COHORT PROFILE

Cohort Profile: The KPSC Children's Health Study, a population-based study of 920 000 children and adolescents in southern California

Corinna Koebnick,[1]* Karen J Coleman,[1] Mary Helen Black,[1] Ning Smith,[1] Jack K Der-Sarkissian,[2] Steven J Jacobsen[1] and Amy H Porter[3]

[1]Department of Research and Evaluation, Kaiser Permanente Southern California, Pasadena, USA, [2]Los Angeles Medical Center, Southern California Permanente Medical Group, Los Angeles, USA and [3]Baldwin Park Medical Center, Southern California Permanente Medical Group, Baldwin Park, USA

*Corresponding author. Department of Research and Evaluation, Kaiser Permanente Southern California, 100 Los Robles, 2nd Floor, Pasadena, CA 91101, USA. E-mail: corinna.koebnick@kp.org

How did the study come about?

Paediatric obesity rates have reached epidemic proportions in the USA, affecting 16.9% of children and adolescents.[1] Although recent nationwide trend analyses suggest that paediatric obesity rates are no longer increasing, data also show a shift towards extreme obesity.[1] Increasing concerns focus on the effects of obesity on morbidity and health-related quality of life issues, particularly in cases of paediatric onset of obesity. However, the economic burden and long-term health consequences of extreme childhood obesity are largely unknown and ill-defined because there have been no large prospective epidemiological studies that included enough extremely obese children to deliver stable estimates of the health consequences.

To address current gaps in the knowledge about the health consequences of extreme childhood obesity, we created a contemporary cohort study of children and adolescents 2–19 years of age in southern California who are actively enrolled in a large, pre-paid, integrated managed healthcare system. The Kaiser Permanente Southern California (KPSC) Children's Health Study is an ongoing cohort study that currently includes >920 000 children and adolescents and continues to enrol new patients joining the health plan. The study is currently supported by a grant from the National Institute of Diabetes and Digestive and Kidney Disorders (R21DK085395; Principal Investigator: Koebnick) and by Kaiser Permanente Direct Community Benefit Funds.

Since 2007, new screening guidelines for paediatric weight management have been implemented at KPSC offering a unique opportunity to fill current knowledge gaps regarding the disease burden associated with childhood obesity. These guidelines include regular screening of blood lipids, liver enzymes and markers of glucose metabolism recommended for children at or above the 95th percentile of BMI-for-age according to the 2000 sex-specific Centers for Disease Control and Prevention (CDC) growth charts[2] and for children at or above the 85th percentile when additional risk factors such as a family history of diabetes or dyslipidaemia are recorded in the electronic health record. Weight and height are measured at almost every medical encounter in KPSC. Thus, there is an almost complete coverage of paediatric KPSC members who had medical encounters since 2007 regarding body mass index (BMI).

What does it cover?

The goal of the KPSC Children's Health Study is to estimate the association between the severity of obesity and adverse health conditions in children and adolescents 2–19 years of age. The adverse health conditions of childhood obesity include those conditions that are related to the metabolic syndrome and potential consequences of the chronic inflammation underlying obesity: impaired glucose tolerance, dyslipidemia, elevated blood pressure and hepatic disease. Ancillary studies include outcomes such as psoriasis, rheumatic diseases, gallstones and idiopathic intracranial hypertension. The study will also assess the potential modification of the association between

the severity of obesity and adverse health conditions by demographic factors such as sex, race/ethnicity and age. Because prospective data are being collected, current studies primarily are cross-sectional studies from the cohort inception.

Besides health consequences of extreme childhood obesity, the aim of the KPSC Children's Health Study is to provide insight into the concept of 'extreme obesity' in children, which currently lacks a standard definition. The 2000 CDC growth charts lack the precision necessary to calculate percentiles <3 and >97;[2,3] however, multiples of other percentiles corresponding to adulthood obesity classes at 18 years of age can be used to define extreme obesity in children. Adulthood obesity class 2 (BMI $\geqslant$ 35 kg/m^2) corresponds approximately to $1.2 \times$ 95th percentile of BMI-for-age after 18 years of age. Therefore, the new definition of extreme childhood obesity proposed by the CDC is a BMI-for-age of $\geqslant 1.2 \times$ 95th percentile or a BMI $\geqslant$ 35 kg/m.[2,3] The population size of the KPSC Children's Health Study and the information available from electronic health records will enable us to provide insight into the value of this new cut-off for risk estimation and clinical practice.

