the weight-height relationship is given by WHE = (WSDS-ρ(HSDS))/$\sqrt{(1-\rho^2)}$, with ρ the age-sex-specific correlation between HSDS and WSDS.[9] Theoretically, this measure should remove any bias in height, age and sex.

Some authors have offered biological explanations for the height bias in BMI.[6,12-14] Buchan *et al*. showed that over a period of 16 years, the secular increase in fatness seen in 3 year old children was much greater in the tall than the short children.[12] Freedman *et al*. showed that height was not only associated with BMI in 5-18 year old children, but also demonstrated similar correlations with the skinfold sum and with percentage body fat. In a related study, the authors argued that taller children are more likely to be obese in adulthood.[14] One may argue that BMI, which preferentially classifies taller young children as overweight, is appropriate because height and adiposity are correlated before the age of 12 years.[6,13]

In this paper, we will elucidate a mechanism that can cause height bias in BMI, and assess its impact on overweight prevalence in childhood. Two definitions for overweight are used: (a) BMI ≥ IOTF cut-off points and (b) WHE ≥ 1.4.

Methods and procedures

Subjects

Individual height and weight data were obtained from three Dutch nation-wide surveys performed in 1980, 1997 and 2003.[15-17] The first survey consists of 41,805 Dutch children aged 0-21 years collected in 1980 (DS1980). The second survey sampled 14,500 Dutch children aged 0-21 years in the year 1997 (DS1997). The third survey contains information on 89,966 children aged 4-15 years and living in the Netherlands (Dutch and other ethnicities) (DS2003). The 1980 and 1997 surveys are representative for the Dutch children. The 2003 survey is not a random sample.

Statistical analyses

For the measures BMI and WHE, we calculated age-sex-specific prevalence estimates of overweight and obesity for children with a short stature (< -1 HSDS), an average stature (-1 to 1 HSDS) and a tall stature (>1 HSDS). Dutch references (1997) were used to calculate HSDS and WSDS. The age-sex-specific correlation (ρ) between HSDS and WSDS was calculated for year classes from data from all three surveys. In this study we chose a cut-off level of 1.4 for WHE, to let the prevalence estimates correspond to a BMI SDS of +2.[10,18]

The odds ratio (OR) of overweight in tall versus average and short children was estimated by logistic regression analyses. Simulation was used to investigate the impact of population height on international comparison of overweight prevalence. All statistical analyses were performed using S-PLUS 7.0 for Windows and SPSS 14.0 for Windows.

The mechanism behind height bias in BMI

By definition, HSDS and WSDS are standard normally distributed in the reference population. If the regression of HSDS on WSDS (and vice versa) is linear then the joint distribution of HSDS and WSDS is bivariate normally distributed.[19] The bivariate normal distribution of HSDS and WSDS describes the proportion of cases at each combination of HSDS and WSDS. We can find all points with the same proportion by taking slices parallel to the bottom plane. The boundaries of such a slice correspond to the equal probability contour. For example, the bivariate 90% confidence interval corresponds to the contour in which 90% of the children are located. For a bivariate normal distribution, contours correspond to ellipses. The shape of the ellipses only depends on the correlation between HSDS and WSDS. As we will show below, the IOTF cut-off points for BMI can be drawn on the HSDS-WSDS plane at a given age. The position of the IOTF

curves relative to the confidence ellipses show that tall children are proportionally more overweight than short children.

Results

Figure 1 plots the (Pearson) correlations between HSDS and WSDS of the three Dutch surveys for boys and girls against age. This figure shows that the correlation between age 2 to 11 years is fairly constant at approximately 0.73. We see a linear decrease of about 0.049 per year for girls older than 11 years, and 0.093 for boys older than 15 years. Notice that the correlation is slightly lower in the 2003 survey.

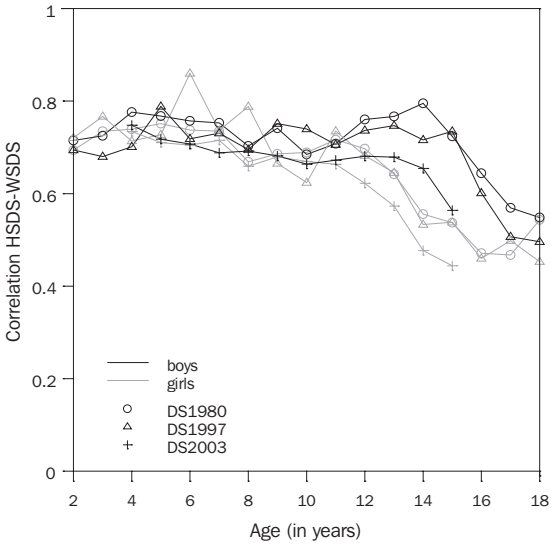


Figure 1 *Pearson correlation between HSDS and WSDS of the three Dutch surveys for boys and girls against age.*

Visualization of the height bias in BMI

There is a methodological explanation of height bias. Figure 2 shows two ellipses that define the area in which 90% and 99% percent of the children are located. This figure plots the case for children of age 6 years, with $\rho = 0.73$. We also plotted the IOTF cut-offs as curves (for boys) onto the same plane by calculating the (HSDS, WSDS) coordinates that correspond to the relevant BMI cut-offs. The crucial observation we make from this

plot is as follows: At -1 HSDS, the IOTF obesity line is close to the 99% ellipse, while at +1 HSDS the IOTF obesity line is close to the 90% ellipse. Therefore, the amount of obesity for tall children is larger than for short children when using BMI. The same holds for the overweight prevalence.

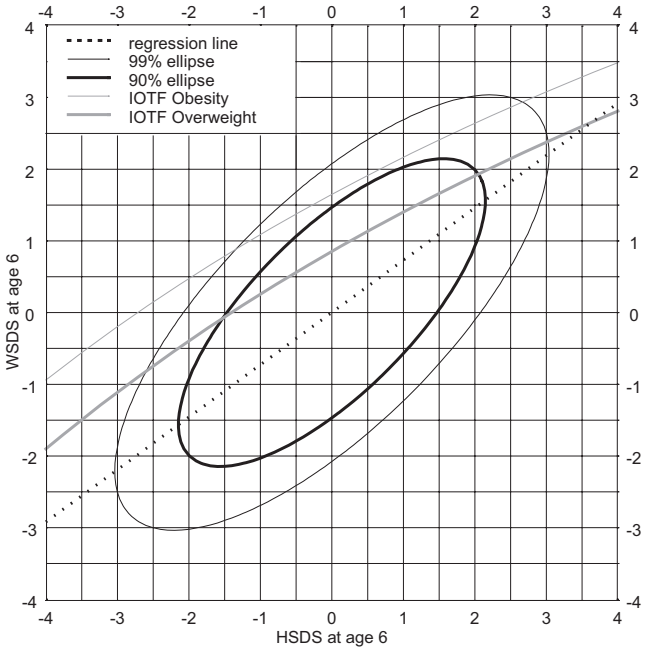


Figure 2 *Two ellipses (with 90% and 99% confidence levels) of the HSDS - WSDS distribution at 6 years of age together with the IOTF cut-off points. Using BMI results in an overestimate of overweight prevalence in tall 6-year olds.*

In contrast, Figure 3 shows the same ellipses for the 18 years olds where $\rho = 0.73-(18-15)*0.093) = 0.45$. At -1 HSDS, the IOTF obesity line is inside the 99% ellipse, while it crosses the 99% ellipse at +1 HSDS. Thus, the prevalence of obesity in tall children is slightly lower than in short children, i.e. the opposite from Figure 2. Notice that in Figure 2, the IOTF line is not parallel to the regression line. In Figure 3, the height bias has almost disappeared, and even seems slightly reversed.

