

Children at High Risk for Overweight: A Classification and Regression Trees Analysis Approach

André Michael Toschke, Andreas Beyerlein, and Rüdiger von Kries

Abstract

TOSCHKE, ANDRÉ MICHAEL, ANDREAS BEYERLEIN, AND RÜDIGER VON KRIES. Children at high risk for overweight: a classification and regression trees analysis approach. *Obes Res.* 2005;13:1270–1274.

Objective: Early identification of children at high risk for childhood overweight is a major challenge in fighting the obesity epidemic. We tried to identify the most powerful set of combined predictors for childhood overweight at school entry.

Research Methods and Procedures: A classification and regression trees analysis on risk factors for childhood overweight in 4289 children 5 to 6 years of age participating in the obligatory school entry health examination 2001/2002 in Bavaria, Germany, was performed. Parental questionnaires asked for children's weight at birth and 2 years, breastfeeding history, maternal smoking in pregnancy, parental education, parental overweight/obesity, nationality, and number of older siblings. Overweight was defined according to sex- and age-specific BMI cut-points proposed by the International Obesity Task Force.

Results: Prevalence of overweight was 11% among the entire study population. Although high early weight gain >10,000 grams was found in about one-half of the overweight children, its positive predictive value reached only 25%, indicating that one of four children with a high early weight gain is overweight at school entry. The best reliable set of predictors included high early weight gain and obese parents and accounted for a likelihood ratio of 3.6, with a

corresponding positive predictive value of 40%, and was found in 4% of all children.

Discussion: A combination of predictors available at 2 years of age could improve predictability of overweight at school entry. However, corresponding low positive predictive values indicate a precision of the prediction that might be insufficient for targeting intervention programs for identified high-risk children.

Key words: epidemiology, diet, prevention and control, energy metabolism, feeding behavior

Introduction

Prevalence of overweight and its related morbidity are increasing in industrialized countries worldwide (1–4). In obese children, interventions rarely show satisfying long-term results (5). Early identification of children at high risk for overweight might offer a chance for early preventive measures (6).

Early high weight gain has been identified as a predictor for later overweight (7–11). The predictive power of this single parameter, however, seems to be limited (12). Multiple parameters might improve early diagnosis of high risk children. Classification and regression trees (CART)¹ analysis, a multivariate analysis method, was applied to provide a useful and precise tool to identify children at high risk for overweight in physician's daily routine without the need of calculations.

Research Methods and Procedures

Study Population and Data Sources

This is a retrospective cohort study of children participating in the obligatory school entry health examination 2001/2002 in six Bavarian communities in Germany (Stadt Ingolstadt, Miesbach, Schwandorf, Kitzingen, Augsburg,

Received for review November 16, 2004.

Accepted in final form May 2, 2005.

The costs of publication of this article were defrayed, in part, by the payment of page charges. This article must, therefore, be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Ludwig-Maximilians-University Munich, Division of Pediatric Epidemiology at the Institute of Social Pediatrics and Adolescent Medicine, Munich, Germany.

Address correspondence to A. M. Toschke, Ludwig-Maximilians-University Munich, Institute of Social Pediatrics and Adolescent Medicine, Division of Pediatric Epidemiology, Heiglhofstr. 63, 81377 Munich, Germany.

E-mail: toschke@biostats.info

Copyright © 2005 NAASO

¹ Nonstandard abbreviations: CART, classification and regression trees analysis.

and Günzburg). Parents ($n = 8741$) were asked to fill out a self-completion questionnaire. Children's age ranged from 4 to 7 years. Approximately 80% of the parents (7026 exactly) returned completed questionnaires involving questions on overweight-related physical, sociodemographic, and lifestyle factors. Data on height and weight at birth and 24 months were obtained by pediatricians or general practitioners performing the examinations of the well baby preventive health program offered to all children in Germany. Parents were asked to copy the data from the well baby check up booklets to questionnaires. Data cleaning and outlier detection excluded weight data at birth of <400 or >5200 grams (0.1%) and at 2 years of <8000 or >17000 grams (0.9%). After these exclusions, only five participants (0.07%) had weight gain data in the first 2 years of life <4000 grams. These were excluded, as well as the four participants with a weight gain $>20,000$ grams. Data were linked with children's stature, and weight was measured in light clothing without shoes.

