Original Article
PEDIATRIC OBESITY

Obesity

Classification Tree Analyses Reveal Limited Potential for Early Targeted Prevention against Childhood Overweight

Andreas Beyerlein[1,2], Dennis Kusian[1,2], Anette-Gabriele Ziegler[1,2,3], Angelika Schaffrath-Rosario[4] and Rüdiger von Kries[5]

Objective: Whether specific combinations of risk factors in very early life might allow identification of high-risk target groups for overweight prevention programs was examined.
Design and Methods: Data of $n = 8981$ children from the German KiGGS study were analyzed. Using a classification tree approach, predictive risk factor combinations were assessed for overweight in 3–6, 7–10, and 11–17-year-old children.
Results: In preschool children, the subgroup with the highest overweight risk were migrant children with at least one obese parent, with a prevalence of 36.6 (95% confidence interval or CI: 22.9, 50.4)%, compared to an overall prevalence of 10.0 (8.9, 11.2)%. The prevalence of overweight increased from 18.3 (16.8, 19.8)% to 57.9 (46.6, 69.3)% in 7–10-year-old children, if at least one parent was obese and the child had been born large-for-gestational-age. In 11–17-year-olds, the overweight risk increased from 20.1 (18.9, 21.3)% to 63.0 (46.4, 79.7)% in the highest risk group. However, high prevalence ratios were found only in small subgroups, containing <10% of all overweight cases in the respective age group.
Conclusions: Our results indicate only a limited potential for early targeted preventions against overweight in children and adolescents.

Obesity (2013) **00**, 00–00. doi:10.1002/oby.20628

Introduction

The prevalence of childhood overweight has been increasing worldwide in recent decades (1,2). This increase seems to be associated rather with a shift in the upper parts of the body mass index (BMI) distribution than with a shift of the medium or lower parts (3). These temporal trends might be explicable by incremental exposure to environmental risk factors, as we have recently shown that well-known risk factors for overweight were more strongly associated with high BMI percentiles than with low or medium BMI percentiles in quantile regression analyses (4,5).

Therefore, identification of specific risk factor profiles might be pivotal for the prevention of childhood overweight. Very early life appears to be a promising period for such prevention efforts, as children's environment at this age is largely under their parents' control and is likely to have a long-term effect on their overweight risk (6,7). A recent review based on prospective studies confirmed associations of a number of early life factors with an increased risk for childhood overweight, implicating a need to identify high-risk groups of infants for clinical practice (8).

Classification trees are useful for this purpose, as they offer an unbiased and easily interpretable statistical approach to split a given dataset into low- and high-risk groups. We have already used a classification tree to identify subgroups of children who are at increased risk to be overweight shortly before school entry, but could not find a specific early risk factor combination which was highly predictive (9). In a similar study, no risk factor combination was found which would have increased prediction of overweight in preschoolers by two times or more (10). However, early life risk factors for childhood overweight might play an increasing role as child's age increases (11), and the prevalence of overweight has been shown to increase substantially after school entry (12-14).

Thus, we wondered whether data with a broader age spectrum would allow determination of potential early high-risk groups for childhood overweight. For this purpose, we analyzed a large contemporary

[1] Institute of Diabetes Research, Helmholtz Zentrum München, Munich, Germany. Correspondence: A. Beyerlein (andreas.beyerlein@helmholtz-muenchen.de) [2] Forschergruppe Diabetes der Technischen Universität München, Munich, Germany [3] Forschergruppe Diabetes e.V. am Helmholtz Zentrum München, Munich, Germany [4] Robert Koch-Institute, Berlin, Germany [5] Institute of Social Paediatrics and Adolescent Medicine, Ludwig-Maximilians University of Munich, Munich, Germany

Disclosure: The authors declare no conflict of interest.
Received: 29 May 2013; **Revised:** 10 September 2013; **Accepted:** 13 September 2013; **Published online 1 October 2013. doi:10.1002/oby.20628**

population-based German dataset containing data on children at different ages during childhood and adolescence.

