Factor Analysis of Clustered Cardiovascular Risks in Adolescence

Obesity Is the Predominant Correlate of Risk Among Youth

Elizabeth Goodman, MD; Lawrence M. Dolan, MD; John A. Morrison, MD, PhD; Stephen R. Daniels, MD, PhD

Background—Clustering of cardiovascular (CV) risks begins in childhood, yet studies of the factor structure underlying this clustering have focused on adults. The increasing rates of obesity and type 2 diabetes and the growing importance of metabolic syndrome among adolescents make assessment of CV risk clustering even more urgent in this age group.

Methods and Results—Exploratory factor analysis (principal components analysis) was performed with data from 1578 healthy seventh to 12th graders from the Princeton School District Study, a school-based study in Cincinnati, Ohio. Measured CV risks included cholesterol, triglycerides, fasting insulin and glucose, body mass index (BMI), waist circumference, fibrinogen, and blood pressure. Factor analysis yielded 4 uncorrelated factors (adiposity [BMI, waist, fibrinogen, insulin], cholesterol [LDL and total cholesterol], carbohydrate-metabolic [glucose, insulin, HDL cholesterol, triglycerides], and blood pressure [systolic and diastolic blood pressure]). These factors explained approximately 67% of the total variance. A summary cumulative risk scale was derived from factor scores, and high risk was defined as scoring in the top 5%. Although insulin loaded onto both the adiposity and carbohydrate-metabolic factors, obesity was a much stronger correlate of high cumulative risk (odds ratio = 19.2; 95% CI, 7.6 to 48.5) than hyperinsulinemia (odds ratio = 3.5; 95% CI, 1.8 to 6.8). A sizable proportion (18.5%; n = 12) of those who were at high cumulative risk were not at high risk for any of the individual factors.

Conclusions—The patterning of CV risk clustering seen among adults is present in healthy adolescents. Among youth, obesity is the predominant correlate of cumulative risk. (Circulation. 2005;111:1970-1977.)

Key Words: atherosclerosis ■ obesity ■ pediatrics ■ risk factors ■ metabolic syndrome X
Methods

Sample Description
The Princeton School District Study takes place in the Princeton City School District, a well-defined geographic area with a socioeconomically and racially diverse public school population in the Greater Cincinnati area. There is 1 junior high school (grades 7 to 8) and 1 high school (grades 9 to 12). Students with known diabetes were not eligible for inclusion in this analysis. Of the eligible subjects, 1578 (57.3%) had both a physical examination and fasting blood sample assayed for CV risks. Because we used a cross-validation technique for the factor analysis, ~20% of the cohort (n = 278) was randomly selected as an exploratory sample. The remaining sample (n = 1300) was used as a validation sample and to explore descriptors of the uncovered factors. Table 1 provides a description of these cohorts.

Examination Procedures
Examinations took place in the school setting with the use of trained staff and standardized protocols. Informed consent was obtained from a parent or legal guardian for those aged <18 years and from students aged ≥18 years before the examination. Student assent was also obtained. After a verified 10-hour fast, subjects had a physical examination and blood draw for laboratory assays. Body mass index (BMI) (kg/m²) was calculated from measured height and weight. BMI percentiles, derived from the Centers for Disease Control and Prevention 2000 growth chart standards, were used to classify

<table>
<thead>
<tr>
<th>TABLE 1. Description of Study Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exploratory Sample (n=278)</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Race/ethnicity</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Black</td>
</tr>
<tr>
<td>Hispanic</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Grade</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>Weight status</td>
</tr>
<tr>
<td>Normal</td>
</tr>
<tr>
<td>Overweight</td>
</tr>
<tr>
<td>Obese</td>
</tr>
<tr>
<td>Pubertal stage</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Peri</td>
</tr>
<tr>
<td>Post</td>
</tr>
<tr>
<td>Missing</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
</tr>
<tr>
<td>BMI %</td>
</tr>
<tr>
<td>Waist circumference, cm</td>
</tr>
<tr>
<td>Glucose, mg/dL</td>
</tr>
<tr>
<td>Insulin, pmol/L</td>
</tr>
<tr>
<td>Total cholesterol, mg/dL</td>
</tr>
<tr>
<td>HDL-C, mg/dL</td>
</tr>
<tr>
<td>LDL-C, mg/dL</td>
</tr>
<tr>
<td>Triglycerides, mg/dL</td>
</tr>
<tr>
<td>Fibrinogen, mg/dL</td>
</tr>
<tr>
<td>SBP, mm Hg</td>
</tr>
<tr>
<td>DBP, mm Hg</td>
</tr>
</tbody>
</table>

Values are n (%) or mean±SD (range).