The analysis was confined to children at least 5 years of age but <7 years of age (234 exclusions) and with full information on anthropometric measures (355 exclusions) and potential predictors available at 2 years of age: high infant weight gain in the first 2 years of life (8–12) (927 exclusions), birth weight (13) (147 exclusions), breastfeeding (14,15) (212 exclusions), maternal smoking in pregnancy (16–19) (212 exclusions), parental education (20) (296 exclusions), parental overweight or obesity (21,22) (1091 exclusions), nationality (23) (25 exclusions), and information on having older siblings (15) (683 exclusions; multiple reasons for exclusion possible). After exclusions, data for 2270 girls and 2219 boys (total $n = 4289$) were available for analysis.

Measures

Overweight and obesity were defined according to sex- and age-specific BMI cut-points proposed by the International Obesity Task Force (24), which are equivalent to the widely used cut-points of 25 kg/m² for adult overweight and cut-points of 30 kg/m² for adult obesity.

The following variables were considered as potential predicting factors for childhood overweight: for weight gain, the cut-point with the highest predictability for overweight at school entry (10,000 grams) (12) was used to differentiate between high and normal weight gain. Two possible cut-points were used for high birth weight, 3800 and 4000 grams, and parental overweight and obesity were defined as BMI of at least one parent ≥ 25 (overweight) or ≥ 30 kg/m² (obesity), respectively. Nationality was differentiated between German and non-German nationality. Number of older siblings was dichotomized into none vs. having older siblings. Parental education was divided into five categories, from <8 years of school up to >13 years of school.

Statistical Analysis

The prevalence of overweight/obesity and 95% exact confidence limits were calculated based on the binomial distribution (25). CART is an approach designed to model the relationship between a response (overweight in our case) and explaining factors measured on different scales (26). Trees are a useful way to express knowledge that may be of help in decision-making. CART analysis allows for all possible interactions and adjustments. The maximum of a two-sample statistic with an asymptotic χ^2 distribution with one degree of freedom for all bipartitions or the minimal p value, respectively, was considered as optimality criterion of every split. Appropriate Bonferroni-corrected χ^2 tests were two sided (26). A further partition of a subset was rejected if the size of the subset was less than the criterion $n_{\min} = \sqrt{4289} = 65.5$ or if the minimal p value was greater than the prespecified value $p_{\text{stop}} = 0.05$ (26).

All calculations were carried out with the software package SAS version 8.2 (SAS Institute, Cary, NC).

Results

At school entry, 12.0% (95% CI, 10.6 to 13.5) of girls and 9.9% (95% CI, 8.7 to 11.2) of boys were classified as overweight. The distributions of predictors are shown in Table 1. For example, a high weight gain ($>10,000$ grams) in the first 2 years of life was observed among $n = 984$ (23%) children (Table 1).

The results of the CART analysis are shown in Figure 1. In the CART analysis, the higher (lower) prevalence after partition corresponds to the positive predictive value (false negatives) because of a positive (negative) test result. A positive predictive value indicates the probability of being ill after a positive test result.

The overall average prevalence for being overweight was 11%. High weight gain in the first 2 years of life was the best predictor for overweight at school entry (Figure 1). Only 7% of the children with a weight gain $<10,000$ grams in the first 2 years of life compared with 25% among children with a high weight gain were overweight (Figure 1). Prediction was improved by further partitions. The lowest prevalence for overweight at school entry was observed among the subgroup of children without a high early weight gain, in the absence of overweight or obese parents, with German nationality, and who were breast-fed (2%; $n = 990$). This corresponds to a negative predictive value of $100\% - 2\% = 98\%$ (base value, $100\% - 11\% = 89\%$) and a likelihood ratio of $2\%/11\% = 0.18$, indicating that the overweight prevalence in the entire study population was reduced to the fifth part among this subgroup.