Methods

The data were collected from May 2003 to May 2006 in the German Health Interview and Examination Survey for Children and Adolescents (KiGGS), a representative nation-wide survey on children and adolescents selected within 167 communities (primary sample points). Within the sample points, addresses of children were drawn randomly from local registries to invite the children and their parents to participate in the survey. The response rate was 66.6% (15). Overall, $n = 17{,}641$ children aged 0–17 years were enrolled. The study was approved by the Institutional Review Board of the Virchow-Klinikum of the Humboldt-University Berlin. A detailed description of the survey has been published elsewhere (15,16).

Children's height was measured, without wearing shoes, by trained staff with an accuracy of 0.1 cm, using a portable Harpenden infantometer or stadiometer (Holtain Ltd., Crymych, UK). Body weight was measured with an accuracy of 0.1 kg, wearing underwear, with a calibrated electronic scale (SECA, Birmingham, UK). We used the BMI values to define overweight (including obesity) according to the sex- and age-specific reference values of the International Obesity Task Force (IOTF) which were derived from six large nationally representative cross-sectional surveys from different countries (17). We excluded all 2805 children aged 0–2 years, because child's length was measured in either lying or standing mode in this age group in the KiGGS data (depending on the child's skills or behaviour), leading to a potential bias in BMI measurements, while in older children height was measured in standing mode only, and because the IOTF values do not pertain to children <2 years.

Information on sociodemographic covariates and life style factors was obtained from a self-administered questionnaire from parents. For non-German families, questionnaires in their native languages were provided. Migration status was defined based on parental origin and nationality (18). Parental BMI was calculated from self-reported height and weight at interview and categorized as overweight ($\geq$25 kg/m^2) and obese ($\geq$30 kg/m^2). Socioeconomic status (SES) was classified based on the parents' professional status, income, and educational achievements and assigned low, middle, or high according to the parent with the higher status (19). Maternal smoking in pregnancy was documented in three categories (never, occasionally, or regularly). For further questions regarding pregnancy, mothers were encouraged to consult their "maternity pass." In Germany, a "maternity pass" is issued to every pregnant woman at her first pregnancy-related visit to the gynecologist for complete documentation of antenatal care visits, including regular weight measurements. Mothers were asked to consult their maternity pass to answer the question how much weight they gained during the pregnancy with the index child. In accordance with a previous publication, we defined a gestational weight gain of >17 kg as high, as this corresponded with the upper quartile of gestational weight gain in the KiGGS data (20). We defined occurrence of gestational diabetes mellitus (GDM) as a positive answer to the question "Has diabetes or gestational diabetes been diagnosed during this pregnancy?" (i.e., the pregnancy with the index child). There was no further question related to diabetes before or after pregnancy. Small-for-gestational-age and large-for-gestational-age were defined in terms of birth weight below or above the respective national 10th or 90th sex- and gestational-age-specific birth weight percentile (21).

We used the following set of predictors with known associations with childhood overweight (4,22-24): Migration status, low parental SES, older biological siblings, maternal smoking during pregnancy (never or any), breastfeeding (never or any), overweight or obesity of one or both parents, high gestational weight gain, occurrence of GDM in the index child's pregnancy, small-for-gestational-age and large-for-gestational-age. The analyses were restricted to children living with their biological mother, with available anthropometric measurements and with full information on all a priori selected predictors, yielding a final sample size of $n = 8981$. We calculated separate analyses for 3–6-year-old ($n = 2673$), 7–10-year-old ($n = 2672$), and 11–17-year-old children ($n = 3636$), corresponding with preschool, primary school, and secondary school age in Germany, respectively. The rationale to use these three subgroups was that prevalence rates of overweight have been shown to differ considerably between these groups (12,13).

