The combination of insulin resistance, dyslipidemia, hypertension, and obesity has been described as a “metabolic syndrome” that is a strong determinant of type 2 diabetes. Factor analysis was used to identify components of this syndrome in 1,918 Pima Indians. Prospective analyses were conducted to evaluate associations of identified factors with incidence of diabetes. Factor analysis identified 4 factors that accounted for 79% of the variance in the original 10 variables. Each of these factors reflected a proposed component of the metabolic syndrome: insulinemia, body size, blood pressure, and lipid metabolism. Among 890 originally nondiabetic participants with follow-up data, 144 developed diabetes in a median follow-up of 4.1 years. The insulinemia factor was strongly associated with diabetes incidence (incidence rate ratio [IRR] for a 1-SD difference in factor scores = 1.81, P < 0.01). The body size and lipids factors also significantly predicted diabetes (IRR 1.52 and 1.37, respectively, P < 0.01 for both), whereas the blood pressure factor did not (IRR 1.11, P = 0.20). Identification of four unique factors with different associations with incidence of diabetes suggests that the correlations among these variables reflect distinct metabolic processes, about which substantial information may be lost in the attempt to combine them into a single entity. 

**RESEARCH DESIGN AND METHODS**

Participants and measurements. The present data come from a longitudinal study of type 2 diabetes that has been conducted among residents of the Gila River Indian Community in central Arizona (27); most of these residents are Pima or Tohono O’odham Indians. Community residents who are ≥5 years of age are invited biennially to have an examination, which includes a 75-g oral glucose tolerance test. Plasma glucose and serum insulin concentrations are measured in specimens collected when the participant is fasting and 2 h after an oral glucose load. Insulin concentrations are measured by immunoassay (Concept 4; ICN, Costa Mesa, CA). Diabetes is diagnosed by World Health Organization (WHO) criteria for epidemiological studies (i.e., 2-h plasma glucose concentration ≥11.1 mmol/l or if the diagnosis is made during routine medical care (28,29). Height, weight, and waist circumference were measured with the participant wearing light clothing and no shoes. Blood pressure is measured to the nearest 2 mmHg with the participant lying supine; first and fourth Korotkoff sounds are taken as systolic and diastolic pressures, respectively. Since 1993, fasting serum triglyceride and HDL cholesterol concentrations have also been measured.

The present factor analyses included the following measurements: fasting and 2-h plasma glucose concentrations (G0, G2), fasting and 2-h serum insulin concentrations (I0, I2), systolic and diastolic blood pressure, body weight, waist circumference, serum triglyceride, and HDL cholesterol concentrations. Because many of these variables change dramatically as children mature, analyses were restricted to individuals who were ≥20 years of age. For each individual, the first examination at which all relevant measurements were available was taken. Individuals who were currently taking insulin or who were pregnant were excluded. Thus, factor analyses included 1,448 nondiabetic individuals (856 women and 592 men, mean age ±SD) 34.6 ± 11.7 years) and 470 diabetic individuals (303 women and 167 men, mean age 43.5 ± 12.8 years). Diabetic and nondiabetic participants were initially analyzed separately, and since similar results were obtained, a combined analysis was conducted. To better represent insulin resistance, particularly when both diabetic and nondiabetic individuals are considered, product (I0/G0, I2/G2) and ratio (I0/G0, I2/G2) indexes were analyzed; these indexes are strongly correlated with measurements of insulin resistance derived from the hyperinsulinemic-euglycemic “clamp” (30,31). However, similar results were obtained.
tained when \( I_0 \), \( G_0 \), \( G_1 \), and \( G_2 \) were analyzed in separate studies of diabetic and nondiabetic individuals. In statistical analyses, the natural logarithm was taken of all variables to reduce skewness. Before factor analyses, all variables were adjusted for age, sex, height, and birth year by linear regression. Analyses were conducted using programs of the SAS Institute (Cary, NC). To examine whether medications that affect blood pressure, lipidemia, or glycemia were influencing results, factor analyses were conducted among the subset of individuals who were not currently taking medicines, and similar results to those presented here were obtained.

**Factor analyses.** The premise underlying factor analysis is that correlations observed among a set of variables can be explained by a small number of unique unmeasured variables or “factors.” Therefore, factor analysis involves two procedures: 1) factor extraction to estimate the number of factors, and 2) factor rotation to determine constituents of each factor in terms of the original variables.

Factor extraction was conducted by the method of principal components. These components are linear combinations of the original variables that are constructed so that each component has a correlation of zero with each of the other components. Each principal component is associated with an “eigenvalue,” which represents the variance in the original variables explained by that component (with each original variable standardized to have a variance of 1). The number of principal components that can be constructed is equal to the number of original variables. In factor analysis, the number of factors is customarily determined by retention of only those components that account for more of the total variance than any single original variable (i.e., those components with eigenvalues > 1).