Who is in the sample?

The KPSC Children's Health Study includes children who are actively enrolled in a large, prepaid, integrated managed healthcare system. The KPSC setting consists of three entities: (i) Kaiser Foundation Health Plan that provides prepaid healthcare insurance, (ii) Kaiser Foundation Hospitals that own and operate 12 hospitals and >130 other medical offices to deliver ambulatory care and (iii) Southern California Permanente Medical Group, a partnership of >6000 physicians who represent the entire range of medical specialists who provide the vast majority of care received by KPSC members. Members and their families enrol through their employer, the employer of a family member, an individual prepaid plan or a state or federal programme such as Medi-Cal, CALPERS or Medicare. Children receive medical care mainly from KPSC paediatricians, and also may be seen by family practitioners, specialists or other healthcare professionals.

The KPSC Children's Health Study cohort (2007–09) is currently composed of 920 034 children and adolescents 2–19 years of age (Figure 1). Between 1 January 2007 and 31 December 2009, KPSC provided comprehensive health services to >3.4 million patients including 1 295 971 children and adolescents 2–19 years of age. After exclusion of 265 241 members who did not have any medical encounters between 2007 and 2009, 1 030 730 patients were eligible for participation in the cohort study. Out of these patients, 920 034 patients (89.2% of eligible patients) had at least one valid weight and height in the 3-year study period. The study protocol was reviewed and approved by KPSC's Institutional Review Board.

Children and adolescents enrolled in the KPSC health plan represent ~12.3% of the population in the counties of southern California served by KPSC. The demographic characteristics of members currently enrolled in the study are shown in Table 1. Compared with the southern California population, KPSC patients were similar in the distribution of sex, but slightly older than the underlying population.[4] Ethnic/racial distribution differed slightly with more Hispanic and fewer non-Hispanic Whites in the health plan than the California population of the same age.[4]
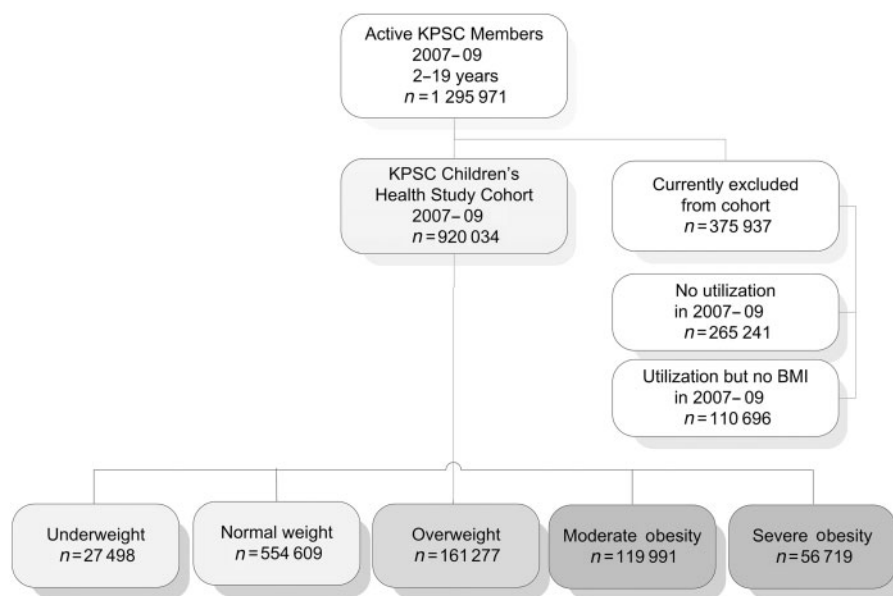

Figure 1 Flow chart of the KPSC Children's Health Study

Table 1 Demographic characteristics of KPSC members 2–19 years of age who are currently included or excluded from the KPSC Children's Health Study