Classification trees are a statistical technique which is helpful to assess and depict the associations between an outcome variable (in this case overweight) and a number of explanatory variables, implicitly considering potential interactions between these variables. Therefore, classification trees provide a powerful tool in questions related to decision-making (25). The classification tree analyses were performed in accordance with our previous study (9): At each node, we calculated 2 $\times$ 2 contingency tables according to child's overweight and each binary predictor remaining at the respective node, together with the corresponding chi-square statistics with one degree of freedom. The maximum chi-square statistic for all bipartitions at the respective node was considered as optimality criterion for every split. A further partition of a subset was rejected if the size of the subset was less than the square root of the initial sample size used for calculation of the respective classification tree or if there was no P-value of <0.05 with respect to any association at this point (26). We calculated 95% binomial confidence intervals (CIs) for the prevalence of overweight in all subgroups with weighted estimates accounting for the unequal inclusion probabilities. The clustering of the children within the primary sample points (communities) was not accounted for in the analysis. For the sake of clarity, all splits of a tree are displayed only up to level three (i.e., three splits). Further partitions are only shown if these resulted in a subgroup with an overweight prevalence of >50% or in the subgroup with the lowest prevalence identified in the respective tree. To get an estimate of the strength of an association in a particular subgroup, we calculated the prevalence ratio, dividing the prevalence estimate for overweight in the respective subgroup by the prevalence of overweight in the whole age group. For each prevalence ratio, we calculated P-values based on chi-square tests.

All calculations were carried out with SAS 9.2 (SAS Institute Inc, Cary, NC), using the *freq* and *surveyfreq* procedures, respectively.

Results

The prevalence of overweight differed between age groups, with 10.0 (95% CI: 8.9, 11.2)% in pre-schoolers, 18.3 (16.8, 19.8)% in 7–10-year-old children and 20.1 (18.9, 21.3)% in 11–17-year-olds (Table 1, Figures 1–3). Exposure rates to the risk factors examined were relatively similar between age groups (Table 1).

In preschool children, the subgroup with the highest overweight risk were migrant children with at least one obese parent, with a prevalence of 36.6 (22.9, 50.4)%, corresponding with a prevalence ratio of 36.6/10.0 = 3.66 (2.29, 5.04) ($P < 0.0001$, Figure 1). This subgroup contained 17 overweight children out of 283 overweight children (6.0%) in the whole preschoolers group. With 4.7 (3.2, 6.1)% [prevalence ratio: 0.47 (0.32, 0.61), $P < 0.0001$], the lowest prevalence was observed in children whose parents were not overweight and had no migration background.

TABLE 1 Study characteristics of the data analyzed, stratified by children's age

	Preschool children ($n = 2673$)		Primary school children ($n = 2672$)		Secondary school children ($n = 3636$)	
	n	%	*n*	%	*n*	%
Child's overweight	283	10.0	506	18.3	736	20.1
Male sex	1339	51.3	1383	52.2	1860	50.9
Migration status	226	11.2	122	4.9	126	3.7
Low parental SES	553	20.4	529	20.0	685	18.8
Smoking in pregnancy	417	16.0	399	16.2	508	14.6
Gestational weight gain >17 kg	700	25.0	555	19.8	612	16.2
GDM	74	2.7	60	2.4	67	1.7
Small-for-gestational-age	232	8.8	290	11.0	425	11.4
Large-for-gestational-age	286	10.6	247	8.9	305	7.9
Breastfeeding	2204	81.8	2168	79.7	2811	76.6
Older biological siblings	1473	56.9	1460	54.4	1941	52.8
Mother is overweight	847	30.8	871	33.2	1350	36.7
Father is overweight	1500	54.9	1574	59.4	2228	60.9
Mother is obese	254	8.9	278	10.7	461	12.4
Father is obese	293	10.1	310	11.9	499	13.3

GDM, gestational diabetes mellitus; SES, socioeconomic status.
Child's overweight was classified using sex- and age-specific cut-off values. Small-for-gestational-age and large-for-gestational-age were defined in terms of birth weight below or above the respective national 10th or 90th sex- and gestational-age-specific birth weight percentile. Migration status was defined based on parental origin and nationality. Proportions were calculated based on weighted estimates accounting for unequal inclusion probabilities.

