Cluster Analysis Methods Help to Clarify the Activity–BMI Relationship of Chinese Youth

Keri L. Monda and Barry M. Popkin

Abstract

MONDA, KERI L. AND BARRY M. POPKIN. Cluster analysis methods help to clarify the activity–BMI relationship of Chinese youth. *Obes Res.* 2005;13: 1042–1051.

Objective: To use cluster analysis to create patterns of overall activity and inactivity in a diverse sample of Chinese youth and to evaluate their use in predicting overweight status.

Research Methods and Procedures: The study populations were drawn from the 1997 and 2000 years of the longitudinal China Health and Nutrition Survey, comprised of 2702 and 2641 schoolchildren in the 1997 and 2000 cross-sectional samples, respectively, and 1175 children in the longitudinal cohort. Cluster analysis was used to group children into nonoverlapping activity/inactivity "clusters" that were subsequently used in models of prevalent and incident overweight. Results were compared with traditional models, with activity and inactivity coded separately, to assess whether further insight was gained with the cluster analysis methodology.

Results: Moderately and highly active youth were shown to have significantly decreased odds of overweight in both cross-sectional and longitudinal analyses using cluster analysis. In incident longitudinal models, youth in the high activity/high inactivity cluster had the lowest odds of overweight [odds ratio = 0.12 (0.03, 0.44)]; in contrast, results from traditional models failed to show any significant relationship between overweight and activity or inactivity.

Discussion: Cluster analysis methods allow researchers to simultaneously capture activity and inactivity in new ways. In this comparative study, only with the clustering methodology did we find a significant effect of activity on incident overweight, furthering our ability to examine this complex relationship. Interestingly, no effect of increasing levels of inactivity was observed using either method, indicating that activity seems to be the more important determinant of overweight in this population.

Key words: cluster analysis, physical activity, sedentary, children, incident overweight

Introduction

Research on physical activity, including studies evaluating physical activity patterns and obesity, typically separates measures of activity during work and leisure time and measures of inactivity such as television viewing and video game playing. Clearly for many individuals, physical activity and sedentary behavior covary in meaningful ways and are not randomly distributed; that is, they are correlated within the same individual. In studying these factors separately, we may miss major overall dimensions of activity and inactivity that affect obesity. While consideration of multidimensionality or patterning of health behaviors has been used regularly in other disciplines, the physical activity field has not yet embraced methodologies to deal with this issue. This paper presents one method for examining these co-occurring behaviors in a joint manner rather than as independent factors and examines how this approach differs from the more traditional method of studying the activity-inactivity-obesity nexus.

Underlying the concern for this study is the worldwide obesity pandemic. Obesity is particularly common among adults but is increasingly found among children and adolescents (1–8). Reduction in physical activity is ascribed as one of the major causes for the increased energy imbalance among children (9,10), yet because of the paucity of longitudinal data on youth physical activity patterns in developing countries, this hypothesis remains relatively unstudied in these locations. Furthermore, with the exception of a few studies examining television viewing and obesity, results of studies evaluating activity and/or inactivity and their relationship to youth obesity are inconsistent and are far from

Received for review July 12, 2004.
Accepted in final form March 15, 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.
Carolina Population Center and School of Public Health, Department of Nutrition University of North Carolina, Chapel Hill, North Carolina.
Address correspondence to Barry M. Popkin, Carolina Population Center, University of North Carolina, 123 W. Franklin Street, Chapel Hill, NC 27516-3997.
E-mail: popkin@unc.edu

clearly defined. Cross-sectional and longitudinal studies have shown that obese children spend less time in moderate and vigorous physical activity and that children who engage in the least vigorous physical activity or the most television viewing tend to be the most overweight (11–18). While it may seem intuitively obvious to combine the roles of activity and inactivity or to examine them together, the focus instead has been on examining each dimension separately, usually in separate studies. Thus, an important methodological consideration is to evaluate the use of methods that combine activity and inactivity in new ways.

This study uses longitudinal data on Chinese youth to examine alternate ways of representing physical activity and inactivity patterns as exposures for child obesity. Cluster analysis can be used to place children in mutually exclusive groups delineated by both their activity and inactivity behaviors. This technique has been used frequently in health behavior–health outcome research, particularly in studies of dietary patterns and relevant health outcomes (19–22); however, thus far in our knowledge, non-hierarchical clustering methodologies have not been similarly used to create patterns of activity and inactivity. Data reduction methodologies such as factor analysis have been used in past studies using activity and inactivity data to try to limit the number of explanatory variables and, thus, provide a more reliable measure of overall activity (23), but this method does not give each child a unique pattern and is more difficult to interpret and use as a basis for developing overall patterning measures. For further information specific to pattern analysis methodology, see studies by Hu (24) and Togo et al. (25). For this paper, we used select measures of activity and inactivity to form patterns, or clusters, potentially addressing an important methodological consideration and evaluated their ability to predict overweight status.