*BMI % = 0.08 was a 14-year-old white male with a weight of 34.2 kg and height of 1.55 m. BMI % = 99.8 was a 13-year-old black male with a weight of 134.3 kg and height of 1.72 m.

†BMI % = 0.01 was a 15-year-old Asian male with a weight of 36.2 kg and height of 1.56 m. BMI % = 99.9 was a 13-year-old black male with a weight of 180.4 kg and height of 1.70 m.
participants as normal weight (BMI <85%), overweight (BMI ≥85% but <95%), or obese (BMI ≥95%). A fiberglass tape crossing over the umbilicus and the superior iliac crests was used to measure waist circumference. The mean of 2 measurements made at the end of a normal expiration was used in analyses. We were able to include blood pressure (BP) measurement only for a convenience subsample of 267 participants (16.9%) because of time constraints in the school setting. The mean of 3 BP readings taken after resting for 5 minutes 2 minutes apart was used. BP was more likely to be measured on those who were overweight (42.0%) or obese (43.9%) compared with youth of normal weight (14.2%, P<0.001) and in those who were hyperinsulinemic (30.5% versus 11.6%; P<0.001). A total of 20.5% of those with BP measurements were in the exploratory sample, and 79.5% (n =212) were in the validation sample. Pubertal stage was defined in a validated method with the use of sex steroid levels and age of menarche for girls and axillary hair distribution for boys.35

**Laboratory Assays**

Plasma insulin concentration was measured by radioimmunoassay with an anti-insulin serum raised in guinea pigs, 125I-labeled insulin as a standard, and a double-antibody method to separate bound from free tracer. Hyperinsulinemia was defined as those scoring in the top quartile.33 Glucose was measured by an enzymatic method. Cholesterol was measured with the Cholesterol/HP kit from Roche Boehringer. A direct measurement of HDL cholesterol (HDL-C) was made with the HDL-C Plus kit from Roche. Lipid profiles were performed on the Hitachi 704. National Cholesterol Education Program performance criteria for accuracy and precision were followed. Triglycerides were measured with a single reagent system from Roche-BMD. LDL cholesterol (LDL-C) was calculated according to the Friedewald equation except in instances in which triglycerides were >350 mg/dL, in which case a direct method was used with the Roche LDL-C Plus reagent. Fibrinogen was measured with the Sysmex CA6000 coagulation analyzer with Dade Behring thrombin reagent (Dade Behring; coefficient of variation 4%). Coefficients of variation were 8% for insulin, 1.6% for glucose, 2.2% for cholesterol, 2.6% for HDL-C, 4% for triglycerides, and 4% for fibrinogen.

**Statistical Analyses**

All analyses were performed with SPSS for Windows.39 For the exploratory factor analysis, we ran principal components analysis (PCA). PCA is a mathematical technique that uses the linear relationships among variables to reduce the collected variables into a smaller number of summary factors that retain as much of the variance in the original variables as possible. PCA assumes that the linear intercorrelations between measured variables reflect a smaller number of interpretable summary factors. The factors are created by a scoring algorithm such that individual variables “load” most strongly onto the factor with which they are the most correlated. Thus, higher factor loadings represent more correlation between the variable and the summary factor. If a single factor explained all the variance, all measured variables would load onto it, each with a factor loading of 1. Usually, multiple factors are extracted, and each variable represents a fraction of the explained variance within that factor. Because factor analysis is often performed on standardized variables, especially when the measurement scale differs between measured items, we created z scores for each measured variable to include in our factor analyses. Three steps are involved: (1) extraction of the factors, which produces the minimum number of factors that retain as much of the total variance in the original data as possible; (2) orthogonal factor rotation to transform the extracted factors into uncorrelated, independent factors, thereby increasing interpretability of the factors; and (3) interpretation based on rotated factor loading. We used an eigenvalue (sum of the squared factor loadings) >1 as the extraction method and varimax rotation. We included only those variables with at least 15% shared variance between the variable and the summary factor (factor loading ≥0.4) in interpreting factors. However, factor scores were based on all variables. We performed the identical PCA on both the exploratory and the validation samples and then compared results across the 2 samples. We ran a third PCA on the subsample of 212 validation sample subjects who also had BP measurements to assess whether addition of BP would alter the factor structure. We did not perform PCA on the exploratory sample subset with BP measures (n=55) because the small number of subjects precluded such analyses. Further analyses relating to the extracted factors were performed only on the larger validation sample.