In children at primary school age, parental weight status was again the first-split criterion (Figure 2). The prevalence of overweight increased to 35.4 (31.2, 39.7)% in children of obese parents and to 57.9 (46.6, 69.3)% [prevalence ratio: 3.16 (2.55, 3.79), $P < 0.0001$], if these were additionally born large-for-gestational-age. However, the latter subgroup contained only 44/506 = 8.7% of all overweight cases recorded in this age group. By contrast, the prevalence of overweight was below 25% in any subgroup of children of nonobese parents, with the lowest risk in children of nonoverweight nonmigrant parents whose mothers did not smoke during pregnancy [5.5 (3.7, 7.3)%, prevalence ratio: 0.31 (0.20, 0.40), $P < 0.0001$].

Similar results were found in 11–17-year-old children (Figure 3). Overweight prevalences of >50% were found in children of overweight or obese parents, if they were born large-for-gestational-age and not breastfed [63.0 (46.4, 79.7)%, prevalence ratio: 3.13 (2.43, 3.75), $P < 0.0001$], or if they were from families with low SES and additionally exposed to other risk factors [migration status: 56.0 (39.6, 72.4)%; smoking in pregnancy and no older siblings: 62.2 (48.9, 75.4)%]. These subgroups together comprised only 56/736 = 7.6% of the observed overweight cases in 11–17-year-olds. The most protective combination in this age group was no parental overweight together with high SES and no GDM [6.6 (5.1, 8.1)%, prevalence ratio: 0.33 (0.25, 0.40), $P < 0.0001$].

Discussion

Our analyses identified specific risk groups amongst school children with an overweight risk >50%. Using the optimal subset of early life predictors for each age group, the prediction of

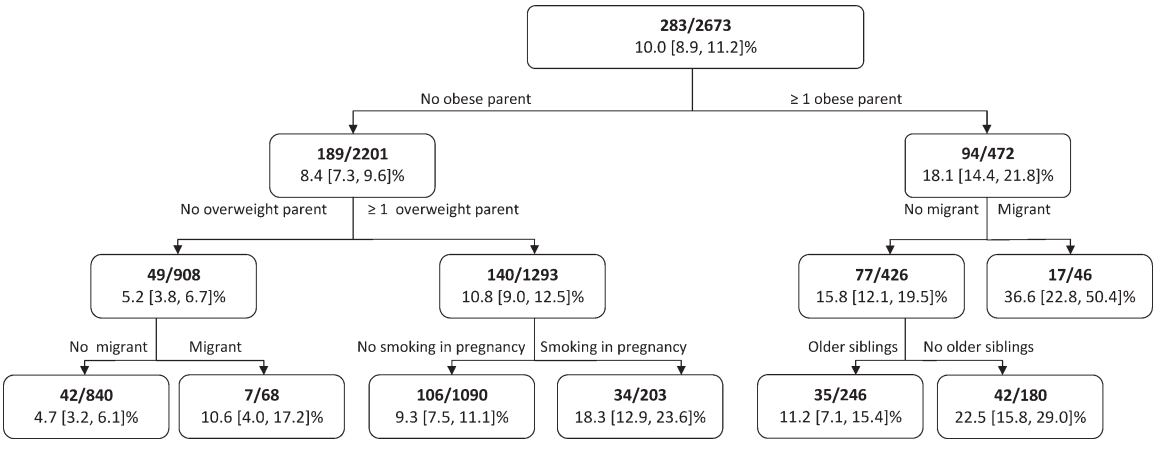

FIGURE 1 Classification tree for prevalence of overweight (95% confidence intervals) in 3–6-year-old children. Proportions are based on weighted estimates accounting for the sampling design.