Once the number of factors has been established, then factor rotation is conducted to determine the composition of factors that has the most parsimonious interpretation in terms of the original variables. In factor rotation, “factor loadings,” which represent correlations of each factor with the original variables, are changed so that these factor loadings are made as close to 0 or 1 as possible (with the constraint that the total amount of variance explained by the factors remains unchanged). A number of methods for factor rotation have been developed; these methods can be distinguished by whether they require the final set of factors to remain uncorrelated with one another (orthogonal methods) or by whether they allow factors to be correlated (oblique methods). In the present analyses, both an orthogonal (varimax rotation) and an oblique (promax rotation) method were used. Because these methods gave similar results and because previous studies of metabolic syndrome variables have used orthogonal methods, results for the varimax method are presented.

In interpretation of factor analysis, the pattern of factor loadings is examined to determine which original variables represent primary constituents of each factor. Conventionally, variables that have a factor loading > 0.4 (or less than −0.4) with a particular factor are considered to be its major constituents.

**Incidence of diabetes.** The examination for which factor analysis was conducted was taken as the baseline examination for analyses of diabetes incidence; (these examinations occurred between February 1993 and May 1998—the time in which all relevant variables were measured). Individuals who were not diabetic at baseline and who had at least one subsequent examination were included. They comprised 890 participants (549 women and 341 men, mean age 33.3 ± 10.3 years). Factor scores were calculated from analysis of all subjects; these scores are linear combinations of the original variables that represent the predicted value of a given factor for an individual.

Person-time was calculated from the baseline examination until diagnosis of diabetes or until the last examination before August 2001, whichever came first. Incidence rates were calculated in events per 1,000 person-years for tertile groups for each factor score. Age- and sex-standardized incidence rates (and SEs) were calculated by the direct method as previously described (28) with the 1980 U.S. population as the referent. The proportional hazards model (32) was used to calculate the incidence rate ratio (IRR) and 95% CI for a 1-SD difference in each of the factor scores with adjustment for age and sex. This allows for comparisons of the influence of factors on risk of diabetes by comparison of the magnitude of the IRR. Validity of the proportionality assumption was assessed by use of time-dependent covariates (33) and by comparison of the IRR before and after median time to development of diabetes (3.3 years).

Comparison of the ability of factor scores to predict incident diabetes was determined by analysis of receiver operating characteristic (ROC) curves, in which sensitivity is plotted as a function of specificity. In the present context, area under the ROC curve represents the probability that an individual randomly selected from among those who developed diabetes has a higher factor score than one selected from among those who did not develop the disease (34). Thus, it is a measure of the ability of each factor score to predict diabetes, with values closer to 1 representing stronger prediction. Statistical significance of the difference between ROC curve areas for pairs of factor scores was tested as described by Hanley and McNeil (34,35). As the purpose of factor analysis is to identify shared components that explain correlations among a set of variables, factor scores are not necessarily optimal predictors of diabetes. For comparative purposes, therefore, a ROC curve was also calculated for an optimally predictive multivariable score calculated as the sum of each of the original variables multiplied by its regression coefficient in a proportional hazards model that simultaneously included all 10 variables.

Incidence of diabetes was also examined according to two recent definitions of metabolic syndrome (36,37). Thus, individuals were defined as having metabolic syndrome by WHO criteria (36) if they had impaired glucose regulation or insulin resistance (fasting plasma glucose ≥ 6.1 mmol/l or 2-h plasma glucose ≥ 7.8 mmol/l or \( I_0 \) \( G_0 \), in the highest quartile for normoglycemic individuals) and at least two of the following: blood pressure ≥ 140/90 mmHg; HDL cholesterol ≤ 1.3 mmol/l (women); waist circumference ≥ 88 cm (women); triglycerides ≥ 1.7 mmol/l; and HDL cholesterol < 1.0 mmol/l (men) or < 1.3 mmol/l (women). Similarly, individuals had metabolic syndrome by National Cholesterol Education Program (NCEP) criteria (37) if they had at least three of the following: blood pressure ≥ 130/85 mmHg; fasting plasma glucose ≥ 6.1 mmol/l; waist circumference > 102 cm (men) or > 88 cm (women); triglycerides ≥ 1.7 mmol/l; and HDL cholesterol < 1.0 mmol/l (men) or < 1.3 mmol/l (women). The present analyses use 1999 WHO criteria, but results are virtually identical if 1998 WHO criteria (38), which have slightly different values for blood pressure and albuminuria, are used.

**RESULTS**

Correlations among variables are shown in Table 1. For most variables correlations were of modest to moderate magnitude (0.2–0.7). A notable exception was that between body weight and waist circumference, which was consistently > 0.9. In addition, correlations of blood pressure variables with the others tended to be < 0.2. The general pattern of correlations was similar in diabetic and nondiabetic individuals.