The correlation between HSDS and WSDS is the only factor that differs between Figure 2 and Figure 3. Therefore, the correlation between weight and height determines the size of the height bias. Since this correlation is not constant with age (see Figure 1), the implication is that the IOTF cut-off points (or for that matter, any cut-offs based on BMI) are sensitive to different height biases at different ages.

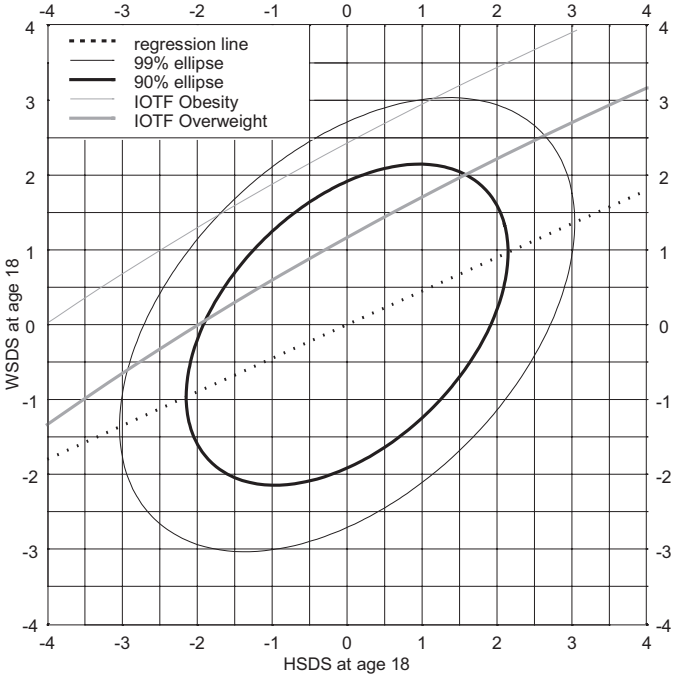


Figure 3 *Two ellipses (with 90% and 99% confidence levels) of the HSDS - WSDS distribution at 18 years of age together with the IOTF cut-off points.*

Empirical evidence of the existence of height bias

Table 1 shows the prevalence of overweight (including obesity) stratified by height (short, average and tall stature), age and sex for the 1997 survey. The strong height bias in BMI in Table 1 is visible as differences between the short, average and tall groups. For example, 7.2% of all short girls (< -1 HSDS) aged 4-7 years has overweight, whereas the overweight prevalence for tall girls (> +1 HSDS) of the same age is 27.5%. When ages are combined, the OR's (95% CI) for the BMI definition is 4.6 (2.6-8.2) for boys aged 4-14 years, and 3.8 (2.3-6.4) for girls aged 4-12 years. In contrast, for the WHE definition, we find 1.2 (0.7-1.8) for the boys and 0.6 (0.4-1.1) for the girls. As predicted by Figure 3 an inverse relation occurs at the age of 18 years. The prevalence of overweight according to BMI was lower for the tall children with OR= 0.2 (0.1-0.7), but not statistically significant for WHE. Table 1 empirically shows the existence of a height bias in overweight prevalence in BMI. Similar results were obtained for the 1980 and 2003 surveys (not shown).

Table 1 *The prevalence of overweight (including obesity) stratified by height (short (S), average (A) and tall (T) stature), age and sex for the DS1997.*

Sex	Age	n	BMI ≥ IOTF cut-off points			WHE ≥ 1.4		
			S	A	T	S	A	T
Boys	2	450	3.2	6.6	11.9	4.8	8.6	5.2
	3	294	4.1	6.1	10.6	6.1	9.1	8.5
	4-7	425	1.5	9.1	19.0*	3.1	8.1	14.3*
	8-9	444	5.3	9.4	13.0	8.8	10.1	5.8
	10	349	1.7	6.3	15.4*	3.3	7.6	15.4*
	11	366	1.9	8.7	15.5*	11.1	12.2	10.3
	12	380	3.4	7.8	9.5	10.2	10.5	6.3
	13	430	7.8	6.6	13.8	11.7	9.5	10.0
	14	414	1.3	9.2*	10.8*	7.7	13.0	4.1
	15	406	5.7	4.4	9.7	10.0	6.6	9.7
	16	354	4.1	8.9	7.1	4.1	10.2	5.7
	17	350	12.5	7.1	9.3	12.5	7.1	9.3
	18	333	12.8	9.7	4.0	10.6	8.5	4.0
Girls	2	447	5.8	6.6	12.2	5.8	6.6	8.1
	3	313	6.8	9.7	17.0	9.1	6.0	11.3
	4-7	423	7.2	10.2	27.5**	7.2	7.4	5.8
	8-9	428	4.8	14.6*	26.2**	8.1	10.0	7.7
	10	365	10.3	13.7	16.7	16.2	13.7	8.3
	11	365	6.3	8.3	25.8**	9.5	9.2	8.1
	12	394	4.4	14.3*	12.3	11.8	11.8	6.2
	13	469	3.2	11.1	10.8	8.1	10.5	6.0
	14	389	5.4	7.8	5.9	8.9	8.2	3.9
	15	402	11.8	9.8	9.4	9.8	9.1	6.3
	16	241	13.3	6.9	11.1	10.0	5.7	8.3
	17	183	14.3	7.3	12.5	5.7	6.5	8.3
	18	214	29.0	10.1*	5.7**	16.1	6.8	5.7

S: < -1 HSDS A:-1 to 1 HSDS T: >1 HSDS. Logistic regression with < -1 HSDS as reference group: **p<0.01 *p<0.05

Impact on international comparison of overweight prevalence

To investigate the impact of height bias in international comparisons of overweight prevalence, we conducted two simulation experiments. The first experiment mimicked the case where US reference[20] children aged 4-11 years were as tall as the Dutch reference[15] children. We first draw a random sample of 2000 children (125 children for each year and sex) from the bivariate standard normal distribution of HSDS and WSDS with a correlation of 0.73. Assuming that HSDS and WSDS were expressed with respect to the US references, we calculated the proportion of children above or at the IOTF overweight cut-off points. The result was equal to 13.8%. We then changed the mean of the random sample from (0,0) to (+0.73, +0.53). The value 0.73 corresponds to the height difference between the US and Dutch 4-11 year olds (4.42 cm = 0.73 US HSDS). The value 0.53 is the expected value of WSDS = ρ*expected(HSDS) = 0.73×0.73, where ρ is 0.73 and where the expected value of HSDS is (by coincidence) also 0.73. Thus, 0.53 corresponds to the difference in WSDS expected for a hypothetically taller

US population. The prevalence of overweight (including obesity) was then estimated as 20.7%. Thus, had the US population of 4-11 year olds been as tall as the Dutch, the overweight prevalence according to the IOTF cut-off points would be 20.7% instead of 13.8%.