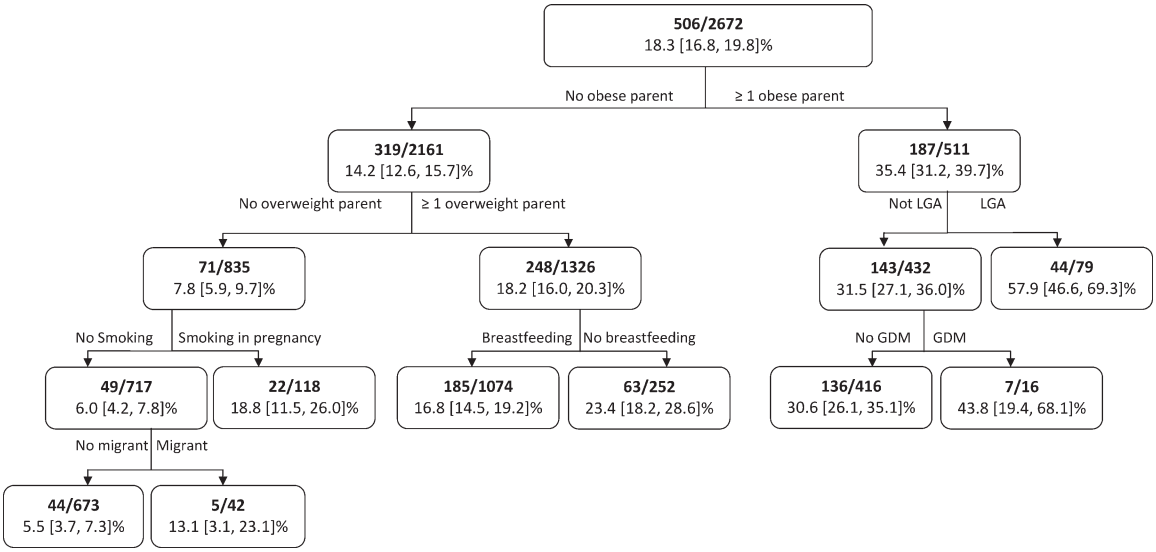

FIGURE 2 Classification tree for prevalence of overweight (95% confidence intervals) in 7–10-year-old children. Proportions are based on weighted estimates accounting for the sampling design. GDM, gestational diabetes mellitus; LGA, large for gestational age.