On the other hand, the combination of a high early weight gain, non-obese or non-overweight parents, parental education <10 years, and birth weight ≥ 3800 grams accounted for the highest overweight prevalence at school entry (positive predictive value; 73%; Figure 1). The respective like-

Table 1. Distribution (prevalence) of potential predictors among complete cases ($n = 4289$)

Predictor	Number	Prevalence (%)
Never breastfed	977	22.8
Breastfed up to 1 month	1718	40.1
Birth weight ≥ 3800 grams	808	18.8
Birth weight ≥ 4000 grams	417	9.7
Weight gain in the first 2 years of life $\geq 10,000$ grams	978	22.8
BMI of at least one parent ≥ 25 kg/m ²	2763	64.4
BMI of at least one parent ≥ 30 kg/m ²	640	14.9
Non-German nationality	163	3.8
No older siblings	1996	46.5
Maternal smoking in pregnancy	883	20.6
Years of parental education		
<8	34	0.8
8 to <10	1195	27.9
10 to <12	1606	37.4
12 to ≤ 13	573	13.4
>13	915	21.3

likelihood ratio of $73\%/11\% = 6.6$ was high, and the subgroup contained only $n = 11$ children (0.3% of entire population). High early weight gain and obese parents accounted for an overweight prevalence (positive predictive value) of 40% among $n = 163$ (4% of entire population) children (likelihood ratio, $40\%/11\% = 3.6$).

A similar tree could be observed for obesity at school entry but with additional splits for smoking in pregnancy as a risk factor for offspring's obesity (data not shown). Prediction of obese children was best among children with a high early weight gain and obese parents, with a positive predictive value of 16%. Although the corresponding likelihood ratio was slightly higher for obesity (5.3) compared with overweight (3.6), the positive predictive value for prediction of obesity at school entry (16%) was lower because of the smaller overall a priori prevalence of obesity (3%).

Discussion

Identifying children at high risk for overweight at school entry by means of predictors detectable at 2 years of age yielded best likelihood ratios of 0.18 (negative) and of 3.6

and 6.6 (positive). Although high early weight gain $>10,000$ grams was found in about one-half of the overweight children, its positive predictive value reached only 25%, indicating that one of four children with a high early weight gain became overweight at school entry. A combination of normal weight gain, in the absence of overweight or obese parents, German nationality, and breastfeeding observed in one-quarter of the population accounted for the best negative predictive value, with 98% probability for being non-overweight at school entry compared with an a priori probability of 89% in the entire population. Weight gain $>10,000$ grams and obese parents accounted for the best reliable positive predictive value of 40%. In this subgroup of 4% of the entire population, two of five children will be overweight at school entry. These results reflect an improved but still insufficient identification of high-risk children even with an optimal set of predictors.

Statistical multivariate methods allow for adjusted assessment of multiple predictors. CART analysis is a multivariate standard method that additionally may provide predictive values (26) and has been used in a number of previous studies (27–30). Predictive values are essential for objective evaluation of the predictive potential of tests under consideration for the general population or test results on the individual level. Furthermore, decision trees provide a useful and precise tool for decision-making in the physician's daily routine by simple visual assessment of disease probability without the need of any calculations.

The insufficient identification of high-risk children even with an optimal set of predictors in our data is unlikely to be explained by biased estimates. In fact, our results are in excellent accordance with other studies on predictors for overweight. Like others, we could identify a high weight gain (8–11), parental overweight and obesity (21,22), lack of breastfeeding (14,15), parental education <10 years (20), high birth weight (13), having older siblings (15), non-German nationality (23), and maternal smoking in pregnancy (16–19) as predictors for later overweight or obesity.

The overall poor positive predictive values need some discussion. For a small subgroup ($n = 11$), a positive predictive value of 73% could be calculated. This might be an artifact, however, because the corresponding split was based on 68 children only, which is just slightly above the final split criterion of 65.5. Additionally, these results are difficult to explain because non-obese parents were identified as contributing to the risk of childhood overweight.