Variable	KPSC Children's Health Study Cohort	Currently excluded KPSC members
n	920 034	375 937
Sex		
Male (%)	49.9	53.3
Age group (%)		
2–5 years	25.2	14.8
6–11 years	28.7	27.7
12–19 years	46.1	57.4
Race/ethnicity (%)		
Non-Hispanic White	21.1	21.9
Hispanic	51.0	43.5
Black	7.5	4.9
Asian or Pacific Islander	6.8	5.2
American Indian/Alaskan Native	0.1	0.1
Multiple races	0.5	0.2
Others	1.2	0.7
Unknown race/ethnicity	11.8	23.6
Service area (%)		
Los Angeles Metropolitan Area	48.0	58.9
Orange County	11.8	5.8
Inland Empire	24.1	23.8
Kern County	3.5	0.0
San Diego	12.7	7.4
Neighborhood education (%)		
Less than high school	28.5	30.2
High school graduate	21.5	21.9
Some college or associate degree	30.3	29.8
Bachelor degree or higher	19.7	18.1
Neighbourhood income (%)		
<$15 000	10.4	11.1
$15 000 to $34 999	19.0	20.0
$35 000 to $49 999	14.3	14.7
$50 000 to $74 999	19.6	19.6
$75 000 to $99 999	13.8	13.4
$100 000 to $149 999	14.6	13.6
$150 000 or more	8.5	7.7
Medi-Cal (%)	13.3	4.0
Born in KPSC hospital (%)	30.0	12.7

How have they been followed up?

The follow up of the cohort is through passive surveillance of clinical care information and utilizes the KPSC electronic health record system called Kaiser Permanente HealthConnect™. Information on virtually all aspects of care delivered is captured and extracted for research purposes. All administrative and clinical data are linked through a unique medical record number and include membership information and benefits, medical encounters, disease codes, health claims, laboratory test results, pharmaceuticals dispensed, treatments administered and procedures performed in pathology as well as pathology and radiology results. Information for the KPSC Children's Health Study is updated on an annual basis.

What has been measured?

Outcome ascertainment

Disease outcomes are extracted from electronic health records using International Classification of Diseases (ICD-9) codes and KPSC internal diagnosis codes. The records per medical encounter contain up to 45 diagnostic codes and up to 45 procedure codes. The validity of diagnosis codes is relatively high in this integrated healthcare system. Due to referral to specialists and removal of wrong diagnosis codes, the validity for most diagnosis codes is well above 90%. Diagnosis validations have been performed in many areas such as psoriasis and psoriatic arthritis,[5] gestational diabetes,[6,7] gastroesophageal reflux disease[8] and cardiovascular disease.[9] Special diagnosis validation in this paediatric population is ongoing.

Cardiovascular risk factors such as blood pressure, blood lipids and glucose levels are extracted from the vital signs and the laboratory results of the electronic health record.

Race and ethnicity

Race and ethnicity information was obtained from birth certificate information for 27.7% of the cohort members. If no birth certificate was available, race and ethnicity information was obtained from health plan administrative records (40.6% of cohort members; Figure 2). The quality of information retrieved from health plan administrative records has been validated and described in detail elsewhere.[10] We briefly validated information on race and ethnicity extracted from health plan administrative records of 325 810 children born between 1998 and 2008 against birth certificate records. Positive predictive values (PPVs) for White, Black and Asian/Pacific Islander were 89.3, 86.6 and 73.8%, respectively. PPV for Hispanic ethnicity was 95.6%. Concordance of racial and ethnic information improved with an increasing number of medical visits.[10]

For unknown race and ethnicity information (31.7% of cohort members), an imputation algorithm was used based on surname lists and address information derived from the US Census Bureau.[7–9] The optimal probability cut-offs were chosen based on receiver operating characteristic curve analyses in patients with known race ethnicity when specificity
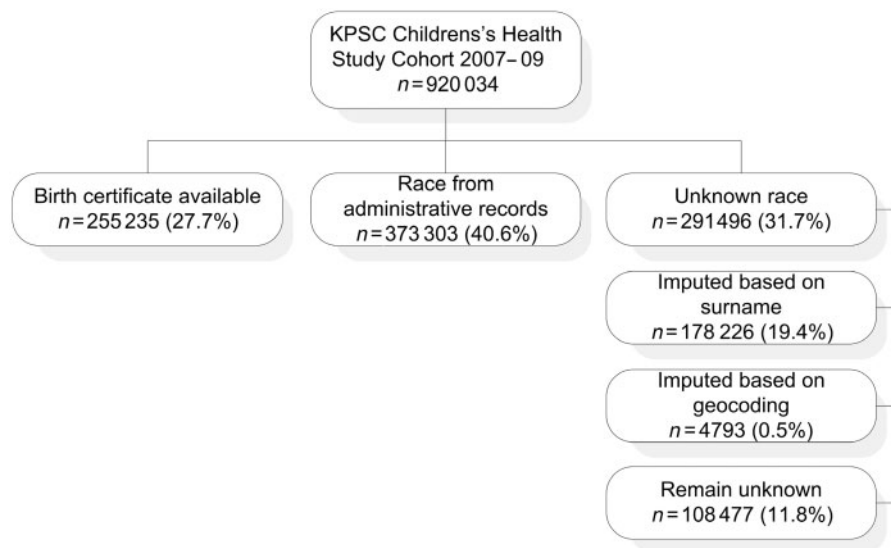
Figure 2 Source of race and ethnicity information in the KPSC Children's Health Study