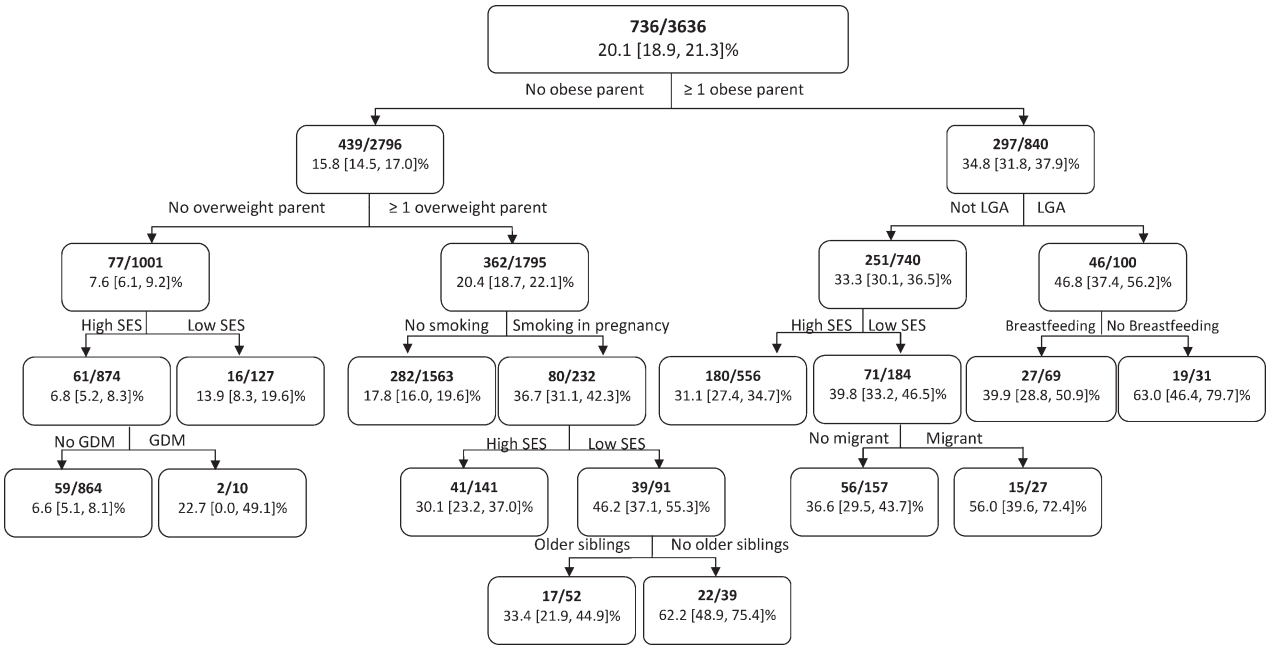

FIGURE 3 Classification tree for prevalence of overweight (95% confidence intervals) in 11–17-year-old children. Proportions are based on weighted estimates accounting for the sampling design. GDM, gestational diabetes mellitus; LGA, large for gestational age; SES, socioeconomic status.

overweight could be increased by about three times, as indicated by respective prevalence ratios. For example, the prevalence of overweight in preschoolers increased from 10.0% overall to 36.6% in migrant children with at least one obese parent. However, prevalence ratios of this size were found only in small subgroups, containing <10% of all overweight cases in the respective age group.

Thus, even if highly effective early prevention programs might be developed for these subgroups, they would not be expected to contribute to a considerable reduction in the overweight prevalence on a population level. Interestingly, another recent analysis indicated that early preventive measurements would be expected to prevent <10% of all overweight cases (27). Future efforts might therefore focus on intervention rather than on early prevention strategies.

Parental obesity was the first split criterion and thus the most important predictor in all age groups. Parental obesity is likely to represent a combination of environmental and genetic factors, as the latter are also known to contribute considerably to one's overweight risk (28). Unfortunately, we were not able to include known obesity-related genes to improve prediction, as no genetic information has been collected within the KiGGS study. However, it appears doubtful that inclusion of such genes would have increased prediction, as this was not the case in a previous study (29).

The prevalence ratios of the most predictive risk factor combinations were similar in each age group. Thus, our data do not allow concluding an increasing impact of early life risk factors by child's age. The fact that we found higher positive predictive rates of overweight in certain subgroups of school children compared to preschoolers might therefore rather have been due to the overall higher prevalence of overweight in school children in our data.

Our data are based on an up-to-date, high-quality national survey on child health in a considerable number of children in Germany, and therefore appear to be generalizable to other high-income countries. A further strength of our study is the broad age spectrum of children considered, which allowed us to explore associations in different age groups. Although KiGGS is basically a cross-sectional study, the data can be interpreted as retrospective cohort data with respect to the impact of early life variables. We focused on pre- and perinatal risk factors and on factors which are likely to be relatively persistent through offspring's childhood and adolescence (such as parental obesity and SES), thus avoiding potential reverse causation issues. Children's weight and height were measured by trained staff, but parents' weight and height had been self-reported. The latter might constitute a limitation of our study, as especially high BMI values tend to be underestimated by self-reporting (30). Therefore, the association of parental obesity and offspring's overweight may have been underestimated. However, as parental obesity was the most important split criterion in all age groups anyway, it seems that this issue is not likely to have biased our main results. There may also be some ascertainment bias with respect to maternal smoking in pregnancy and GDM, as the prevalence of both factors may have been underestimated (31,32). Again, this would likely have led to only a slight underestimation of their associations with offspring's overweight (31,33) and should thus not have affected our main results substantially. Unfortunately, the dataset contained no information about early weight gain, which was an important predictor in our previous study (9). In order to consider potential effects of early catch-up growth (34), we included small-for-gestational-age in our analyses, which was, however, no significant predictor in any classification tree. A potential limitation of classification trees is that they consider classification variables one at a time, that is, without simultaneous adjustment for the other classification variables which have not been selected for partition previous nodes. Thus, this method might be less appropriate than, for example, regression methods to quantify the effect of a certain predictor which might be confounded by another one.