Research Methods and Procedures

Study Population

The China Health and Nutrition Survey (CHNS)[1] is a longitudinal monitoring survey with data collections in 1989, 1991, 1993, 1997, and 2000 (26–29). The study population represents a diverse mixture with respect to socioeconomic factors such as income, employment, and education, as well as other health, nutritional, and demographic factors. Additional details are available on the Internet (http://www.cpc.unc.edu/projects/china). This paper focuses on data collected in the survey years of 1997 and 2000. Detailed physical activity data on children and adolescents were first collected in 1997, with the first follow-up data collected in 2000.

[1] Nonstandard abbreviations: CHNS, China Health and Nutrition Survey.

The analysis sample of children and adolescents consisted of subjects between 6 and 18 years of age who were enrolled in school at the time of data collection. The 1997 and 2000 cross-sectional samples were comprised of 2282 and 2174 subjects, respectively. A total of 1175 individuals between 6 and 18 years of age who were surveyed in both years made up the longitudinal cohort. Children not attending school were excluded before analysis because of differences in the survey questions asked of those respondents attending school vs. those not attending school. Further exclusions were made on the basis of incomplete or inconsistent height, weight, demographic, and activity data.

Variable Creation

Respondents attending school were asked to report hours and minutes per week spent in each of a number of active and inactive pursuits; active pursuits were queried separately for time spent outside of and during school hours, whereas inactive pursuits were queried only regarding time before and after school. Descriptive analyses showed that the majority of school-aged youths did not participate in physical activity outside of school hours. For those activities completed while at school, the questionnaire did not distinguish between organized physical education and informal activities at recess or lunchtime. Activity variables were categorized as 0, $>$0 to $<$30, and $\geq$30 min/d; inactivity variables were categorized as 0, $>$0 to $<$60, and $\geq$60 min/d. For further descriptive details regarding activity and inactivity measures in the CHNS, see Tudor-Locke et al. (30).

Four physical activity variables (gymnastics in school, team sports in school, track or swimming in school, and active commuting) and two physical inactivity variables (television/video watching, and reading/writing/drawing/homework) were used in final cluster formation. We did not use the remaining activity/inactivity variables because of the low prevalence in the sample, where the overwhelming majority of respondents indicated 0 min/wk participation. Initially, all activity and inactivity variables were included in cluster formation because we thought there might be an important subset of children who participated in these specific less prevalent activities; inclusion, however, did not illuminate such a pattern and increased the noise of the resulting clusters to a degree that made them difficult to interpret. Thus, we limited our usage to variables that had $>$10% prevalence within the analysis sample. The majority of the excluded variables were those representing physical activity and not inactivity. Specifically, because the majority of youths did not engage in physical activity outside of school hours, these activities were excluded from cluster formation. Similarly, the variables querying the inactive pursuits "board games" and "listening to the radio" were excluded from cluster formation.

Height and weight were measured directly by trained health workers following a standard protocol. Body weight was measured in lightweight clothing to the nearest 0.1 kg and height to the nearest 0.1 cm. BMI was calculated as weight (kilograms) over height squared (meters squared). Individual BMI percentile was computed by standardizing to the Centers for Disease Control 2000 age- and sex-specific growth charts for individuals (31). For this paper, overweight was considered to be ≥85th percentile.

Cluster Analysis Strategy

Cluster analysis methods attempt to find natural groupings in data by clustering respondents into groups (clusters) based on their variable values (24,32). Partition, or nonhierarchical, cluster analysis methods break the data into a user-specified number of nonoverlapping groups. Computationally, the goal is to minimize variability within clusters and maximize variability between clusters. Cluster variability is measured with respect to their means for classifying variables such that when more than one variable is used, the distances between clusters are measured in multidimensional space. After generation, cluster membership becomes a variable with which to perform further analyses—in our case, the primary exposure variable.

We created physical activity/inactivity clusters using the 1997 data, specifying three-, four-, five-, and six-cluster solutions. There was little benefit in using the less parsimonious five- or six-cluster solution over the four-cluster solution. Inspection of the resulting clusters revealed that the inactivity variables were more apt to differentiate between clusters than activity variables. Thus, to enable further differentiation on the activity level, we reran the clustering methods using only the four activity variables and achieved substantial differentiation with a three-cluster solution. The four inactivity clusters were cross-tabbed with the three activity clusters, and relevant categories based on level (low, moderate, high) of activity and inactivity were grouped to form a final seven-cluster solution. The cluster analysis algorithms provide the researcher with a set of clusters that are internally consistent but completely driven by the data and somewhat complex to interpret. From earlier studies, we learned that a critical final step is to use the clusters to design groupings—in this case, based on the variables involved in the cluster analysis—that are readily interpretable (19,20,22,33). It is important to note that cluster analysis will always produce a classification solution whether or not it makes any intuitive sense or has any predictive benefit. Thus, the predictive validity of the clusters was assessed cross-sectionally using overweight status as a criterion variable that is expected to vary across cluster but was not used in cluster generation. All clusters were originally created with the use of PROC FASTCLUS algorithms (version 8.1; SAS Institute, Cary, NC); however, to track activity patterns longitudinally, we manually recoded the computer-generated clusters so that respondents could be assigned to clusters using the same protocol in both survey years.