was >98%. Asian/Pacific Islander race was assigned when the probability cut-off for surname was >40% for Asian/Pacific Islander race. Hispanic ethnicity was assigned when the probability cut-off for surname was >50% for Hispanic ethnicity. For Blacks and non-Hispanic Whites, members' addresses were geocoded to the US Census Block level. Race and ethnicity were assigned on the basis of the composition of the block, applying a probability cut-off >75% for Black and >45% for White race. We categorized race/ethnicity as non-Hispanic White, Hispanic or Latino, Black, Asian or Pacific Islander, other or multiple race/ethnicity and unknown due to missing information. The distribution of the population remained essentially unaltered after race imputation. Out of the 31.7% of cohort members, race information of 19.4% was imputed based on surname-based race probability, 0.5% was imputed based on geocoding information and 11.8% remained unknown due to insufficient probability for imputation (Figure 2). In a validation study, imputed race and ethnicity from this algorithm were validated in a subset of KPSC Children's Health Study participants who were born in KPSC hospitals, and for which race and ethnicity information from birth certificates was available for both parents ($n = 250\,921$ children). Imputation of race and ethnicity information based on surname and geocoding was possible for only 68.3% of children, whereas 31.7% remained unclassified. Compared with birth certificate information, the imputed race and ethnicity information was correct for 85.5% of children.

Body weight and height
Information from electronic medical charts was utilized to extract body weight and height from the same day. BMI was calculated as weight (kilograms) divided by the square of the height (metres). For each year (2007–09), the median BMI-for-age of all encounters for a patient from the most recent available year was used for these analyses. Based on a validation study including 15 000 patients with 45 980 medical encounters, the estimated error rate in body weight and height data was <0.4%.[11]

Definitions for overweight and obesity in children and adolescents are based on the sex-specific BMI-for-age growth charts developed by the CDC and based on World Health Organization definitions for overweight and obesity in adults.[12–14] Children were categorized as underweight (BMI-for-age <5th percentile), normal weight (BMI-for-age $\geqslant$5th and <85th percentile), overweight (BMI-for-age $\geqslant$85th percentile or a BMI $\geqslant 25\,\mathrm{kg/m}^2$), moderately obese (BMI-for-age $\geqslant$95th percentile or a BMI $\geqslant 30\,\mathrm{kg/m}^2$) and extremely obese (BMI-for-age $\geqslant 1.2 \times$ 95th percentile or a BMI $\geqslant 35\,\mathrm{kg/m}^2$).

How much attrition has there been?

Currently, 89.2% of eligible patients had at least one valid weight and height in the 3-year study period and were included in the final cohort. By the end of 2009, 10.8% of KPSC members 2–19 years of age were not included in the final cohort because of missing weight and height information. Because weight and height are now measured at almost every medical encounter, these members will be included at a later point in time after they have had a medical encounter. The demographic characteristics of participants of the KPSC Children's Health Study and KPSC members who are currently excluded from the cohort are shown in Table 1. Currently excluded KPSC members

Table 2 Percentage of active KPSC members on 15 January 2005, 0–24 years of age and the percentage retained in the KP plan between 2005 and 2010

Age (years)	Retention of active members (%) in years					
	2005	2006	2007	2008	2009	2010
0–2	100	89.9	82.6	76.8	71.8	66.5
3–6	100	91.0	84.4	79.1	74.1	68.8
7–12	100	92.0	86.0	80.9	76.3	71.1
13–17	100	92.5	83.1	73.6	63.8	51.5
18–24	100	80.4	67.5	56.1	45.6	35.8

are older, more likely to be male and of unknown race and ethnicity.