In summary, our results indicate only a limited potential for targeted preventions against overweight in children and adolescents in early childhood. High positive predictive rates were found only in small subgroups, suggesting that even highly effective prevention programs for these subgroups would not considerably reduce childhood overweight on a population level. Further, our data do not indicate an increasing impact of early life risk factors by child's age. O

Acknowledgments

AB had the initial idea to perform the study, supervised the analyses and wrote the first and final draft of the manuscript. DK performed the statistical analyses. AGZ and ASR contributed to subsequent drafts of the manuscript. RvK contributed to the first and final draft of the manuscript.
The German Health Interview and Examination Survey for Children and Adolescents (KiGGS) study was funded by the German Ministry of Health, the Ministry of Education and Research, and the Robert Koch Institute. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Wang Y, Lobstein T. Worldwide trends in childhood overweight and obesity. *Int J Pediatr Obes* 2006;1:11-25.
2. de Onis M, Blossner M, Borghi E. Global prevalence and trends of overweight and obesity among preschool children. *Am J Clin Nutr* 2010;92:1257-1264.
3. Flegal KM, Troiano RP. Changes in the distribution of body mass index of adults and children in the US population. *Int J Obes Relat Metab Disord* 2000;24:807-818.
4. Beyerlein A, Toschke AM, von Kries R. Risk factors for childhood overweight: shift of the mean body mass index and shift of the upper percentiles: results from a cross-sectional study. *Int J Obes (Lond)* 2010;34:642-648.
5. Beyerlein A, Toschke AM, Schaffrath Rosario A, von Kries R. Risk factors for obesity: further evidence for stronger effects on overweight children and adolescents compared to normal-weight subjects. *Plos One* 2011;6:e15739.
6. Anzman SL, Rollins BY, Birch LL. Parental influence on children's early eating environments and obesity risk: implications for prevention. *Int J Obes (Lond)* 2010; 34:1116-1124.
7. Dattilo AM, Birch L, Krebs NF, Lake A, Taveras EM, Saavedra JM. Need for early interventions in the prevention of pediatric overweight: a review and upcoming directions. *J Obes* 2012;2012:123023.
8. Weng SF, Redsell SA, Swift JA, Yang M, Glazebrook CP. Systematic review and meta-analyses of risk factors for childhood overweight identifiable during infancy. *Arch Dis Childhood* 2012;97:1019-1026.
9. Toschke AM, Beyerlein A, von Kries R. Children at high risk for overweight: a classification and regression trees analysis approach. *ObesiRes* 2005;13:1270-1274.
10. Kitsantas P, Gaffney KF. Risk profiles for overweight/obesity among preschoolers. *Early Hum Dev* 2010;86:563-568.
11. Pryor LE, Tremblay RE, Boivin M, et al. Developmental trajectories of body mass index in early childhood and their risk factors: an 8-year longitudinal study. *Arch Pediatr Adolesc Med* 2011;165:906-912.
12. Hughes AR, Sherriff A, Lawlor DA, Ness AR, Reilly JJ. Incidence of obesity during childhood and adolescence in a large contemporary cohort. *Prev Medicine* 2011;52:300-304.
13. von Kries R, Beyerlein A, Müller MJ, et al. Different age-specific incidence and remission rates in pre-school and primary school suggest need for targeted obesity prevention in childhood. *Int J Obes (Lond)* 2012;36:505-510.
14. von Kries R, Reulen H, Bayer O, Riedel C, Diethelm K, Buyken AE. Increase in prevalence of adiposity between the ages of 7 and 11 years reflects lower remission rates during this period. *Pediatr Obes* 2013;8:13-20.
15. Kurth B, Scheidt-Nave C, Schlaud M, et al. The challenge of comprehensively mapping children's health in a nation-wide health survey: design and first results of the German KiGGS-Study. *BMC Public Health* 2008;8:196-204.
16. Kamtsiuris P, Lange M, Schaffrath Rosario A. [The German Health Interview and Examination Survey for Children and Adolescents (KiGGS): sample design, response and nonresponse analysis]. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz* 2007;50:547-556.
17. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ* 2000;320:1240-1243.
18. Schenk L, Ellert U, Neuhauser H. [Children and adolescents in Germany with a migration background. Methodical aspects in the German Health Interview and Examination Survey for Children and Adolescents (KiGGS)]. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz* 2007;50:590-599.
19. Winkler J, Stolzenberg H. [Social class index in the Federal Health Survey]. *Gesundheitswesen* 1999;61 Spec No:S178-S183.
20. von Kries R, Ensenauer R, Beyerlein A, Amann-Gassner U, Hauner H, Rosario AS. Gestational weight gain and overweight in children: results from the cross-sectional German KiGGS study. *Int J Pediatr Obes* 2011;6:45-52.