Data Analyses

Physical activity/inactivity clusters were entered as indicator variables into the model. We first conducted cross-sectional analyses among children 6 to 18 years of age for both 1997 and 2000 to examine the association between overweight status and activity/inactivity cluster. Second, we conducted longitudinal analyses to examine whether activity/inactivity cluster membership in 1997 significantly predicted overweight status 3 years later. Third, to evaluate whether further insight was gained using cluster analysis methods, we ran logistic regression models in which total minutes per day spent in active and inactive pursuits were entered as separate variables into the model (called herein the “traditional” model). Both continuous and categorical coding schemes were evaluated. To test for interaction of the activity and inactivity variables, we performed a likelihood ratio test of the interaction terms. For both the cross-sectional and longitudinal models, we failed to reject the null hypothesis that the coefficients of the interaction terms equaled zero, and, thus, we removed the interaction terms from the models and proceeded with the reduced (main effects) models as our final models. All cross-sectional logistic regression models were adjusted for age, sex, and urban/rural residence; longitudinal models of incident overweight were further adjusted for baseline overweight status.

Results

Selectivity Analysis

The cross-sectional analysis sample from 1997 comprised 1440 boys and 1262 girls. Complete height, weight, demographic, and activity data were available for 1060 (84.0%) girls and 1222 (84.9%) boys, bringing the final total to 2282 individuals. For 2000, the analysis sample was comprised of 1400 boys and 1241 girls. Complete height, weight, demographic, and activity data were available for 1158 (82.7%) boys and 1016 (81.9%) girls, bringing the final total to 2174 individuals. Of the children with complete anthropometric and demographic data in 1997 and 2000, 412 and 310 were, respectively, excluded because they were not in school. In both time periods, children not in school were significantly older, more rural, of lower income, and, in 1997 only, less likely to be overweight. In the longitudinal sample, of the 1678 schoolchildren between the ages of 6 and 18 in both collection periods, 503 were excluded because of missing or inconsistent height, weight, or demographic data, bringing the final total to 1175. Individuals in the longitudinal analysis sample were younger than those in the cross-sectional sample, but no other significant differences were found. Table 1 presents the distributions of sociodemographic and

Table 1. Sociodemographic and anthropometric characteristics and selectivity analysis of children 6 to 18 years of age in the cross-sectional and longitudinal analysis samples from the China Health and Nutrition Survey, 1997 and 2000

	Cross-sectional 1997		Cross-sectional 2000		Longitudinal analysis*
	Analysis (n = 2282)	Excluded (n = 412)	Analysis (n = 2174)	Excluded (n = 310)	(n = 1175)
Sociodemographic characteristics					
Age (years)	11.1 ± 3.1	14.6 ± 4.1†	11.7 ± 2.9	15.2 ± 3.4†	9.7 ± 2.2§
Age group, 6 to 11 years (%)	58.0	20.4†	47.0	12.0†	79.4§
Sex (% girls)	53.6	48.5	53.3	46.1	45.5
Urban or rural residence (% urban)	25.5	17.7†	26.4	16.8†	24.8
Household income: low/middle/high (%)	40.1/34.6/25.3	54.9/28.2/17.0†	37.9/36.5/25.6	53.1/34.6/12.3†¶	39.8/34.6/25.6
Anthropometric measures					
Weight (kg)	34.5 ± 12.3	45.7 ± 14.3†	37.5 ± 13.0	49.4 ± 13.3†	29.3 ± 9.2§
Height (cm)	139.2 ± 16.7	152.1 ± 19.9†	143.7 ± 16.5	156.1 ± 16.8†	131.9 ± 14.1§
BMI (kg/m^2)	17.2 ± 3.2	19.1 ± 2.9†	17.7 ± 3.8	19.9 ± 3.6†	16.5 ± 2.9§
Overweight (%)	8.0	4.6‡	8.5	6.8	9.0
Underweight (%)	13.9	13.4	13.2	11.6	13.2

* Based on 1997 measures.
† $p < 0.001$ from analysis sample within the same year.
‡ $p < 0.05$ from analysis sample within the same year.
§ $p < 0.001$ between cross-sectional 1997 and longitudinal analysis samples.
¶ Frequency based on $n = 130$ because of missing data.

anthropometric characteristics of the 1997 and 2000 cross-sectional analysis and excluded samples and those of the combined longitudinal sample. Age group was defined as <12 years and ≥12 years, a cut-point that has been used in earlier analyses on this sample of Chinese youth (30). It is at this age that Chinese children leave primary school and enter middle school. Between 1997 and 2000, mean height, weight, BMI, and percentage overweight increased, whereas percentage underweight decreased, although none of these differences was significant.