The second experiment modeled the reverse situation. What would happen if Dutch 4-11 year olds were as short as US children? Assuming that HSDS and WSDS were expressed with respect to the Dutch references, we found an overweight prevalence (incl. obesity) equal to 10.2%. We changed the mean to (-0.77, -0.56) since -4.42 cm = -0.77 Dutch HSDS and -0.56=0.73 x -0.77. The estimated prevalence is now 7.3%. Thus, had the Dutch population of 4-11 year olds been as short as the US, the overweight prevalence according to the IOTF cut-off points would be 7.3% instead of 10.2%.

Discussion

BMI does not properly adjust weight for height in children. Criteria for overweight and obesity that are based on BMI will exhibit height bias, because the correlation between weight and height changes during childhood. BMI-based definitions overestimate overweight and obesity prevalence in tall children, and underestimate prevalence in short children (4-12/14 years). We found that the height bias present in the IOTF cut-off points for overweight was quite strong in terms of differences in overweight prevalence between short and tall children. The height bias may be partially explained by biological mechanisms, but the observed height bias is also related to peculiar behavior of BMI as a measure for overweight in childhood. The existence of height bias in BMI-based measures can distort international comparison over childhood overweight prevalence if populations differ in height.

Height bias in BMI can confound the comparison of prevalence if one population is taller than the other. We suggest using WHE to remove the impact of height bias on the comparison of prevalence estimates. Applying WHE to define overweight or obesity (as well as underweight) requires three steps:

1. choice of the correlations ρ between HSDS and WSDS stratified by age and sex,
2. choice of a reference standard used to calculate HSDS and WSDS,
3. choice of the cut-off level ρ to define the appropriate level.

With respect to choice 1, we suspect that the pattern of ρ will be similar across different populations. One study showed that at the age of three years, the weight-height correlation rose from 0.59 to 0.71 between 1988 and 2003.[12] In our study, however, the weight-height correlation was almost constant between 1980 and 2003. The correlation was slightly lower in the 2003 survey, but this is probably due to the fact that the

measurements in the 2003 survey were noisier. A sensible choice for a reference standard would be the population-specific or the global WHO anthropometric references for height and weight.[21] A reasonable choice for ρ involves the probability levels that match well-accepted BMI categories (16, 17, 18.5, 25, 30, 35, 40 kg/m^2) at the age of 18 years for the selected standard. This amounts to the same strategy as was employed by the IOTF group. Note that for a fixed ρ the prevalence estimates of overweight and obesity are assumed to be the same over all ages in the chosen reference population.

References

1. James WP. The epidemiology of obesity: the size of the problem. J Intern Med 2008;263:336-52.
2. McArdle WD, Katch FI, Katch VL. Exercise Physiology: Energy, Nutrition and Human Performance. Lippingcott Williams and Wilkins: Philadelphia, 2006.
3. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ 2000;320:1240-3.
4. Barlow SE, Dietz WH. Obesity evaluation and treatment: Expert Committee recommendations. The Maternal and Child Health Bureau, Health Resources and Services Administration and the Department of Health and Human Services. Pediatrics 1998;102:E29.
5. Mulligan J, Voss LD. Identifying very fat and very thin children: test of criterion standards for screening test. BMJ 1999;319:1103-4
6. Freedman DS, Khan LK, Serdula MK, Dietz WH, Srinivasan SR, Berenson GS. Inter-relationships among childhood BMI, childhood height, and adult obesity: the Bogalusa Heart Study. Int J Obes Relat Metab Disord 2004;28:10-6.
7. Cole TJ. Weight/heightp compared to weight/height2 for assessing adiposity in childhood: influence of age and bone age on p during puberty. Ann Hum Biol 1986;13:433-51.
8. Fung KP, Lee J, Lau SP, Chow OK, Wong TW, Davis DP. Properties and clinical implications of body mass indices. Arch Dis Child 1990;65:516-9.
9. Cole TJ. A new index of child weight-for-height based on weight and height Z scores [abstract]. Ann Hum Biol 1994**;**21:96.
10. Cole TJ. A chart to link child centiles of body mass index, weight and height. Eur J Clin Nutr 2002;56:1194-9.
11. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med 1992;11:1305-19.

12. Buchan IE, Bundred PE, Kitchiner DJ, Cole, TJ. Body mass index has risen more steeply in tall than in short 3-year olds: serial cross-sectional surveys 1988-2003. Int J Obes 2007;31:23-9.
13. Freedman DS, Thornton JC, Mei Z, Wang J, Dietz WH, Pierson RN Jr et al. Height and adiposity among children. Obes Res 2004;12:846-53.
14. Freedman DS, Khan LK, Mei Z, Dietz WH, Srinivasan SR, Berenson GS. Relation of childhood height to obesity among adults: the Bogalusa Heart Study. Pediatrics 2002;109:E23.
15. Fredriks AM, van Buuren S, Wit JM, Verloove-Vanhorick SP. Body index measurements in 1996-7 compared with 1980. Arch Dis Child 2000;82:107-12.
16. Roede MJ, van Wieringen JC. Growth diagrams 1980; Netherlands third nationwide survey. Tijdschrift voor Sociale Gezondheidszorg 1985;63(suppl.).
17. van den Hurk K, van Dommelen P, van Buuren S, Verkerk PH, Hirasing RA. Prevalence of overweight and obesity in the Netherlands in 2003 compared to 1980 and 1997. Arch Dis Child 2007;92:992-5.
18. de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. Bulletin of the World Health Organization 2007;85:660–7.
19. Arnold BC, Castillo E, Sarabia JM. Conditional specification of statistical models. Springer: Berlin, 1999.
20. Ogden CL, Kuczmarski RJ, Flegal KM, Mei Z, Guo S, Wei R et al. Centers for Disease Control and Prevention 2000 growth charts for the United States: improvements to the 1977 National Center for Health Statistics version. Pediatrics 2002;109:45-60.
21. WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards based on length/height, weight and age. Acta Paediatr 2006;450:76–85.

Summary and recommendations

12

The aim of this thesis was to answer the research question ***"What is the validity of referral criteria in growth monitoring?"*** First, we summarize the results of this thesis. Second we discuss the impact of the results on practice and consequences for future research.

Summary of results

Chapter 2

Growth charts are needed to assess whether the growth pattern of a child deviates from that of the reference population. Growth charts for twins were not available, while it is known that the growth pattern of twins differs from that of singletons during infancy. In chapter 2 we established growth charts for monozygotic and dizygotic twins aged 0-2.5 years. In the first six months of life, length and weight of twins were approximately -1.2 SDS below the median of the Dutch reference population, while this was -0.6 SDS for BMI. Approximately half of this growth retardation was attributable to gestational age. Twins grow faster after birth, but do not yet achieve the same height and weight at age 2.5 years.