The 5-year retention rate for children and adolescents from recent years at KPSC is a function of health plan membership and varies by age group with the highest 5-year retention in 7- to 12-year-old children and the lowest 5-year retention in adolescents and young adults (Table 2). Due to recent policy changes that grant longer insurance coverage for young adults under their parent's health insurance, we expect these retention rates to increase over the next few years. Although long-term follow up is dependent on continued membership, some events such as death and cancer can be traced through probabilistic linkage with the California state and national death files and cancer registry files. If younger participants lose health coverage due to loss of employment of their parents, they may continue health care under a state or federal programme such as Medi-Cal, CALPERS or Medicare.

Most children included in the study have been members for many years. The average length of membership before study enrolment (mean ± SD) is 2.6 ± 1.6 years for children 2–5 years of age, 5.8 ± 3.7 years for children 6–11 years of age and 9.3 ± 6.1 years for children 12–19 years of age. Thirty per cent of children in the cohort were born in KPSC hospitals and have been members since birth (Table 1).

What are the key findings and publications

Previous results from the study have shown that 7.3% of boys and 5.5% of girls in southern California can be classified as extremely obese.[4] The prevalence of extreme obesity peaked at 10 years of age in boys and at 12 years of age in girls, with a bimodal distribution in girls (second peak at 18 years, P for sex $\times$ age interaction = 0.036). The prevalence of extreme obesity varied among ethnic/racial and age groups with the highest prevalence in Hispanic boys (up to 11.2%) and Black girls (up to 11.9%).[4]

What are the main strengths and weaknesses?

The KPSC Children's Health Study is unique with respect to the population size and the broad information provided by an integrated healthcare system with access to laboratory results and medical diagnoses. The study sample currently includes 89.2% of eligible KPSC members. One particular advantage of this setting is that all clinical data represent routine healthcare delivered to KPSC patients by KPSC providers; these data are therefore indicative of typical healthcare decisions. All subjects have relatively consistent health insurance benefits, and clinicians are influenced by a consistent incentive system and centralized efforts to standardize practice. Thus, there are minimal effects on utilization and outcomes from insurance coverage that may confound outcomes in other settings. The KPSC Children's Health Study represents ∼12.3% of the population in the counties of southern California served by KPSC, includes a large number of children from low-income groups and is roughly comparable regarding most sociodemographic factors.

The study is also unique in its inclusion of a large number of severely obese children, which will enable us to conduct stratified analyses for important potential effect modifiers. Additionally, the study setting allows for collection of data on laboratory markers and diagnoses made by physicians, which are not available in other cohort studies, such as National Health and Nutrition Examination Survey.

A potential strength and limitation of the proposed study is the follow up through passive surveillance which offers unique opportunities to follow an extremely large population. However, it also implies some restrictions arising from real life as opposed to research settings. These restrictions include a potential variation in the proportion of children who are screened for obesity-related conditions within KPSC. Physician's screening behaviour may vary across medical centres. Because most of these conditions are silent conditions without symptoms, children may not seek a doctor's attention and may not be tested. The true prevalence of some conditions may not be estimable since data were not generated by a comprehensive screening of the entire KPSC population of the targeted age range. However, the increasing number of screening tests, such as glucose and cholesterol, performed each year suggests a high awareness of KPSC physicians for populations at risk for obesity-related conditions. Computer-assisted aids implemented in the medical record system make compliance with the screening recommendations feasible by BMI-triggered 'best practice' alerts that make the physicians aware of the condition (i.e. obesity) and recommend screening for risk factors. Laboratory orders can then be placed by simply clicking on the alert flag.

Another potential limitation of the study is the restricted availability of cardiovascular screening results in children who are normal weight and healthy. Screening for abnormalities in fasting glucose, blood lipids and liver enzymes is usually only performed in children who are overweight unless other conditions require surveillance of these factors. Therefore, silent conditions such as liver disease may only be diagnosed in children who were screened (=obese children) but not in others (=normal-weight children), which would lead to an overestimation of the association between obesity and adverse health outcome. It is noteworthy, however, that some vital signs such as blood pressure are measured at every visit regardless of body weight and will be available in all children. Using overweight children who were screened as reference as well as multiple imputation methods can partially compensate for these restrictions.