Cluster Breakdown

After clustering, we needed a simple way to identify and represent the clusters in our analyses. Because the clusters did not differentiate on type of activity (e.g., there was not a "gymnastics" cluster), we chose to differentiate on overall amount of time spent in active and inactive pursuits (Figure 1; Table 2). Thus, "low" activity refers to the clusters of children who did not participate in gymnastics, team sports, or track/swimming but did actively commute to school. Furthermore, the differentiation between "moderate" and "high" activity levels was based on a child's participation in team sports and slightly higher average levels of time spent in track/swimming. Inactivity was also more easily described using amount of time spent rather than type of inactive pursuit. All children with either moderate or high levels of inactivity spent time both watching television/videos and reading/homework; they merely differentiated in the average amount of time spent. Oddly, we discerned no "low" inactivity cluster; rather our sample distinguished those individuals who reported no time spent in inactive pursuits. We acknowledge that there is likely something unique about these children, that it is unlikely they spend no time in inactive pursuits, but that we failed to capture them with the relevant explanatory variables available to us. In models using activity/inactivity cluster as the exposure variable, clusters were entered in as indicator variables with the low activity/no inactivity cluster serving as referent. We selected this cluster as the referent because not only did it have the lowest mean level of overall physical activity, but also its members showed the highest levels of prevalent overweight. We felt that these characteristics would aid in the meaningful interpretation of results.

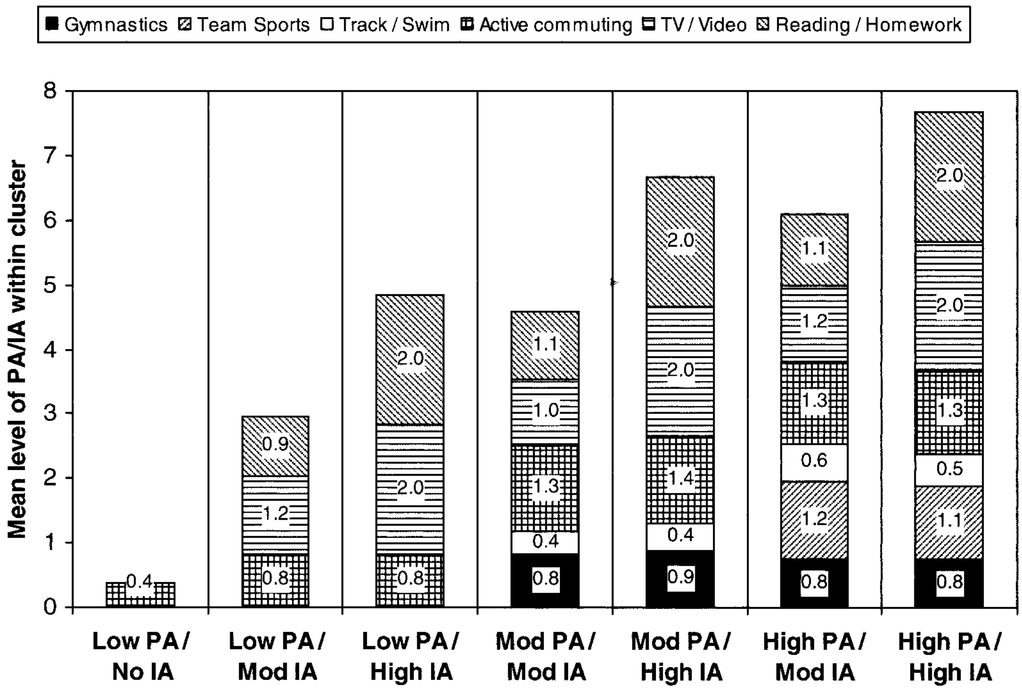

Figure 1: Overall physical activity/inactivity patterns of Chinese schoolchildren in the longitudinal cohort (1997 values). Mod, moderate; PA, physical activity; IA, inactivity.

Table 2 outlines the descriptors and respective frequencies per cross-sectional and longitudinal sample of the final seven clusters these analyses are based on. Cluster breakdown did not change drastically among the three samples; the moderate physical activity/moderate inactivity cluster comprised the majority of individuals, and the low physical activity/high inactivity cluster comprised the least in all three. Figure 1 is a representation of the mean levels of each of the six used activity and inactivity variables per cluster within the longitudinal sample. As seen in Figure 1, the majority of time captured was in inactive pursuits, especially reading and homework, and the most common activity was walking or biking to school. Distribution of the active and inactive variables did not differ substantially for the 1997 or 2000 cross-sectional samples (data not shown).

Table 2. Cluster breakdown with frequencies for the cross-sectional and longitudinal samples

Cluster descriptor	1997 cross-sectional [*N* (%)]	2000 cross-sectional [*N* (%)]	Longitudinal* [*N* (%)]
Mod PA/mod IA	635 (27.8)	600 (27.6)	389 (33.1)
Mod PA/high IA	346 (15.2)	332 (15.3)	235 (20.0)
High PA/mod IA	426 (18.7)	458 (21.1)	207 (17.6)
High PA/high IA	235 (10.3)	344 (15.8)	136 (11.6)
Low PA/mod IA	269 (11.8)	165 (7.6)	79 (6.7)
Low PA/high IA	165 (7.2)	105 (4.8)	56 (4.8)
Low PA/no IA	206 (9.0)	170 (7.8)	73 (6.2)
Total	2282	2174	1175

* Frequencies based on 1997 activity and inactivity values.
Mod, moderate; PA, physical activity; IA, inactivity.