**Cumulative Risk Scale Creation and Definition of “High Risk”**

In addition to data reduction, PCA allows assessment of cumulative risk by adding factor scores into a summary risk score. We created such a summative scale in the validation sample. Factor scores created during the factor analytic process were added together to create the risk score. High-risk categories were then created for each individual factor and for the cumulative risk scale. Because our factor analyses were performed with standardized variables, a score >2.0 on an individual factor was categorized as “high risk.” For cumulative risk, those scoring in the top fifth percentile (score >3.31) were assigned to the high-risk category. Logistic regression analysis with adjustment for age, gender, and race/ethnicity assessed independent correlates of high cumulative risk. Odds ratios and 95% CIs are reported. Because so few participants of Hispanic or other races/ethnicities were present, the logistic regression analysis was restricted to the 1232 non-Hispanic black and white adolescents in the validation sample.

**Results**

Correlations among measured variables, excluding BP, are presented in Table 2. The degree of correlation supported the use of factor analysis. In the subsample of those who had BP measured, systolic BP (SBP) was significantly associated with all CV risks but insulin; diastolic BP (DBP) was significantly associated with BMI, waist circumference, fibrinogen, LDL-C, and total cholesterol (P<0.05 for all).

Table 3 provides the results of the factor analyses that are presented graphically in the Figure. Three factors were extracted in both the exploratory and the validation samples with remarkably similar factor loadings. An adiposity factor, which accounted for the largest proportion of the total variance, was the initial factor extracted. A cholesterol factor was second factor, and a carbohydrate-metabolic factor was the third factor extracted. Together, these factors accounted for ~67% of the variance in the measured variables. Although the factor structure was consistent across samples, there were 2 notable differences between the exploratory and validation samples. Insulin had a factor loading >0.4 for only the carbohydrate-metabolic factor in the exploratory sample but for both the adiposity and the carbohydrate-metabolic factors in the validation sample. Additionally, HDL-C had a factor loading >0.4 for the adiposity factor in the exploratory sample and the carbohydrate-metabolic factor in the validation sample. For both samples, Bartlett’s test of sphericity was highly significant (<0.0001 for both), indicating good model acceptability.

In the subset with BP measures, factor analysis yielded a 4-factor solution. The first 3 factors were nearly identical to those extracted when BP was not included. The fourth factor was composed of the BP measures. Glucose loaded negatively onto the adiposity factor (~0.409). Glucose also loaded
negatively onto the adiposity factor in the validation sample, but the factor loading was much lower (−0.110).

**Description of Factor Scores and Cumulative Risk Scale Scores**

Scores on the adiposity factor ranged from −2.08 to 4.95. The median was −0.21. For the cholesterol factor, scores ranged from −2.74 to 7.63; the median was −0.07. For the carbohydrate-metabolic factor, scores ranged from −3.28 to 9.29, with a median of −0.13. Cumulative risk scale scores ranged from −3.54 to 12.58, with a median on −0.32. The majority (n=1148, 88.3%) were not at high risk for any factor. However, 10.9% (n=142) were at high risk for 1 of the 3 factors, and 10 subjects (0.8%) were at high risk for 2 factors. No one was at high risk for all 3 factors. The adiposity factor had the highest number of high-risk subjects (n=66, 5.1%). Female subjects were more likely to be at high risk for the adiposity factor (P=0.002), as were non-Hispanic

![Table](https://example.com/table2.png)
black compared with non-Hispanic white subjects \((P=0.001)\). The cholesterol and carbohydrate-metabolic factors had nearly equal numbers of high-risk subjects \((n=46, 3.5\% \text{ for the cholesterol factor and } n=50, 3.8\% \text{ for the carbohydrate-metabolic factor})\). Female subjects were less likely to be at high risk on the carbohydrate-metabolic factor \((P=0.009)\), as were prepubertal subjects \((P=0.003)\). Among those at high cumulative risk, a sizable proportion \((n=12, 18.5\% \text{ were not high risk for any of the 3 factors. Most scoring in the top fifth percentile of cumulative risk had only 1 factor with a score } >2.0 \text{ (n=43, 66.2\%). Among those at high cumulative risk, scores on the adiposity factor ranged from } -0.88 \text{ to } 4.95. \text{ Scores on the cholesterol factor ranged from } -1.84 \text{ to } 7.62, \text{ and scores on the carbohydrate-metabolic factor ranged from } -0.071 \text{ to } 9.29.\)