In addition, utilization may vary due to slight variations in the health plans including variations in co-payments for office visits or medical procedures such as laboratory tests or imaging. High co-payments may negatively affect the likelihood of undergoing certain preventive screening procedures, and consequently, the likelihood of being diagnosed with some health conditions. However, highly deductible plans with varying benefits and high medication co-pays were not offered comprehensively at KPSC until 2006, and mostly do not affect health care for families with children. On the other hand, the KPSC population also includes a large number of children receiving full or additional coverage through Medi-Cal and other subsidized plans. Therefore, we do not expect this to have an impact on our findings.

Race and ethnicity were not systematically collected on every member of KPSC. Multiple sources were used to supplement existing race/ethnicity data, including imputation of race and ethnicity information for 19.9% of the current cohort, which was primarily based on surname probabilities. However, we expect the potential bias that may be introduced by the imputation procedure to be minimal and to outweigh the potential bias in classifying these members as an unknown race.

Can I get access to the data? Where can I find out more?

Information on the KPSC Children's Health Study is available at http://xnet.kp.org/research/. Because the information contained in electronic health records is protected and confidential, it can be extremely difficult to exchange individual level information. Therefore, individual level information from several sources is usually managed, analysed and aggregated at the KPSC data centre in projects with external collaborators. Active collaborations are welcome and the principal investigator of the study, Dr Corinna Koebnick, can be contacted for more information.

References

[1] Ogden CL, Carroll MD, Curtin LR, Lamb MM, Flegal KM. Prevalence of high body mass index in US children and adolescents, 2007–2008. *JAMA* 2010;**303:**242–49.

[2] Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ* 2000;**320:**1240–43.

[3] Flegal KM, Wei R, Ogden CL, Freedman DS, Johnson CL, Curtin LR. Characterizing extreme values of body mass index for age by using the 2000 Centers for Disease Control and Prevention growth charts. *Am J Clin Nutr* 2009;**90:**1314–20.

[4] Koebnick C, Smith N, Coleman KJ *et al.* Prevalence of extreme obesity in a multiethnic cohort of children and adolescents. *J Pediatr* 2010;**157:**26–31 e2.

[5] Wu JJ, Porter AH, Black MH, Smith N, Jacobsen SJ, Koebnick C. Low prevalence of psoriasis among children and adolescents in a large multiethnic cohort in southern California. *J Am Assoc Dermatol* 2010. in press.

[6] Lawrence JM, Black MH, Hsu JW, Chen W, Sacks DA. Prevalence and timing of postpartum glucose testing and sustained glucose dysregulation after gestational diabetes mellitus. *Diabetes Care* 2010;**33:**569–76.

[7] Lawrence JM, Contreras R, Chen W, Sacks DA. Trends in the prevalence of preexisting diabetes and gestational diabetes mellitus among a racially/ethnically diverse population of pregnant women, 1999–2005. *Diabetes Care* 2008;**31:**899–904.

[8] Koebnick C, Getahun D, Porter AH, DerSarkissian JK, Jacobsen SJ. Extreme childhood obesity is associated with increased risk for gastroesophageal reflux disease. *Int J Pediatr Obes* 2011;**6:**e257–63.

[9] Karter AJ, Ferrara A, Liu JY, Moffet HH, Ackerson LM, Selby JV. Ethnic disparities in diabetic complications in an insured population. *JAMA* 2002;**287:**2519–27.

[10] Smith N, Iyer R, Langer-Gould A *et al*. Health plan administrative records versus birth certificate records: quality of race and ethnicity information in children. *BMC Health Serv Res* 2010;**10:**316.

[11] Smith N, Coleman KJ, Lawrence JM *et al*. Body weight and height data in electronic medical records of children. *Int J Pediatr Obes* 2010;**5:**237–42.

[12] World Health Organization. Technical Report Series 894. *Obesity: Preventing and Managing the Global Epidemic*. Geneva: World Health Organization, 2000. ISBN 92-4-120894-5.

[13] Flegal KM, Wei R, Ogden CL, Freedman DS, Johnson CL, Curtin LR. Characterizing extreme values of body mass index for age by using the 2000 Centers for Disease Control and Prevention growth charts. *Am J Clin Nutr* 2009;**90:**1314–20.

[14] Kuczmarski RJ, Ogden CL, Guo SS *et al*. 2000 CDC Growth Charts for the United States: methods and development. *Vital Health Stat* 2002;**11:**1–190.