Chapter 3

Chapter 3 shows that additive genetic factors, environmental factors shared by a twin pair, environmental factors unique to each twin individual and gestational age affect different parameters of a growth curve from birth to 2.5 years of age. Genetic and common environmental factors affecting different growth parameters were also correlated over time, indicating that some genes and shared environmental factors are expressed during early growth. Variation in size at birth was mostly explained by gestational age. Size at one (particularly for boys) and two years of age as well as weight velocity at one year of age were more influenced by heritability than common environmental factors. However, environmental factors affecting height and weight during the growth period were also important.

Chapters 4-9

In the chapters 4 to 9 we studied longitudinal anthropometric measurements from patients with Turner's syndrome, short stature due to pathology, cystic fibrosis, celiac disease, hypernatraemic dehydration, as well as children from the general population. Our outcome measures were sensitivity, specificity and age of referral for referral criteria that were of practical value for child health care programs. We established referral criteria based upon short stature, failure to thrive and body mass index (BMI).

Short stature

In chapters 4 and 6 we showed that validated referral criteria can be formulated and we proposed a guideline for short stature. The guideline for short stature warrants referral if a child is 0-3 years old and:

- he/she has an extreme short stature (<-3 SDS) and birth weight ≥ 2500 g,
- he/she has a very short stature (<-2.5 SDS) at least two times over a time period of six months to one year and birth weight ≥ 2500 g. With these criteria, 15% of the girls with Turner's syndrome can be detected, at the account of less than 1% false-positives.

Furthermore, a 3-10 year old child should be referred if:

- he/she has a very short stature (<-2.5 SDS),
- he/she has a short stature (<-2 SDS) in combination with a large distance to target height (> 2 SD) or a height deflection > 1 SD. These (two) rules were able to detect 86% of the girls with Turner's syndrome, 77% of the children with short stature due to pathology and 27% of the children with celiac disease at the account of less than 2% false-positives.
- he/she has a short stature (<-2 SDS) and medical history reveals that birth weight and/or length was low, then the diagnosis of persistent short stature after small for gestational age can be made.

In the case of an unusual growth pattern, certainly if associated with clinical symptoms or signs (such as disproportion and/or dysmorphic features, emotional deprivation), even if it did not comply with the rules for referral or the recommendations, doctors should still be free to follow their clinical judgment.

Chapter 5 shows that analysis of multiple measurements over time forms a substantial improvement in diagnostic performance over a single measurement. We summarized longitudinal height measurements of children with Turner's syndrome and children from the general population from birth to five and eight years of age by parameters of the Jenss-Bayley growth curve. A model was established by the parameters of the growth curve, parental height and gestational age. This model explains the grouping of the reference and children with Turner's syndrome. With this model sensitivity is about 11 to 30% higher at the same level of specificity compared to conventional rules.

Failure to thrive

Chapter 9 contains a growth chart for relative weight change (RWC) for healthy, breast-fed infants between postnatal days 2 and 11. We proposed a guideline for the detection of neonates with hypernatraemic dehydration using the RWC growth chart. This guideline suggests intervention when weight loss exceeds 10% in the first week of life and referral to a hospital when weight loss exceeds approximately 12% in the first week (< -2.5 SDS) or 10% after the first week. In total, 90% of the children with hypernatraemic dehydration can be detected at the account of 2% percent false-positives by this guideline.

Body Mass Index

Chapter 7 and 8 show that the most sensitive growth parameter to detect children aged 0-2.5 years with symptomatic celiac disease or with cystic fibrosis is a decrease in BMI. BMI is a better predictor than weight or length. Thirty percent of the symptomatic celiac disease children and 2% of the reference children had a BMI SDS less than -1.5 and a decrease in BMI SDS of at least -2.5. Children with celiac disease detected by screening grow normally between birth and 2.5 years of age. Furthermore, 32% of the cystic fibrosis children and 1.7% of the reference children had a BMI SDS less than -2.5 and a decrease in BMI SDS of at least -0.5.

Chapter 10

In chapter 10, the prevalence of childhood overweight and obesity was calculated with IOTF cut-off points. In the Netherlands, the prevalence of childhood overweight and obesity rose between the year 1980 and 1997. Chapter 10 shows that the prevalence rose even faster between 1997 and 2003. On average, 14.5% of the boys and 17.5% of the girls were overweight in 2003. More girls than boys were overweight and obese at all ages.

Chapter 11

Chapter 11 shows that BMI has a height bias in childhood at different ages. Tall children are proportionally more overweight than short children. The correlation between height and weight varied with age and closely corresponded to this height bias. The odds ratio of overweight in tall boys aged 4-14 years compared to short boys is 4.6. This is 3.8 for tall girls aged 4-12 years of age compared to short girls. The following "Weight for Height Estimate" (WHE) can be used if one wants to remove the height bias: WHE = $(WSDS-\rho(HSDS))/\sqrt{(1-\rho^2)}$ with ρ the age-sex-specific correlation between height SDS and weight SDS.

This is the first evidence-based investigation of referral criteria in growth monitoring. The results of the chapters indicate that the validity is high. Proper growth monitoring leads to early identification of children with conditions.

Recommendations

- **Growth monitoring should be evaluated along all UK national screening committee (NSC) quality criteria.**[1] This thesis has focused on the properties of growth monitoring as a screening test. We recommend further investigation of the other elements of the NSC quality criteria, i.e. the epidemiology of the condition, any treatment options, and the acceptability of the screening program.
- **The determination of optimal ages for growth monitoring is needed.** An important element of the NSC quality criteria is that the screening program should be cost-effective. It is, therefore, important to investigate how often and when a child should be measured. Less than one-third of growth monitoring programs assess the growth of children beyond six year of age.[2] It has been predicted that referral based on a single height measurement ($<0.4^{th}$ centile) at 5-year olds would miss 1:7,143 girls with Turner's syndrome and 1:20,000 cases with growth hormone deficiency.[3] No numbers were presented for other conditions. This thesis shows that involving multiple measurements (≥ 3) over time form a substantial improvement in diagnostic performance over a single measurement in girls with Turner's syndrome.
- **Further research is needed to assess the validity of referral criteria in other growth-related conditions, like growth hormone deficiency.** While a great diversity of disorders and adverse conditions can cause growth impairment, only hypernatraemic dehydration, Turner's syndrome, celiac disease, cystic fibrosis and obesity were taken into account as these are the largest groups to be detected next to growth hormone deficiency. Growth hormone deficiency is heterogeneous in its clinical presentation[4], and it is difficult to obtain sensitivity and specificity without having a gold standard. However, we recommend further research to investigate referral criteria in this and other groups with conditions that manifest themselves through an abnormal growth.
- **The performance of referral criteria based on other methodology should be evaluated.** Most referral criteria in this thesis use SDS based on a general population as an indication for referral. However, as mentioned in chapter 1, there are also syndrome-specific growth charts available. With these growth charts, one can evaluate the probability of a specific disorder given one or more anthropometric measurements. This can be compared by the probability of being healthy given one or more anthropometric measurements. The likelihood ratio statistic tells us how much more likely the observation is in the disorder than in the reference group. Theoretical and empirical evidence exists that screening rules based on likelihood ratio statistic discriminate better than rules based on SDS based on a general population.[5] The likelihood ratio statistic can only be calculated when syndrome-specific growth charts are available.