**Association of Weight Status and Hyperinsulinemia to High Risk**

Weight status was related to being in the high-risk category for all 3 extracted factors, whereas hyperinsulinemia was associated with being at high risk for 2 of the 3 factors (Table 4). In logistic regression analyses, both obesity and hyperinsulinemia remained independent correlates of high cumulative risk. However, obesity was a much more powerful correlate of high cumulative risk \((\text{odds ratio } [OR]=19.2; 95\% \text{ CI, 7.6 to 48.5})\) than hyperinsulinemia \((\text{OR}=3.5; 95\% \text{ CI, 1.8 to 6.8})\). When the cut point for hyperinsulinemia was increased to the top fifth percentile, obesity remained the strongest correlate of high cumulative risk \((\text{OR}_{\text{obesity}}=24.5 [95\% \text{ CI, 9.9 to 60.2}]; \text{OR}_{\text{hyperinsulinemia}}=12.4 [95\% \text{ CI, 6.0 to 26.0}])\).

**Discussion**

This study demonstrates that obesity is the predominant correlate of CV risk among adolescents regardless of whether that risk is defined in terms of individual physiological variables, factor scores assessing domains of metabolic clustering, or a cumulative risk scale. BMI and obesity were associated with every risk we measured. Insulin and hyperinsulinemia were associated with most of these risks, but not as consistently and to a lesser degree. These findings are consistent with data from the Bogalusa Heart Study, which demonstrated that BMI was a stronger predictor than insulin of developing a phenotype consistent with the metabolic syndrome.\(^{13,15}\) We also found that the factor patterning seen among adults is present in healthy adolescents and that the factors could be interpreted in a manner similar to those resulting from PCA in adults.\(^{36}\) Thus, the interrelationships between these physiological variables appear to be established early in the life course. Whether high factor scores on any particular factor will predict development of CV disease in adulthood remains to be determined through longitudinal analyses.

This study not only confirms but also extends prior work by developing a cumulative risk scale from the factor scores. To our knowledge, such a cumulative scale has not been used in prior studies, despite the fact that PCA is an excellent tool for creating a summary risk measure. We found that high cumulative risk did not reflect high risk across all factor domains. Most subjects with high cumulative risk were high risk on only 1 factor, and nearly 1 in 5 were not high risk for any factor. These findings highlight the importance of a global approach to assessing risk and the need for studies that elucidate how these various risk domains interact over time to create clinical disease.

These findings also add depth to the small amount of literature on factor analyses of CV risk in pediatric populations. PCA from the Bogalusa Heart Study revealed only 2 factors: a BP factor and a metabolic syndrome factor.\(^{7}\) In contrast, using almost identical physiological variables, methods, and cut points for interpretations, Lambert et al\(^{30}\) demonstrated 3 uncorrelated factors \((\text{glucose, lipids, and BP})\) in children and adolescents from the Quebec Child and Adolescent Health and Social Survey. Our population is more similar to that of the Bogalusa Heart Study than the Quebec study. However, our methods and data are more consistent with the work of Lambert et al. We and Lambert et al\(^{30}\) used standardized variables in our analyses, whereas Chen et al\(^{7}\) used raw variables. As in the Quebec study, we found
multiple factors in addition to a BP factor and found that glucose loaded negatively onto the first factor.