**Factor analysis in nondiabetic individuals.** Among nondiabetic individuals, principal components analysis identified four factors with an eigenvalue > 1; the largest eigenvalue among the remaining components was 0.64. These four factors, which accounted for 81.2% of the variance in the original 10 variables, were retained for factor rotation. Factor loadings of the original variables with each factor after varimax rotation are shown in Table 2. The first factor was strongly correlated with \( I_0 \), \( G_0 \), \( I_2 \), \( G_2 \), \( I_1/G_0 \), and \( I_2/G_2 \) as indicated by a factor loading > 0.4 with each of these variables. Likewise, the second factor was strongly correlated with body weight and waist circumference, the third factor was strongly correlated with systolic and diastolic blood pressure, and the fourth factor was strongly correlated with serum triglyceride concentration and inversely with HDL cholesterol concentration. Thus, these four factors can be interpreted as representing insulinemia, body size, blood pressure, and lipid metabolism.

**Factor analysis in diabetic individuals.** Among diabetic participants, there were also four factors that had eigenvalues > 1. The largest eigenvalue among the remaining components was 0.69, and the four factors accounted for 80.7% of the variance among the original variables. Factor loadings after varimax rotation are shown in Table 3. The first factor was correlated with \( I_0 \) \( G_0 \), \( I_2 \), \( G_2 \), \( I_0/G_0 \), and \( I_2/G_2 \); the second factor was correlated with body weight and waist circumference; the third factor was correlated with systolic and diastolic blood pressure, and the fourth factor was correlated with serum triglycerides and \( I_0 \) \( G_0 \) and inversely correlated with HDL cholesterol. Thus, similar to results in nondiabetic individuals, results in
diabetic participants can be interpreted as representing insulinemia, body size, blood pressure, and lipid metabolism.

### TABLE 2
Factor loadings for original variables with rotated factors in nondiabetic participants

<table>
<thead>
<tr>
<th></th>
<th>Insulinemia</th>
<th>Body size</th>
<th>Blood pressure</th>
<th>Lipidemia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body weight</td>
<td>0.25</td>
<td>0.93</td>
<td>0.11</td>
<td>0.10</td>
</tr>
<tr>
<td>Waist circumference</td>
<td>0.23</td>
<td>0.93</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>HDL cholesterol</td>
<td>-0.15</td>
<td>-0.18</td>
<td>0.18</td>
<td>-0.78</td>
</tr>
<tr>
<td>Triglycerides</td>
<td>0.10</td>
<td>-0.01</td>
<td>0.22</td>
<td>0.82</td>
</tr>
<tr>
<td>Systolic blood pressure</td>
<td>0.06</td>
<td>0.06</td>
<td>0.86</td>
<td>0.00</td>
</tr>
<tr>
<td>Diastolic blood pressure</td>
<td>0.05</td>
<td>0.10</td>
<td>0.85</td>
<td>0.04</td>
</tr>
<tr>
<td>Fasting insulin × glucose</td>
<td>0.82</td>
<td>0.37</td>
<td>0.06</td>
<td>0.11</td>
</tr>
<tr>
<td>2-h Insulin × glucose</td>
<td>0.89</td>
<td>0.08</td>
<td>0.09</td>
<td>0.13</td>
</tr>
<tr>
<td>Fasting insulin/glucose</td>
<td>0.83</td>
<td>0.33</td>
<td>0.03</td>
<td>0.08</td>
</tr>
<tr>
<td>2-h Insulin/glucose</td>
<td>0.92</td>
<td>0.05</td>
<td>0.02</td>
<td>0.09</td>
</tr>
<tr>
<td>Percentage of variance</td>
<td>31.5</td>
<td>20.4</td>
<td>15.8</td>
<td>13.5</td>
</tr>
</tbody>
</table>

Factor loadings with an absolute value ≥0.4, which indicates that the variable can be considered a major constituent of the factor, are shown in bold. Each factor is named according to the variables which are its major constituents. The percentage of the variance represents the proportion of variance in the original 10 variables accounted for by each factor.