Table 3. Associations of overweight and activity/inactivity cluster from cross-sectional and longitudinal logistic regression models

	Cross-sectional (n = 2282) Activity 1997 and OW 1997*		Longitudinal (n = 1175) Activity 1997 and OW 2000*		Longitudinal (n = 1175) Activity 1997 and OW 2000†	
Cluster descriptor	**OR (95% CI)**	***p***	**OR (95% CI)**	***p***	**OR (95% CI)**	***p***
Low activity						
No inactivity	Reference		Reference		Reference	
Moderate inactivity	0.75 (0.44, 1.30)	0.919	0.23 (0.08, 0.63)	0.005	0.20 (0.07, 0.60)	0.004
High inactivity	0.63 (0.34, 1.19)	0.943	0.48 (0.18, 1.25)	0.133	0.39 (0.13, 1.14)	0.085
Moderate activity						
Moderate inactivity	0.29 (0.17, 0.49)	0.001	0.18 (0.09, 0.37)	0.000	0.23 (0.11, 0.50)	0.000
High inactivity	0.47 (0.27, 0.82)	0.006	0.17 (0.08, 0.38)	0.000	0.20 (0.08, 0.46)	0.000
High activity						
Moderate inactivity	0.36 (0.20, 0.63)	0.022	0.22 (0.10, 0.50)	0.000	0.26 (0.11, 0.62)	0.003
High inactivity	0.16 (0.07, 0.37)	0.000	0.08 (0.02, 0.27)	0.000	0.12 (0.03, 0.44)	0.002

* Controlled for sex, age, and urban/rural residence.
† Additionally controlled for overweight status in 1997.
OW, overweight; OR, odds ratio; CI, confidence interval.

Logistic Analyses of Overweight Status using Activity/Inactivity Cluster

Logistic regression models were used to explore the relationship between activity/inactivity cluster and overweight status both cross-sectionally and longitudinally. All models were controlled for age, sex, and urban or rural residence; incident models were additionally controlled for baseline overweight. Table 3 summarizes these findings. Using the 1997 data in cross-sectional analyses, we found a significant reduction in the odds of overweight among those individuals who were either moderately or highly active compared with those with low levels of activity and no reported inactivity (low physical activity/no inactivity). Those with low levels of activity and either moderate or high levels of inactivity (low physical activity/moderate inactivity and low physical activity/high inactivity) were not different from the referent group. Activity level seems to be driving this relationship because we see no noticeable effect caused by level of inactivity. Results for the cross-sectional relationship between activity/inactivity cluster and overweight in 2000 were similar to those seen in 1997, although none were significant (data not shown). Longitudinal models used to predict prevalent overweight in 2000 from activity/inactivity cluster in 1997 among a cohort of schoolchildren revealed reduced odds of overweight for those moderately or highly active regardless of level of inactivity. Additionally, children in the low physical activity/moderate inactivity cluster were significantly less likely to be overweight than those in the low physical activity/high inactivity cluster, who were not significantly different from the referent cluster. Note, however, that these two clusters are not significantly different from each other; thus, we cannot ascribe the reduced odds of overweight within those who have low activity to decreased inactivity. Nonetheless, this difference is meaningful because the low physical activity/moderate inactivity cluster had, on average, higher mean levels of active commuting than the low physical activity/no inactivity cluster (Figure 1), although they are both labeled as "low" active. The longitudinal results did not change when predicting incident overweight in 2000 from 1997 activity/inactivity cluster by controlling for baseline overweight in the models.

Logistic Analyses of Overweight Status using Traditional Modeling Techniques

To evaluate whether there was any benefit to clustering activity and inactivity, we ran traditional logistic regression models where total times spent in active and inactive pursuits were modeled as separate exposure variables within the same model. We summed the total times spent in the four physical activity variables and the two inactivity variables and tested their respective associations with overweight using both continuous and categorical coding strategies. Total activity was categorized as <30, 30 to 60, and >60 min/d; total inactivity was categorized as <1, 1 to 3, and >3 h/d. Categorizations were based on the univariate distributions of the summary variables. Results are summarized in Table 4. Analyses of the cross-sectional relationship

Table 4. Associations of overweight and total time in active and inactive pursuits from cross-sectional and longitudinal traditional logistic regression models

	Cross-sectional (n = 2282) Activity 1997 and OW 1997*		Longitudinal (n = 1175) Activity 1997 and OW 2000*		Longitudinal (n = 1175) Activity 1997 and OW 2000†	
	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p
Categorical coding						
Total activity (min/d)						
<30	Reference		Reference		Reference	
30 to 60	0.46 (0.32, 0.65)	0.000	0.64 (0.38, 1.06)	0.082	0.85 (0.49, 1.46)	0.550
>60	0.37 (0.23, 0.61)	0.000	0.41 (0.19, 0.91)	0.029	0.67 (0.29, 1.54)	0.350
Total inactivity (h/d)						
<1	Reference		Reference		Reference	
1 to 3	0.87 (0.59, 1.27)	0.464	0.60 (0.35, 1.05)	0.073	0.60 (0.33, 1.07)	0.084
>3	0.97 (0.64, 1.48)	0.893	0.72 (0.39, 1.33)	0.294	0.65 (0.34, 1.26)	0.205
Continuous coding‡						
Total activity	0.93 (0.89, 0.96)	0.000	1.00 (0.99, 1.00)	0.003	1.00 (1.00, 1.00)	0.112
Total inactivity	1.00 (0.99, 1.01)	0.712	1.00 (1.00, 1.00)	0.297	1.00 (1.00, 1.00)	0.277