- **Growth charts are needed for specific groups of children.** For example, syndrome-specific growth charts, or growth charts for South Asian children living in the Netherlands. The validity of growth monitoring may increase when using growth charts for specific groups of children, because children with a growth disorder of a relatively tall subpopulation are more likely to be missed when compared with a non-specific reference, while children without a growth disorder from a relatively short subpopulation would be referred too often.[6]
- **The collection of parental heights, especially in short children, should be a routine part of growth monitoring.** The determination of parental heights is essential in the guideline for short stature. In this thesis, we used reported parental heights instead of measured. As a discrepancy between measured and reported heights has been noticed, the validity of the screening test with parental heights may be higher using measured heights instead of reported heights.[7] Men tend to overestimate and women tend to underestimate their height. The difference between reported and measured height is positively associated with age, and there is a wide individual variation between reported and measured heights in both sexes.[8,9] Furthermore, one needs a formula that calculates the target height (TH). Several formulas were put forward that were all based on the mid-parental height (MPH), which is the average of the heights of the two parents. A correction for sex, secular trend and regression to the mean was introduced to the MPH. More research has to be done to estimate the diagnostic performance of these definitions for TH in growth monitoring.
- **Measurements should be accurate.** Accurate measurements include a standardized measurement technique, quality equipment which is regularly calibrated and accurate, and trained measurers.[10,11] Information on the appropriate equipment and techniques for accurate weighing and measuring of infants, children, and adolescents are needed and is available from the International Organization for Standardization; a worldwide federation of national standards bodies.[12]
- **Evidence-based guidelines should be introduced to the user.** Once the evidence-based guidelines are developed, the next step is to introduce them to the users: people working in the primary care midwife practices or the child health care centers, depending on the guideline. The introduction of innovations, such as guidelines, is recognised as a complex process. In general, four main stages are distinguished in innovation processes.[13-15] These stages are: 1) dissemination, 2) adoption, 3) implementation and 4) continuation. The transition from one stage to the next can be affected by various determinants.[14] Before introducing a guideline, it is essential to identify the most relevant determinants and accommodate these in the innovation strategy.[13,16]
- **The guidelines should be implemented and accessible in a computer system.** A guideline implemented in a computer system has several advantages. It can save time

(the computer can perform calculations), it increases validity of detecting conditions (by modeling growth patterns), it eliminates error, and each record is continuously updated and accessible for colleagues. In 2009, the Electronic Child Record (EKD) will be implemented in the Netherlands. The EKD is a medical file for children used in an automated system in child health care.

References

1. Gray JA. New concepts in screening. Br J Gen Pract 2004;54:292-8.
2. Fayter D, Nixon J, Hartley S, Rithalia A, Butler G, Rudolf M, Glasziou P, Bland M, Stirk L, Westwood M. Effectiveness and cost-effectiveness of height screening programmes during the primary school years: a systematic review. Arch Dis Child 2008;93:278-84.
3. de Onis M. The use of anthropometry in the prevention of childhood overweight and obesity. Int J Obes Relat Metab Disord 2004;28 Suppl 3:S81-5.
4. Pinto G, Adan L, Souberbielle JC, Thalassinos C, Brunelle F, Brauner R. Idiopathic growth hormone deficiency: presentation, diagnostic and treatment during childhood. Ann Endocrinol (Paris) 1999;60:224-31.
5. van Buuren S. Improved accuracy when screening for human growth disorders by likelihood ratios. Stat Med 2008;27:1527-38.
6. Chinn S, Cole TJ, Preece MA, Rona RJ. Growth charts for ethnic populations in UK. Lancet 1996;347:839-40.
7. Braziuniene I, Wilson TA, Lane AH. Accuracy of self-reported height measurements in parents and its effect on mid-parental target height calculation. BMC Endocr Disord 2007;7:2.
8. Cizmecioglu F, Doherty A, Paterson WF, Young D, Donaldson MD. Measured versus reported parental height. Arch Dis Child 2005;90:941-2.
9. Engstrom JL, Paterson SA, Doherty A, et al. Accuracy of self-reported height and weight in women: an integrative review of the literature. J Midwifery Womens Health 2003;48:338-45.
10. World Health Organization. Physical Status: The Use and Interpretation of Anthropometry. Report of a WHO Expert Committee. Geneva: World Health Organization Tech Rep Ser 854; 1995.
11. Henry JJ. Routine growth monitoring and assessment of growth disorders. J Pediatr Health Care 1992;6(5 Pt 2):291-301.
12. Basic human body measurements for technological design (ISO 7250:1996).
13. Fleuren M, Wiefferink K, Paulussen T. Determinants of innovation within health care organizations: literature review and Delphi study. Int J Qual Health Care 2004;16:107-23.

14. Rogers EM. Diffusion of Innovations. New York: The Free Press, 2003.
15. Logan J and Graham ID, Towards a comprehensive interdisciplinary model of health care research use, Science Communication 1998;20:227-46.
16. Bartholomew KL, Parcel GS, Kok G and Gottlieb NH, Intervention mapping: designing theory- and evidence-based health promotion program, McGraw-Hill, New York, 2001.

Samenvatting

Het monitoren van groei vormt al meer dan een eeuw een onderdeel van de preventieve gezondheidszorg voor kinderen. Het volgen van de groei is belangrijk om de gezondheids- en voedingssituatie van een individueel kind of van een populatie in kaart te brengen. Op kindniveau bestaat het monitoren van groei uit het meten van de lengte, het gewicht en de hoofdomtrek van het kind en het registreren van de metingen op het groeidiagram. De positie van de metingen op het groeidiagram geeft aan in hoeverre het groeipatroon van het kind afwijkt ten opzichte van de leeftijdsgenoten. Een belangrijk doel van het monitoren van groei is het tijdig opsporen van genetische aandoeningen, ziekten en andere aandoeningen die gepaard gaan met een afwijkende groei. Voorbeelden hiervan zijn het syndroom van Turner, groeihormoon deficiëntie, coeliakie, taaislijmziekte, hypertone dehydratie en obesitas. Merk op dat obesitas niet een aandoening op zichzelf is, maar een risicofactor voor verminderde morbiditeit en mortaliteit. Op populatieniveau bestaat het monitoren van groei uit de studie van groei van een steekproef uit de gehele populatie. Zo weten we bijvoorbeeld dat sinds 1980 het percentage overgewicht onder Nederlandse kinderen enorm is gestegen.

Ondanks de brede acceptatie en uitvoering is er weinig bekend over hoeveel het monitoren van groei bijdraagt aan de tijdige opsporing ("screening") van aandoeningen. Screening of bevolkingsonderzoek is het onderzoeken van een populatie om gevallen van een ziekte of aandoening op het spoor te komen, in de veronderstelling dat deze aandoening in een vroeg stadium beter te behandelen is. Een screeningsprogramma moet aan verschillende eisen voldoen. Eén van die eisen is dat er een valide test (verwijscriterium) moet bestaan. Met andere woorden, er moet een test zijn waarbij het percentage kinderen dat wordt opgespoord met een ziekte of aandoening hoog is (hoge sensitiviteit) en waarbij het percentage gezonde kinderen dat *niet* wordt verwezen hoog is (hoge specificiteit).