Childhood overweight was obtained at school entry, the age of adiposity rebound (31). It has been reported that overweight beginning at the age of adiposity rebound (4 to 7 years) increases the risk of persistent overweight and its complications (32). This underlines the importance of detecting risk factors for childhood overweight at the age of school entry.

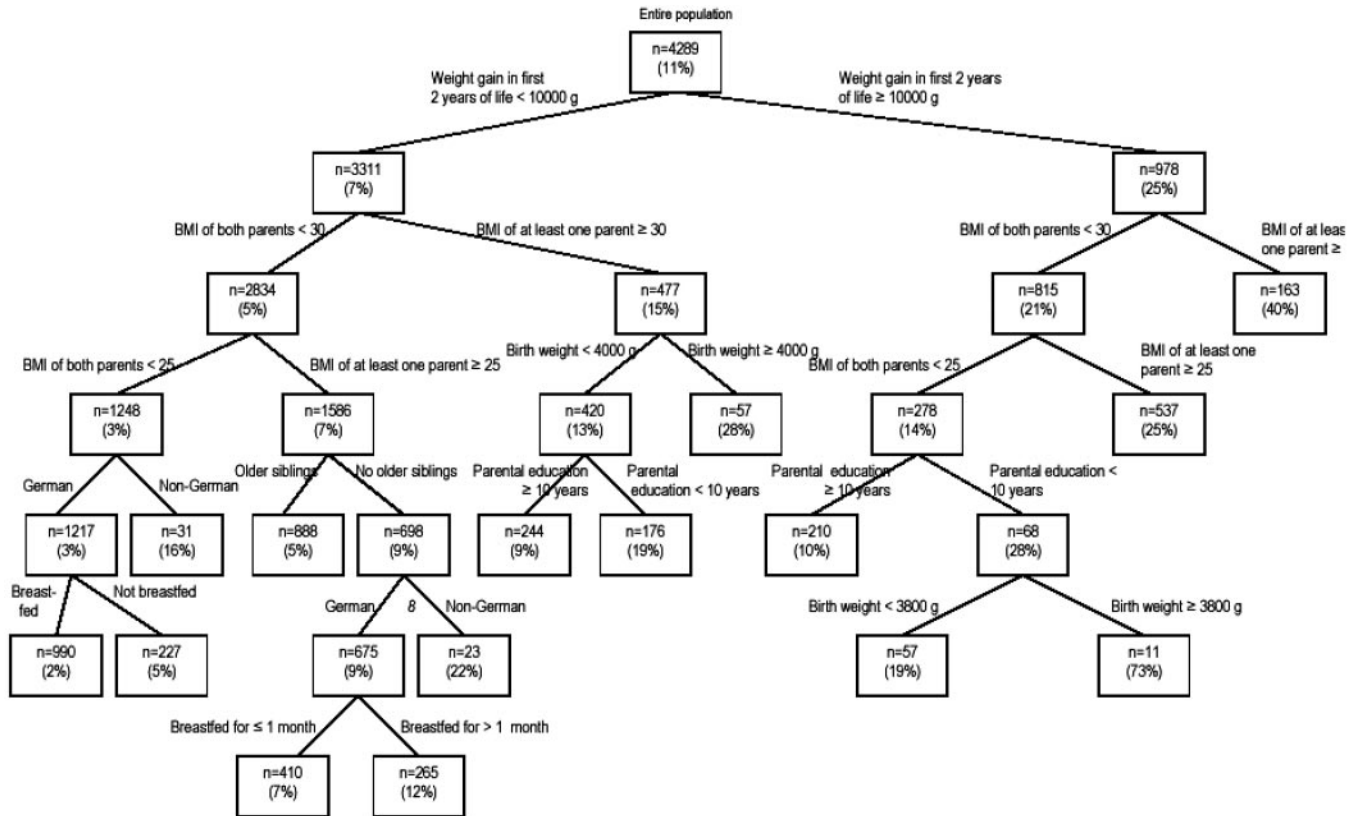


Figure 1: Classification tree for overweight at school entry. Prevalence of overweight in brackets.