* Controlled for sex, age, and urban/rural residence.
† Additionally controlled for overweight status in 1997.
‡ Refers to an additional 30 min/wk.
OW, overweight; OR, odds ratio; CI, confidence interval.

(1997) among physical activity, inactivity, and overweight revealed decreased odds of overweight with increasing amounts of activity, in both categorical and continuous models, although in categorical models, there was no significant difference between the 30 to 60 and >60 min/d levels. Curiously, increasing amounts of inactivity also showed an inverse relationship with odds of overweight; however, these relationships were not significant using either categorical or continuous coding. The cross-sectional results did not particularly deviate from those seen using cluster analysis methods. In longitudinal models predicting prevalent overweight, we saw a similar decrease in the odds of overweight with increasing activity, although this relationship was significant only at the >60 min/d category. Additionally, we again saw the curious, although nonsignificant, inverse relationship between increased inactivity and odds of overweight. Arguably, it is incident overweight that we are most concerned with in modeling this relationship, and, thus, it is an important methodological finding that incident longitudinal models, in contrast to similar models using cluster analysis, failed to show any relationship of activity or inactivity, coded in either manner, and overweight.

Discussion

Profound social and economic developments in China over the last 20 years have resulted in shifts in nutrition and physical activity, predominantly in the urban areas, that undoubtedly play a role in the emerging obesity problem (34). Although researchers intuitively understand the energy in–energy out equation, teasing out the relationship among activity, inactivity, and overweight is complex because of its multifactorial nature and many modifying factors, particularly during the growing years of childhood and adolescence. The rapid economic and acculturation transformations in China include increases in motorized transport, which we have shown elsewhere to be associated with increases in adult overweight and obesity (35); thus, individual patterns of activity and inactivity would be expected to vary and be quite heterogeneous depending on a number of socioeconomic and other factors. Respondents were queried about home production activities including cleaning, childcare, gardening, and other chores, and we might expect their values to differentiate among socioeconomic strata; however, although these are important sources of physical activity in other developing countries (36), they do not have the same relevance in China, with a majority of respondents reporting that they do not perform these activities.

Activity and inactivity in Chinese youth take on different forms than in developed countries such as the United States. Chinese children spend very little time in before- or after-school sports and other physical activities; the majority of this time is spent in the rigors of preparing for school because they are under extreme pressures in this regard (30). The most frequently reported energy-expending activities for Chinese schoolchildren are in walking or biking to and from school and sports and exercise that take place during school hours. For example, 76.9% of the longitudinal sample reported participating in physical activity during school, whereas only 8.8% reported participating in any physical activity before or after school hours. In terms of sedentary behaviors, although >90% of households in China own a television set, time spent watching television and videos is still more constrained than in the United States, where there is greater opportunity and allowance for these types of behaviors. However, American-style satellite programming has only recently been introduced on a large scale, and we expect future research by this group and others to find television viewing rapidly increasing among children and, furthermore, to have the type of impact on overweight found in Singapore, Thailand, and Mexico (15,37,38). Furthermore, as television viewing displaces other forms of inactivity (e.g., homework) in the before- and after-school hours of Chinese youth, subsequent research should additionally inform the question on the specificity of television viewing on overweight status.

Using cluster analysis techniques in our longitudinal analyses, we were able to uncover an effect of activity/inactivity patterns on childhood prevalent and incident overweight over and above that which we found using traditional multivariate regression. With clustering techniques, we found that, regardless of level of inactivity, moderate or high levels of activity resulted in reduced prevalence and incidence of overweight in our longitudinal cohort. We were unable to reproduce this relationship using traditional methodologies. Interestingly, we found no measurable increases in benefit with high levels of activity over moderate levels of activity within any strata of inactivity. Lack of an effect at this level could be caused by cluster definition and lack of variability between the two levels. Furthermore, we were unable to distinguish any differentiation in overweight prevalence or incidence on the basis of inactivity level. Earlier cross-sectional studies done by this group also found lack of an effect caused by inactivity in 6- to 11-year-old Chinese children (39). In the analyses herein, lack of an effect may be caused, in part, by the lesser variability seen within inactivity variables; thus, our clusters tell us more about, and are driven in formation by, activity. It is notable that we found activity level to be protective against overweight regardless of level of inactivity. In fact, the lowest odds of incident overweight were observed in the cluster of youth with the highest levels of both physical activity and inactivity. In a population that is likely to increase their sedentary behaviors as the economic transformation goes forward, continued dominance of school-based physical activity and active commuting may provide an important hindrance to increasing levels of overweight and obesity.