De centrale vraag van dit proefschrift is: 'Wat is de validiteit van verwijscriteria voor groei?' Om deze vraag te kunnen beantwoorden hebben we het groeipatroon van kinderen met het syndroom van Turner, coeliakie, taaislijmziekte en hypertone dehydratie (uitdroging) bestudeerd. Dit groeipatroon hebben we afgezet tegen het groeipatroon van kinderen uit de algemene populatie (referentiekinderen). In dit proefschrift zijn de groeigegevens gestandaardiseerd naar leeftijd en geslacht in zogenoemde standaard deviatie scores (SDS). Een SDS is een maat die de afwijking van de mediaan van de algemene populatie uitdrukt als het aantal standaard deviaties onder (negatieve SDS) of boven (positieve SDS) de mediaan.

Het is bekend dat tweelingen anders groeien dan eenlingen, maar tot op heden bestonden er geen aparte groeidiagrammen voor deze groep. Hoofdstuk 2 van dit proefschrift presenteert groeidiagrammen voor groei van tweelingen tussen 0 en 2,5

jaar. Hiermee is het mogelijk om de groei van een tweeling ten opzichte van andere tweelingen te beoordelen. De validiteit van verwijscriteria hangt af van factoren die een significant effect kunnen hebben op groei. Hoofdstuk 3 laat zien dat genetische- en omgevingsfactoren beiden een invloed hebben op het groeiproces van 0 tot 2,5 jarigen. Omdat de invloed van genen sterk is op de groei van 2 jarigen is ervoor gekozen in de rest van het proefschrift de ouderlengte te gebruiken als parameter voor genetische achtergrond van de groei.

In de hoofdstukken 4 t/m 6 zijn verwijscriteria ontwikkeld ten behoeve van de screening op klinisch relevante groeistoornissen met een kleine lengte. Een richtlijn voor kleine lengte wordt gepresenteerd in hoofdstuk 6. Volgens deze richtlijn is er sprake van verwijzing bij kinderen van 0-3 jaar oud als:

- hij/zij een extreem kleine lengte (<-3 SDS) heeft en het geboortegewicht is ≥ 2500 g,
- hij/zij herhaaldelijk een zeer kleine lengte (<-2,5 SDS) heeft en het geboortegewicht is ≥ 2500 g. Met herhaaldelijk wordt hier bedoeld dat er na een half jaar, maar uiterlijk binnen een jaar een tweede lengtemeting moet zijn. Als hierbij óók een zeer kleine lengte wordt gevonden, dient verwezen te worden. Met bovenstaande criteria kan 15% van de meisjes met het syndroom van Turner worden opgespoord. Dit gaat ten koste van het onterecht verwijzen van minder dan 1% van de gezonde kinderen.

Er is sprake van verwijzing bij kinderen van 3-10 jaar oud als:

- hij/zij een zeer kleine lengte (<-2,5 SDS) heeft,
- hij/zij een kleine lengte (<-2 SDS) heeft in combinatie met een lengte die meer dan twee standaard deviaties afwijkt van de target height of als er een afbuiging plaatsvindt van meer dan één standaard deviatie. Met deze criteria kan 86% van de meisjes met het syndroom van Turner worden opgespoord en 27% van de kinderen met coeliakie. Dit gaat ten koste van het onterecht verwijzen van 2% van de gezonde kinderen.
- hij/zij een kleine lengte (<-2 SDS) heeft met een geschiedenis van een laag geboortegewicht of -lengte (<-2 SDS) gecorrigeerd voor zwangerschapsduur: small for gestational age.

Indien een kind bepaalde symptomen heeft (zoals van emotionele deprivatie, disproportie en/of dysmorfe kenmerken) kan al eerder tot verwijzing worden overgegaan

In Hoofdstuk 5 wordt de validiteit onderzocht van criteria gebaseerd op meerdere metingen (≥ 3) over een bepaalde periode, bijvoorbeeld van geboorte tot aan vijf jaar. Met die metingen kan een groeicurve voor het kind worden opgesteld met de Jenss-Bayley formule. De parameters van deze groeicurve kunnen zeer goed voorspellen of het kind wel of niet het syndroom van Turner heeft.

Bij kinderen met coeliakie en taaislijmziekte speelt de verhouding tussen gewicht en lengte een grote rol (H7 en H8). Uit het onderzoek blijkt dat body mass index (BMI: kg/m^2) de meest informatieve groeiparameter is om deze kinderen op te kunnen sporen. In de leeftijd van 0-2,5 jaar heeft ongeveer één op de drie kinderen met symptomatische coeliakie of taaislijmziekte een lage BMI (<-1,5 SDS en <-2,5 SDS respectievelijk) in combinatie met een afbuiging in BMI (>2,5 SDS en >0,5 SDS respectievelijk), tegen 2% van de gezonde kinderen.

In de eerste levensweken is het gewichtsverlies van borstgevoede baby's met hypertone dehydratie (ernstige uitdroging) beduidend groter dan van borstgevoede baby's zonder hypertone dehydratie. Om te kunnen bepalen of een baby een risico loopt op hypertone dehydratie stelt hoofdstuk 9 een groeidiagram op van het gewichtsverlies van gezonde, borstgevoede baby's in de eerste twee weken. Aan de hand van het groeidiagram is er een richtlijn ontwikkeld om baby's met hypertone dehydratie in een zo vroeg mogelijk stadium op te kunnen sporen. Deze richtlijn geeft aan dat er een interventie moet plaatsvinden als het gewichtsverlies meer dan tien procent is in de eerste week. Als het gewichtsverlies meer dan 12% is in de eerste week of het gewichtsverlies is < -2,5 SDS, dan is er sprake van een groot risico op hypertone dehydratie. In totaal kan de richtlijn 90% van de baby's met hypertone dehydratie opsporen ten koste van 2% van de gezonde kinderen.

Met het monitoren van groei kan worden bepaald of een kind overgewicht of obesitas heeft. De prevalentie van overgewicht en obesitas bij Nederlandse kinderen kan worden berekend aan de hand van BMI met International Obesity Task Force (IOTF) afkappunten. In Nederland is de prevalentie van overgewicht en obesitas in de periode 1980 tot 1997 sterk toegenomen. De prevalentie tussen 1997 en 2003 is zelfs nog sterker toegenomen (H10). Gemiddeld genomen had 14,5% van de jongens en 17,5% van de meisjes tussen de 4 en 15 jaar overgewicht in 2003.

Hoofdstuk 11 is een methodologisch artikel dat aantoont dat BMI niet correct corrigeert voor lengte voor jongens van 4-14 jaar en meisjes van 4-12 jaar. De prevalentie van overgewicht van lange kinderen is aanzienlijk hoger dan de prevalentie van overgewicht van korte kinderen. De odds ratio van overgewicht in lange jongens ten opzichte van korte jongens is 4,6. Dit is 3,8 voor meisjes. Deze hoge odds ratio heeft een deels methodologische verklaring. Een alternatief voor BMI is de "Weight for Height Estimate" (WHE) met $WHE = (WSDS-\rho(HSDS))/\sqrt{(1-\rho^2)}$ met ρ de correlatie tussen lengte SDS en gewicht SDS. Deze definitie geeft bij kinderen een betere lengte-correctie van gewicht dan BMI.