Although our findings are consistent with those of Lambert et al.,30 there are 3 significant differences between our study and both prior community-based pediatric studies. First, both prior studies used lower cut points for interpretation of the factor loadings than our cut point of 0.4. This difference influenced their conclusions. Chen et al7 suggested that insulin was a key linking entity because it loaded on multiple factors. However, the insulin loadings onto the BP factor were <0.4 in their adolescent and young adult samples. Lambert et al30 suggested a unifying role for both insulin and BMI because these measured variables loaded onto multiple factors, yet the second factor loading for BMI was consistently <0.4. We chose a cut point of 0.4, which has been used in adult factor analysis,21 because this indicates that the measured variable shares at least 15% of the variance with the factor. Second, our aim was to use PCA as a means of summarizing information across a broad array of measured variables to assess cumulative risk, not to explain the structure underlying the metabolic syndrome. As noted above, there are methodological difficulties inherent in using PCA to assess etiologic processes underlying CV risk clustering. Third, fibrinogen, LDL-C, and total cholesterol were included in our factor analyses. Fibrinogen loaded onto the adiposity factor, which is consistent with the pattern seen in adults and suggests that this variable is a marker of inflammation more than hemostasis.21 The separation of LDL-C and total cholesterol is also consistent with adult data.23 When we restricted our PCA to those variables related to the metabolic syndrome, as the other pediatric studies have done, we also extracted 3 factors (carbohydrate-metabolic [insulin, glucose, HDL-C, triglycerides], adiposity [BMI, waist, glucose], and BP [SBP, DBP]). These analyses suggest that the factor-loading patterns we demonstrate are robust.

The inverse factor loading for glucose on the first factor in both our study and the large, representative Canadian study30 suggests that the linear relationship responsible for this pattern is real. Its etiology, however, remains unclear. Care must be taken when any causal relationship between variables and factors in PCA is suggested. Although prior studies have used the factor-loading patterns from principal components analyses to explain the structure underlying the metabolic syndrome and CV risk clustering, this statistical technique does not provide a test of a hypothetical causal model or underlying structure. The factors derived by PCA are a mathematical transformation of the measured variables. There is no broader meaning that can be ascribed to them than those inherent in the original measured variables.55 Confirmatory factor analysis allows investigators to ascribe the factor pattern to an underlying hypothetical structure or causal model. Only 1 study among adults has used confirmatory factor analyses to test the factor structure underlying the metabolic syndrome.29 Results supported the 4-factor model. Whether this structure is consistent throughout development is unclear, as is the relationship of a broader array of CV risks to this factor pattern.

Although our findings are internally robust and consistent with prior work, there are some limitations that should be noted. This study has poor representation from racial/ethnic groups other than non-Hispanic blacks and whites. Whether this factor structure would generalize other population subgroups is unclear, although work among adults suggests that the factor pattern among Hispanics is similar to that seen among non-Hispanic black and white populations.28 In addition, we know of no reason to postulate that the linear relationships between these measured CV risks, on which PCA is based, would differ on the basis of demographic characteristics such as race/ethnicity, gender, or age. Prior factor analysis studies that have performed subgroup analyses have not found significant differences in factor patterns on the basis of these demographic characteristics.7,17,25,28,30 Last,
although we measured a broad array of CV risks, we were missing BP on the majority of our sample. Virtually every study indicates that BP loads onto a separate factor, and our subgroup analysis supported this separation. Thus, we do not believe that the lack of BP measurement for the entire sample influenced our factor analyses. Whether the addition of BP to the cumulative risk score would change our findings is not clear. Balancing these limitations are the strengths of this study: a large sample with nearly equal representation of non-Hispanic black and white boys and girls, measurement of a broad range of CV risks, careful study design, including use of a split sample technique to reduce the likelihood that the factor patterns we describe are due to some stochastic process particular to this data set, and innovative use of PCA to develop a cumulative risk scale.

In conclusion, the present study demonstrates that obesity is a powerful correlate of CV risk in healthy adolescents. Obesity was associated with individual CV risks, summary factors derived from these risks, and cumulative risk on the basis of the factor scores. The associations of obesity with these risks were stronger and more consistent than the associations between hyperinsulinemia and CV risk. In addition, these data indicate that the clustering of CV risks can be used to derive summary factors and a cumulative risk score in pediatric populations. These summary factors and risk score may enable longitudinal studies to establish trajectories of risk, thereby enhancing our understanding of the natural history of CV disease.

Acknowledgments

This work was supported by National Institutes of Health grants HD41527, DK59183, and M01 RR 08084 and by the William T. Grant Foundation. The authors thank Bin Huang, PhD, for statistical advice and support; the students, parents, teachers, and administration of the Princeton City Schools; and the staff of the Princeton School District Study.

References


Factor Analysis of Clustered Cardiovascular Risks in Adolescence: Obesity Is the Predominant Correlate of Risk Among Youth
Elizabeth Goodman, Lawrence M. Dolan, John A. Morrison and Stephen R. Daniels

doi: 10.1161/01.CIR.0000161957.34198.2B
Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2005 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/111/15/1970

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation is online at:
http://circ.ahajournals.org//subscriptions/