Other potential sources of bias have been considered: the overall overweight prevalence of 12% in the population with available data on age, sex, stature, and weight compared with 11% in the study population with full information on predictors does not indicate selection bias. Reverse causality seems to be unlikely because data available at 2 years of age always precedes overweight at school entry.

The weight data at birth and during the first 2 years were copied from doctors' documentation at the time of the respective well baby check-up visit, whereas all other parameters were collected with self-administered questionnaires. Misclassification might result in biased estimates. Any decision-making based on these parameters, however, will be based on documentation in a routine setting. Therefore, our data provide estimates that likely reflect the degree of prediction attainable in physicians' practices.

We chose 2 years of life as the time-point for deciding about the potential usefulness of possible interventions in high-risk children for two reasons. First, the most important predictor, high weight gain, was more powerful at the age of 2 years than at earlier ages (12). Second, children who are 2 years old might still be young enough for application of intervention strategies to modify a potentially underlying risk behavior. It might be possible that later time-points in

the lives of children might allow for better prediction. However, our data set did not allow us to study this question.

In conclusion, although the optimal set of parameters to predict childhood overweight yielded likelihood ratios of 2 to 6, which have been described as indicating small (but sometimes important) changes in disease probability (33), corresponding positive predictive values might be insufficient to allow for decision-making regarding specific interventions targeted at high-risk children: most children would undergo an unnecessary intervention with potential side effects if intervention were based on the sets of predictors assessed in this study.

Acknowledgments

Financial support was obtained from the Bavarian State Ministry of the Environment, Public Health and Consumer Protection, Munich, Germany.

References

- Ogden CL, Troiano RP, Briefel RR, Kuczmarski RJ, Flegal KM, Johnson CL. Prevalence of overweight among preschool children in the United States, 1971 through 1994. *Pediatrics*. 1997;99:E1.