Voor het eerst is onderzocht wat de validiteit is van verwijscriteria voor groei. Tezamen geven de hoofdstukken in dit proefschrift aan dat de validiteit hoog is. Het op de juiste wijze monitoren van groei leidt tot vroege opsporing van kinderen met aandoeningen.

List of abbreviations

A	additive genetic effects
AIC	Akaike's information criterion
BIC	Bayesian Information criterion
BMI	body mass index
C	common environmental effects
CD	celiac disease/coeliac disease
CDC	Centers for Disease Control and Prevention
CF	cystic fibrosis
CI	confidence interval
Defl	deflection rule
DOS	dizygotic opposite sex
DZ	dizygotic
DZB	dizygotic boys
DZF	dizygotic females
DZG	dizygotic girls
DZM	dizygotic males
EKD	Electronic Child Record/ Elektronisch Kind Dossier
FTT	failure to thrive
GA	gestational age
GHD	growth hormone deficiency
GH	growth hormone
HD	hypernatraemic dehydration
HSDS	height standard deviation score
ICP	infancy-childhood-puberty growth curve model
IOTF	International Obesity Task Force
JB	Jenss-Bayley growth curve
Limburg	cohort of children born in the years 1989 and 1990 in Landgraaf and Kerkrade
LMS	skewness/median/coefficient of variation curves
LR	likelihood ratio statistic
MPH	mid-parental height
MZ	monozygotic
MZB	monozygotic boys
MZF	monozygotic females
MZG	monozygotic girls
MZM	monozygotic males
NBS	newborn screening
NCHS	National Center for Health Statistics
NFI	Normed Fit Index
NSC	National Screening Committee
NTR	Netherlands Twin Register
OR	odds ratio
P	Percentile
PHC	parental height corrected rule
PS	prognostic score
P4	polynomial of degree four
ρ	correlation
RWC	relative weight change
S	specific environmental effects
SDS	standard deviation score
SGA	small for gestational age
SMOCC	Social Medical Survey of Children Attending Child Health Clinics cohort
SSP	short stature due to pathology
TH	target height
UK	United Kingdom
US	United States
WHO	World Health Organization
WSDS	weight standard deviation score
ZHN	cohort children born in 1985-1988, attending school doctors between 1998 and 2000 in Leiden and Alphen aan den Rijn

Acknowledgements

This thesis would not have been accomplished without the help and contribution of other people. First and foremost I would like to express my gratitude to my supervisor, Stef van Buuren, for his inspiring discussions and his smooth (or at least piece-wise continuous) way to convince me to write this thesis. I would also like to greatly acknowledge the valuable contributions that were made by Paul Verkerk. The analyses for the twin studies were performed with the help of the friendly and helpful people from the VU University, Amsterdam: Mathisca de Gunst, Aad van der Vaart and Dorret Boomsma. I thank Floor Grote for sharing her medical knowledge as well as her biscuits and licorice candy. In particular, I would like to thank Jan Maarten Wit. I was very pleased that I had the privilege of working with him. In addition, I want to thank Wilma Oostdijk for her medical advice and support. I would like to thank Katja Hurk for her work. Special thanks are directed to Ko van Wouwe, as he helped me to conquer my fear of public speaking, he introduced me into the world of midwives, and simply for being such a friendly person. I also greatly appreciate the work Jacqueline Breuning-Boers has done in finding all cases of hypernatraemic dehydration. The evidence I found was all because of you. Jaap van der Plas, thank you for the layout of this thesis. I would like to thank Ab Rijpstra for his nice conversations about all topics that are not related to my thesis. I would also express my gratitude to my parents who always supported me in trying to get the best education, irrespective of whether that was in the Netherlands or at the other side of the world. And finally, my two little men Jason and Olivier. Being a wife and mother is more important than being a Doctor of Philosophy. Thank you Jason, for your patience. Olivier, don't become a nerd like your mother and father!

Curriculum vitae

Paula van Dommelen was born on the 2nd of January, 1978 in Linschoten, the Netherlands. She attended secondary school at the 'De Bruijne Lyceum' in Utrecht, where she passed her exam in 1996. Then she started her study of mathematics at the VU University in Amsterdam. In 1999, she followed several courses in mathematics and statistics at the Canterbury University in Christchurch, New Zealand. In 2001, she finished her degree in mathematics. In 2001 she started to work at TNO at the department of statistics as statistical advisor and researcher.

List of peer-reviewed publications

1. Balk AH, Davidse W, van Dommelen P, Klaassen E, Caliskan K, van der Burgh P, Leenders CM. Tele-guidance of chronic heart failure patients enhances knowledge about the disease. A multi-centre, randomised controlled study. Eur J Heart Fail 2008 Sep 12.
2. van Dommelen P, Grote FK, Oostdijk W, Keizer-Schrama SM, Boersma B, Damen GM, Csizmadia CG, Verkerk PH, Wit JM, van Buuren S. Screening rules for growth to detect celiac disease: a case-control simulation study. BMC Pediatrics 2008;8:35.
3. van Wouwe JP, Lanting CI, van Dommelen P, Treffers PE, van Buuren S. Breastfeeding duration related to practised contraception in the Netherlands.Acta Paediatr 2008 Sep 2.
4. van Dommelen P, Schuller A, Detmar S, van den Berg B, Groothuis Oudshoorn K. Preference-based measure of social participation from the Handicap Scale for Children. Expert Rev Pharmacoeconomics Outcomes Res 2008;8:309-17.
5. Grote FK, Oostdijk W, De Muinck Keizer-Schrama SM, van Dommelen P, van Buuren S, Dekker FW, Ketel AG, Moll HA, Wit JM. The diagnostic work up of growth failure in secondary health care; an evaluation of consensus guidelines. BMC Pediatr 2008;8:21.
6. van Dommelen P, de Gunst M, van der Vaart A, van Buuren S, Boomsma D. Growth references for height, weight and body mass index of twins aged 0-2.5 years. Acta Paediatr 2008;97:1099-104.
7. van den Berg B, Van Dommelen P, Stam P, Laske-Aldershof T, Buchmueller T, Schut FT. Preferences and choices for care and health insurance. Soc Sci Med 2008;66:2448-59.
8. Grote FK, van Dommelen P, Oostdijk W, de Muinck Keizer-Schrama SM, Verkerk PH, Wit JM, van Buuren S. Developing evidence-based guidelines for referral for short stature. Arch Dis Child 2008;93:212-7.
9. van den Berg B, Van Dommelen P, Stam P, Laske-Aldershof T, Schut FT. Preferenties voor zorg en zorgverzekeraars. TSG 2008;1:39-46.
10. Wijlhuizen GJ, du Bois P, van Dommelen P, Hopman-Rock M. Effect evaluation of a multifactor community intervention to reduce falls among older persons. Int J Inj Contr Saf Promot 2007;14:25-33.
11. van den Hurk K, van Dommelen P, van Buuren S, Verkerk PH, Hirasing RA. Prevalence of overweight and obesity in the Netherlands in 2003 compared to 1980 and 1997. Arch Dis Child 2007;92:992-5.
12. Grote FK, Oostdijk W, De Muinck Keizer-Schrama SM, Dekker FW, van Dommelen P, van Buuren S, Lodder-van der Kooij AM, Verkerk PH, Wit JM. Referral patterns of children with poor growth in primary health care. BMC Public Health 2007;11;7:77.