Some limitations of this study warrant consideration. We had a substantial number of children lost to follow-up (n = 344). Those lost to follow-up were more likely to be older, more urban, and of higher income (data not shown). While overweight prevalence decreased with increasing age in this sample, there were also increases in overweight in urban, higher-income populations. Thus, it is difficult to predict what effect losing these data points might have had. Additionally, because the physical activity/inactivity clusters that were generated are sample-specific, our results are somewhat limited in overall use and specifically limited to schoolchildren. Another limitation is that the data collection method left an amount of time unaccounted for, and we were unable to assess what type(s) of inactivity our cluster of low physical activity/no inactivity children engaged in. Fortunately these children clustered together and formed a relevant group; however, without being able to define the ways in which they were inactive, we are unable to infer behavioral traits and draw logical conclusions based on these characteristics. Finally, all measures of activity and inactivity were self-reported, and, thus, they are susceptible to the weaknesses of such data collection techniques. Furthermore, parents assisted younger children (<10 years of age) with the survey and may have provided a proxy report of their physical activity and inactivity habits. Although the validity of proxy reports is uncertain, their use is widespread and bypasses the biases that may be associated with recall in younger subjects.

There are also many strengths of this study. We were able for the first time to analyze physical activity and inactivity data in a longitudinal cohort of diverse children from eight provinces in China, an understudied population that is undergoing rapid transition that will likely affect their nutritional status. We were able to show that the cluster analysis methodology is useful in capturing extant behavioral patterns of activity and inactivity, their predictive ability was established by association with overweight status, and we were able to find an effect of activity longitudinally where we could not if restricted to traditional regression methodologies. By examining activity patterns as they relate to overweight status cross-sectionally and longitudinally in an exposure-subsequent outcome or exposure-incidence manner, this paper advances the science in ways previously unstudied. Cluster analysis allows us to simultaneously capture activity and inactivity, dimensions of behavior that are clearly linked, and we have shown for this sample of Chinese youth that the proposed approach of creating overall patterns of activity and inactivity offer major advantages over earlier approaches, where distinct categories of activity and inactivity are examined separately. Employing these methods has furthered our ability to examine the relationship between activity patterns and child overweight status in China and is a potentially valuable approach for others to explore.

Acknowledgments

This study was supported by NIH Grants R01-HD30880 and R01-HD38700. The authors thank Frances L. Dancy for administrative assistance, Cathy Cross for programming support and advice, and Tom Swasey for work on the graphics.