2. **Bundred P, Kitchiner D, Buchan I.** Prevalence of overweight and obese children between 1989 and 1998: population based series of cross sectional studies. *BMJ.* 2001;322:326–8.
3. **Dietz WH.** Childhood weight affects adult morbidity and mortality. *J Nutr.* 1998;128(2 Suppl):411S–4S.
4. **Koletzko B, Girardet JP, Klish W, Tabacco O.** Obesity in children and adolescents worldwide: current views and future directions—Working Group Report of the First World Congress of Pediatric Gastroenterology, Hepatology, and Nutrition. *J Pediatr Gastroenterol Nutr.* 2002;35(Suppl 2):S205–12.
5. **Feldman W.** Periodic health examination, 1994 update: 1. Obesity in childhood. Canadian Task Force on the Periodic Health Examination. *Can Med Assoc J.* 1994;150:871–9.
6. **Lederman SA, Akabas SR, Moore BJ, et al.** Summary of the presentations at the Conference on Preventing Childhood Obesity, December 8, 2003. *Pediatrics.* 2004;114:1146–73.
7. **Eid EE.** Follow-up study of physical growth of children who had excessive weight gain in first six months of life. *Br Med J.* 1970;2:74–6.
8. **Stettler N, Kumanyika SK, Katz SH, Zemel BS, Stallings VA.** Rapid weight gain during infancy and obesity in young adulthood in a cohort of African Americans. *Am J Clin Nutr.* 2003;77:1374–8.
9. **Stettler N, Bovet P, Shamlaye H, Zemel BS, Stallings VA, Paccaud F.** Prevalence and risk factors for overweight and obesity in children from Seychelles, a country in rapid transition: the importance of early growth. *Int J Obes Relat Metab Disord.* 2002;26:214–9.
10. **Stettler N, Zemel BS, Kumanyika S, Stallings VA.** Infant weight gain and childhood overweight status in a multicenter, cohort study. *Pediatrics.* 2002;109:194–9.
11. **Ong KKL, Ahmed ML, Emmett PM, Preece MA, Dunger DB.** Association between postnatal catch-up growth and obesity in childhood: prospective cohort study. *BMJ.* 2000;320:967–71.
12. **Toschke AM, Grote V, Koletzko B, von Kries R.** Identifying children at high risk for overweight at school entry by weight gain during the first 2 years. *Arch Pediatr Adolesc Med.* 2004;158:449–52.
13. **Rasmussen F, Johansson M.** The relation of weight, length and ponderal index at birth to body mass index and overweight among 18-year-old males in Sweden. *Eur J Epidemiol.* 1998;14:373–80.
14. **von Kries R, Koletzko B, Sauerwald T, et al.** Breast feeding and obesity: cross-sectional study. *BMJ.* 1999;319:147–50.
15. **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;141:764–9.
16. **Toschke AM, Montgomery SM, Pfeiffer U, von Kries R.** Early intrauterine exposure to tobacco-inhaled products and obesity. *Am J Epidemiol.* 2003;158:1068–74.
17. **von Kries R, Toschke AM, Koletzko B, Slikker W Jr.** Maternal smoking during pregnancy and childhood obesity. *Am J Epidemiol.* 2002;156:954–61.
18. **Montgomery SM, Ekblom A.** Smoking during pregnancy and diabetes mellitus in a British longitudinal birth cohort. *BMJ.* 2002;324:26–7.
19. **Vik T, Jacobsen G, Vatten L, Bakketeig LS.** Pre- and post-natal growth in children of women who smoked in pregnancy. *Early Hum Dev.* 1996;45:245–55.
20. **Gnavi R, Spagnoli TD, Galotto C, Pugliese E, Carta A, Cesari L.** Socioeconomic status, overweight and obesity in prepubertal children: a study in an area of Northern Italy. *Eur J Epidemiol.* 2000;16:797–803.
21. **Locard E, Mamelle N, Billette A, Miginiac M, Munoz F, Rey S.** Risk factors of obesity in a five year old population. Parental versus environmental factors. *Int J Obes Relat Metab Disord.* 1992;16:721–9.
22. **Stunkard AJ, Sorensen TI, Hanis C, et al.** An adoption study of human obesity. *N Engl J Med.* 1986;314:193–8.
23. **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 Relat Metab Disord.* 2002;26:1211–7.
24. **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.
25. **Clopper CJ, Pearson ES.** The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika.* 1934;26:404–13.
26. **Lausen B, Sauerbrei W, Schumacher M.** Classification and regression trees (CART) used for the exploration of prognostic factors measured on different scales. In: Dirschedl P, Ostermann R, eds. *Computational Statistics.* Heidelberg, Germany: Physica-Verlag; 1994, pp. 483–97.
27. **Rovlias A, Kotsou S.** Classification and regression tree for prediction of outcome after severe head injury using simple clinical and laboratory variables. *J Neurotrauma.* 2004;21:886–93.
28. **Miller PD, Barlas S, Brenneman SK, et al.** An approach to identifying osteopenic women at increased short-term risk of fracture. *Arch Intern Med.* 2004;164:1113–20.
29. **Thwaites GE, Chau TT, Stepniewska K, et al.** Diagnosis of adult tuberculous meningitis by use of clinical and laboratory features. *Lancet.* 2002;360:1287–92.
30. **Ip WF, Dupuis A, Ellis L, et al.** Serum pancreatic enzymes define the pancreatic phenotype in patients with Shwachman-Diamond syndrome. *J Pediatr.* 2002;141:259–65.
31. **Rolland-Cachera MF, Deheeger M, Bellisle F, Sempe M, Guillaud-Bataille M, Patois E.** Adiposity rebound in children: a simple indicator for predicting obesity. *Am J Clin Nutr.* 1984;39:129–35.
32. **Dietz WH.** Critical periods in childhood for the development of obesity. *Am J Clin Nutr.* 1994;59:955–9.
33. **Jaeschke R, Guyatt GH, Sackett DL.** Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA.* 1994;271:703–7.