13. Rijnders MEB, van den Akker-van Marle ME, van Dommelen P, Amelink-Verburg MP, van Wouwe JP, Verkerk PH. Kosteneffectiviteitsanalyse van verschillende strategieën ter preventie van neonatale infectie met groep B-streptokokken. Ned Tijdschr Geneeskd 2007;151:189-93
14. van Dommelen P, van Wouwe JP, Breuning-Boers JM, van Buuren S, Verkerk PH. Reference chart for relative weight change to detect hypernatraemic dehydration. Arch Dis Child 2007;92:490-4.
15. Breuning-Boers JM, van Dommelen P, van Wouwe JP, Verkerk PH. Gewichtsverlies, serumnatriumconcentratie en restverschijnselen bij patiëntjes met hypertone dehydratie door onvoldoende borstvoeding. Ned Tijdschr Geneeskd 2006;150:904-8.
16. van Dommelen P, van Buuren S, Zandwijken GR, Verkerk PH. Individual growth curve models for assessing evidence-based referral criteria in growth monitoring. Stat Med 2005;24:3663-74.
17. Akker-van Marle ME, Rijnders ME, Dommelen P, Fekkes M, Wouwe JP, Amelink-Verburg MP, Verkerk PH. Cost-effectiveness of different treatment strategies with intrapartum antibiotic prophylaxis to prevent early-onset group B streptococcal disease. BJOG 2005;112:820-6.
18. Anthony S, Buitendijk SE, Offerhaus PM, Dommelen P, Pal-de Bruin KM. Maternal factors and the probability of a planned home birth. BJOG 2005;112:748-53.
19. van Dommelen P, de Gunst MC, van der Vaart AW, Boomsma DI. Genetic study of the height and weight process during infancy. Twin Res 2004;7:607-16.
20. van Dommelen P, De Gunst MC, Van Der Vaart AW, Van Buuren S, Boomsma DI. Groeidiagrammen voor lengte, gewicht en 'body mass index' van tweelingen in de peutertijd. Ned Tijdschr Geneeskd 2004;148:1345-50.
21. van Buuren S, van Dommelen P, Zandwijken GR, Grote FK, Wit JM, Verkerk PH. Towards evidence based referral criteria for growth monitoring. Arch Dis Child 2004;89:336-41.

Other publications

1. van Dommelen P, van Wouwe JP, Breuning-Boers JM, van Buuren S, Verkerk PH. Gewichtsverlies bij borstgevoede pasgeborenen. Tijdschrift voor Verloskundigen jan 2008;23-5.
2. van Dommelen P, van Wouwe JP, Breuning-Boers JM, van Buuren S, Verkerk PH. Gewichtsverlies bij borstgevoede pasgeborenen. NVL info 2007;10:24-5.
3. van Putten IKF, van Dommelen P, van Wouwe JP, de Wolf BS. Voorspelt de mictiefrequentie het gewichtsverlies van borstgevoede kinderen. Tijdschrift voor verloskundigen mei 2007;25-9.
4. Straaten van HLM, Dommelen van P, Verkerk PH. Jaarverslag neonatale gehoorscreening in de Neonatale Intensive Care Units 2006. Zwolle: Isala Klinieken, 2007.
5. Fleuren MAH, Dommelen van P, Kamphuis M, Velzen-Mol van HWM. Landelijke Implementatie JGZ-standaard Vroegtijdige Opsporing van Aangeboren Hartafwijkingen 0-19 jaar. Leiden: TNO Kwaliteit van Leven, 2007. Publ. nr. 07.119.
6. Campagne A, Verkerk PH, Kamphuis M, Dommelen van P. Het BET-project: primaire preventie van bedplassen: hoe effectief zijn twee veel door ouders toegepaste technieken bij het ‘s nachts droog krijgen van hun kind. Leiden: TNO Kwaliteit van Leven, 2007.
7. Bruil J, Vries de SI, Dommelen van P, Chorus A, Baan-Slootweg van der OH. Evaluatie van Victory Camp en Real Victory 2004-2005: resultaten van het zomerkamp en de nazorg voor jongeren op overgewicht, gedrag en kwaliteit van leven. Leiden: TNO Kwaliteit van Leven, 2006. Publ. nr. 06.048.
8. Fleuren MAH, Verlaan ML, Velzen-Mol van HWM, Dommelen van P. Zicht op het gebruik van de JGZ-standaard Opsporing van Visuele Stoornissen 0-19 jaar: een landelijk implementatieproject. Leiden: TNO Kwaliteit van Leven, 2006. Publ. nr. 06.036.
9. Tak ECPM, Hespen van ATH, Dommelen van P, Hopman-Rock M. Ontwikkeling en evaluatie van het incontinentie trainingsprogramma INCOnditie voor vrouwen in verzorgingstehuizen. Leiden: TNO Preventie en Gezondheid, 2004. Publ.nr. 04.114.
10. Öry F, Bouhalhoul R, Bengi L, Belhaj K, Dahhan N, Wit de J, Arslan H, Taoujni N, Tellegen B, Dommelen van P, Schothorst Y, Veltman M, Yildirim S, Steketee K. Toegankelijkheid van de Ouder- en Kindzorg voor Marokkaanse en Turkse gezinnen. Leiden: TNO Preventie en Gezondheid, 2003. Publ. nr. 03.035.
11. Hoejenbos MJJ, Mulder YM, Reijneveld SA, Dommelen van P. Het Health Monitoring Instrument: ontwikkeling van een screeningsinstrument voor gezondheidsmonitoring bij militairen. Ned Militair Geneesk Tijdschr 2003;56:96-101.

12. Herschderfer KC, Anthony S, Dommelen van P, Pal-de Bruin van der KM. Het effect van gemelde problematiek in de eerstelijns verloskunde op de thuisbevalling in Nederland. Leiden: TNO Preventie en Gezondheid, 2002. Publ. nr. 02.294.
13. Reeuwijk-Werkhorst van H, Dommelen van P, Rosendal H. Patiënttevredenheid bij transmurale zorg: de ontwikkeling van een Transmurale Tevredenheidsvragenlijst voor Patiënten (TTP). Leiden: TNO Preventie en Gezondheid, 2002. Publ. nr. 02.100.
14. Bruin de SR, Quak ABWM, Dommelen van P. Vergelijking directe kosten Remicade en Enbrel. Leiden: TNO Preventie en Gezondheid, 2002. Publ.nr. 02.065.
15. Buuren van S, Dommelen van P. Documentatie berekening SD-scores. Leiden: TNO Preventie en Gezondheid, 2002. Publ. nr. 02.52.