References

1. **Popkin BM, Doak CM.** The obesity epidemic is a worldwide phenomenon. *Nutr Rev*. 1998;56:106–14.
2. **Flegal KM, Carroll MD, Ogden CL, Johnson CL.** Prevalence and trends in obesity among US adults, 1999–2000. *JAMA*. 2002;288:1723–7.
3. **Ogden CL, Flegal KM, Carroll MD, Johnson CL.** Prevalence and trends in overweight among US children and adolescents, 1999–2000. *JAMA*. 2002;288:1728–32.
4. **Hedley AA, Ogden CL, Johnson CL, Carroll MD, Curtin LR, Flegal KM.** Prevalence of overweight and obesity among US children, adolescents, and adults, 1999–2002. *JAMA*. 2004;291:2847–50.
5. **Troiano RP, Flegal KM.** Overweight children and adolescents: description, epidemiology, and demographics. *Pediatrics*. 1998;101:497–504.
6. **Wang Y, Monteiro C, Popkin BM.** Trends of obesity and underweight in older children and adolescents in the United States, Brazil, China, and Russia. *Am J Clin Nutr*. 2002;75:971–7.
7. **Monteiro CA, Conde WL, Popkin BM.** The burden of disease from undernutrition and overnutrition in countries undergoing rapid nutrition transition: a view from Brazil. *Am J Public Health*. 2004;94:433–4.
8. **Livingstone MB.** Childhood obesity in Europe: a growing concern. *Public Health Nutr*. 2001;4:109–16.
9. **Hill JO, Peters JC.** Environmental contributions to the obesity epidemic. *Science*. 1998;280:1371–4.
10. **Steinbeck KS.** The importance of physical activity in the prevention of overweight and obesity in childhood: a review and an opinion. *Obes Rev*. 2001;2:117–30.
11. **Trost SG, Kerr LM, Ward DS, Pate RR.** Physical activity and determinants of physical activity in obese and non-obese children. *Int J Obes Relat Metab Disord*. 2001;25:822–9.
12. **Trost SG, Sirard JR, Dowda M, Pfeiffer KA, Pate RR.** Physical activity in overweight and nonoverweight preschool children. *Int J Obes Relat Metab Disord*. 2003;27:834–9.
13. **Berkey CS, Rockett HR, Gillman MW, Colditz GA.** One-year changes in activity and in inactivity among 10- to 15-year-old boys and girls: relationship to change in BMI. *Pediatrics*. 2003;111:836–43.
14. **Lazzer S, Boirie Y, Bitar A, et al.** Assessment of energy expenditure associated with physical activities in free-living obese and nonobese adolescents. *Am J Clin Nutr*. 2003;78:471–9.
15. **Hernandez B, Gortmaker SL, Colditz GA, Peterson KE, Laird NM, Parra-Cabrera S.** Association of obesity with physical activity, television programs and other forms of video viewing among children in Mexico city. *Int J Obes Relat Metab Disord*. 1999;23:845–54.
16. **Gortmaker SL, Must A, Sobol AM, Peterson K, Colditz GA, Dietz WH.** Television viewing as a cause of increasing obesity among children in the United States, 1986–1990. *Arch Pediatr Adolesc Med*. 1996;150:356–62.
17. **Andersen RE, Crespo CJ, Bartlett SJ, Cheskin LJ, Pratt M.** Relationship of physical activity and television watching with body weight and level of fatness among children: results from the Third National Health and Nutrition Examination Survey. *JAMA*. 1998;279:938–42.
18. **Robinson TN.** Reducing children's television viewing to prevent obesity: a randomized controlled trial. *JAMA*. 1999;282:1561–7.
19. **Haines PS, Popkin BM, Guilkey DK.** Methods of patterning eating behaviors of American women. *J Nutr Educ*. 1990;22:124–32.
20. **Patterson RE, Haines PS, Popkin BM.** Health lifestyle patterns of U.S. adults. *Prev Med*. 1994;23:453–60.
21. **Martikainen P, Brunner E, Marmot M.** Socioeconomic differences in dietary patterns among middle-aged men and women. *Soc Sci Med*. 2003;56:1397–410.
22. **Liese AD, Schulz M, Moore CG, Mayer-Davis EJ.** Dietary patterns, insulin sensitivity and adiposity in the multi-ethnic Insulin Resistance Atherosclerosis Study population. *Br J Nutr*. 2004;92:973–84.
23. **Prochaska JJ, Sallis JF, Sarkin JA, Calfas KJ.** Examination of the factor structure of physical activity behaviors. *J Clin Epidemiol*. 2000;53:866–74.
24. **Hu FB.** Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol*. 2002;13:3–9.
25. **Togo P, Osler M, Sorensen TI, Heitmann BL.** Food intake patterns and body mass index in observational studies. *Int J Obes Relat Metab Disord*. 2001;25:1741–51.
26. **Popkin BM.** Social change and its nutritional impact: the China Health and Nutrition Survey. *China Exchange News*. 1994;22:9–11.
27. **Wang Y, Popkin B, Zhai F.** The nutritional status and dietary pattern of Chinese adolescents, 1991 and 1993. *Eur J Clin Nutr*. 1998;52:908–16.
28. **Wang Y, Ge K, Popkin BM.** Tracking of BMI from childhood to adolescence: a 6-y follow-up study in China. *Am J Clin Nutr*. 2000;72:1018–24.
29. **Popkin BM, Paeratakul S, Ge K, Zhai F.** Body weight patterns among the Chinese: results from the 1989 and 1991 China Health and Nutrition Surveys. *Am J Public Health*. 1995;85:690–4.
30. **Tudor-Locke C, Ainsworth BE, Adair LS, Popkin BM.** Physical activity and inactivity in Chinese school-aged youth: the Chinese Health and Nutrition Survey. *Int J Obes Relat Metab Disord*. 2003;27:1093–9.
31. **Kuczmarski RJ, Ogden CL, Grummer-Strawn LM, et al.** CDC growth charts: United States. *Adv Data* 2000;314:1–27.
32. **Everitt BS, Landau S, Leese M.** *Cluster Analysis.* 4th ed. London: Arnold; 2001.
33. **Bisgrove EZ, Popkin BM, Barba C.** Infant feeding in the Philippines: a cluster analysis approach. *Ecol Food Nutr*. 1989;23:75–90.
34. **Popkin BM.** The nutrition transition and its health implications in lower-income countries. *Public Health Nutr*. 1998;1:5–21.
35. **Bell AC, Ge K, Popkin BM.** The road to obesity or the path to prevention: motorized transportation and obesity in China. *Obes Res*. 2002;10:277–83.

36. **Tudor-Locke C, Ainsworth BE, Adair LS, Popkin BM.** Physical activity in Filipino youth: the Cebu Longitudinal Health and Nutrition Survey. *Int J Obes Relat Metab Disord*. 2003;27:181–90.
37. **Schmidt GJ, Stensel DJ, Walkuski JJ.** Blood pressure, lipids, lipoproteins, body fat and physical activity of Singapore children. *J Paediatr Child Health*. 1997;33:484–90.
38. **Ruangdaraganon N, Kotchabhakdi N, Udomsubpayakul U, Kunanusont C, Suriyawongpaisal P.** The association between television viewing and childhood obesity: a national survey in Thailand. *J Med Assoc Thai* 2002;85(Suppl 4):S1075–80.
39. **Waller CE, Du S, Popkin BM.** Patterns of overweight, inactivity, and snacking in Chinese children. *Obes Res*. 2003;11:957–61.