21. Voigt M, Schneider KT, Jahrig K. [Analysis of a 1992 birth sample in Germany. 1: new percentile values of the body weight of newborn infants]. *Geburtshilfe Frauenheilkd* 1996;56:550-558.
22. Oken E, Gillman MW. Fetal origins of obesity. *Obes Res* 2003;11:496-506.
23. Toschke AM, Vignerova J, Lhotska L, Osancova K, Koletzko B, Von Kries R. Overweight and obesity in 6- to 14-year-old Czech children in 1991: protective effect of breast-feeding. *J Pediatr* 2002;141:764-769.
24. Kalies H, Lenz J, von Kries R. Prevalence of overweight and obesity and trends in body mass index in German pre-school children, 1982–1997. *Int J Obes* 2002;26: 1211-1217.
25. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software; 1984.
26. Lausen B, Sauerbrei W, Schumacher M. Classification and regression trees (CART) used for the exploration of prognostic factors measured on different scales. 1994. In: Dirschedl P, Ostermann R, (eds). *Computational Statistics*. Heidelberg, Germany: Physica-Verlag; 1994, pp. 483-497.
27. Plachta-Danielzik S, Kehden B, Landsberg B, et al. Attributable risks for childhood overweight: evidence for limited effectiveness of prevention. *Pediatrics* 2012;130: e865-e871.
28. Beyerlein A, von Kries R, Ness AR, Ong KK. Genetic markers of obesity risk: stronger associations with body composition in overweight compared to normal-weight children. *PloS One* 2011;6:e19057.
29. Morandi A, Meyre D, Lobbens S, et al. Estimation of newborn risk for child or adolescent obesity: lessons from longitudinal birth cohorts. *PloS One* 2012;7:e49919.
30. Stommel M, Schoenborn CA. Accuracy and usefulness of BMI measures based on self-reported weight and height: findings from the NHANES & NHIS 2001–2006. *BMC Public Health* 2009;9:421.
31. Beyerlein A, Nehring I, Schaffrath Rosario A, von Kries R. Gestational diabetes and cardiovascular risk factors in the offspring: results from a cross-sectional study. *Diabet Med* 2012;29:378-384.
32. Shipton D, Tappin DM, Vadiveloo T, Crossley JA, Aitken DA, Chalmers J. Reliability of self reported smoking status by pregnant women for estimating smoking prevalence: a retrospective, cross sectional study. *BMJ* 2009;339:b4347.
33. Beyerlein A, Rückinger S, Toschke AM, Schaffrath Rosario A, von Kries R. Is low birth weight in the causal pathway of the association between maternal smoking in pregnancy and higher BMI in the offspring? *Eur J Epidemiol* 2011;26:413-420.
34. Ong KK, Ahmed ML, Emmett PM, Preece MA, Dunger DB. Association between postnatal catch-up growth and obesity in childhood: prospective cohort study. *BMJ* 2000;320:967-971.