### TABLE 3
Factor loadings for original variables with rotated factors in diabetic participants

<table>
<thead>
<tr>
<th></th>
<th>Insulinemia</th>
<th>Body size</th>
<th>Blood pressure</th>
<th>Lipidemia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body weight</td>
<td>0.19</td>
<td>0.95</td>
<td>0.04</td>
<td>-0.05</td>
</tr>
<tr>
<td>Waist circumference</td>
<td>0.14</td>
<td>0.95</td>
<td>0.02</td>
<td>-0.05</td>
</tr>
<tr>
<td>HDL cholesterol</td>
<td>-0.04</td>
<td>-0.05</td>
<td>0.89</td>
<td>0.00</td>
</tr>
<tr>
<td>Triglycerides</td>
<td>-0.13</td>
<td>-0.19</td>
<td>0.10</td>
<td>0.78</td>
</tr>
<tr>
<td>Systolic blood pressure</td>
<td>0.01</td>
<td>-0.05</td>
<td>0.89</td>
<td>0.00</td>
</tr>
<tr>
<td>Diastolic blood pressure</td>
<td>0.00</td>
<td>0.12</td>
<td>0.88</td>
<td>0.03</td>
</tr>
<tr>
<td>Fasting insulin × glucose</td>
<td>0.52</td>
<td>0.33</td>
<td>0.11</td>
<td>0.45</td>
</tr>
<tr>
<td>2-h Insulin × glucose</td>
<td>0.93</td>
<td>0.07</td>
<td>-0.01</td>
<td>0.10</td>
</tr>
<tr>
<td>Fasting insulin/glucose</td>
<td>0.91</td>
<td>0.25</td>
<td>0.03</td>
<td>-0.02</td>
</tr>
<tr>
<td>2-h Insulin/glucose</td>
<td>0.88</td>
<td>0.05</td>
<td>-0.05</td>
<td>-0.23</td>
</tr>
<tr>
<td>Percentage of variance</td>
<td>28.2</td>
<td>20.5</td>
<td>16.1</td>
<td>15.9</td>
</tr>
</tbody>
</table>

Factor loadings with an absolute value ≥0.4, which indicates that the variable can be considered a major constituent of the factor, are shown in bold. Each factor is named according to the variables that are its major constituents. The percentage of the variance represents the proportion of variance in the original 10 variables accounted for by each factor.
identified with eigenvalues >1. The largest eigenvalue among the remaining components was 0.62, and the four factors accounted for 79.1% of the variance in the original variables. Factor loadings after varimax rotation are shown in Table 4. Results were comparable with those obtained when factor scores were analyzed as continuous variables in proportional hazards models. The age- and sex-adjusted IRR for a 1-SD increase in the insulinemia factor score was 1.81 (95% CI 1.48–2.20, \( P < 0.01 \)), whereas the corresponding IRR for the body size factor score was 1.52 (1.30–1.77, \( P < 0.01 \)). The lipids factor score had a more modest, but still statistically significant, association with risk of developing diabetes (IRR 1.37, 95% CI 1.16–1.62, \( P < 0.01 \)), whereas the blood pressure factor score did not significantly predict diabetes (1.11, 0.95–1.31, \( P = 0.20 \)). The risk associated with the lipids factor score was not proportional over the follow-up period, with an IRR of 1.75 (95% CI 1.39–2.21) in the initial period and of 1.10 (0.89–1.38) in the second period.

ROC curves for each of the four factors and for the multivariable score derived from all 10 variables are shown in Fig. 2. In analyses that compared the ability of factor scores to predict diabetes, area under the ROC curve for the insulinemia factor was not significantly greater than that for the body size factor (\( P = 0.77 \) for comparison of areas) and was not significantly greater than that for the lipids factor (\( P = 0.14 \)). However, the insulinemia factor was a stronger predictor of incident diabetes than the blood pressure factor (\( P < 0.01 \)). Similarly, the body size and lipids factors did not differ

![Image](https://via.placeholder.com/150)

**Fig. 1.** Age- and sex-standardized incidence rates of diabetes (with 95% CIs) according to tertile groups defined by factor scores for each of the four factors identified in factor analysis.

---

**TABLE 4**

Factor loadings for original variables with rotated factors in nondiabetic and diabetic participants

<table>
<thead>
<tr>
<th>Factor</th>
<th>Insulinemia</th>
<th>Body size</th>
<th>Blood pressure</th>
<th>Lipidemia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body weight</td>
<td>0.22</td>
<td><strong>0.94</strong></td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>Waist circumference</td>
<td>0.19</td>
<td><strong>0.95</strong></td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>HDL cholesterol</td>
<td>−0.11</td>
<td>−0.16</td>
<td>0.14</td>
<td><strong>−0.79</strong></td>
</tr>
<tr>
<td>Triglycerides</td>
<td>0.01</td>
<td>−0.03</td>
<td>0.19</td>
<td><strong>0.83</strong></td>
</tr>
<tr>
<td>Systolic blood pressure</td>
<td>0.05</td>
<td>0.03</td>
<td><strong>0.87</strong></td>
<td>0.01</td>
</tr>
<tr>
<td>Diastolic blood pressure</td>
<td>0.03</td>
<td>0.12</td>
<td><strong>0.86</strong></td>
<td>0.05</td>
</tr>
<tr>
<td>Fasting insulin × glucose</td>
<td><strong>0.63</strong></td>
<td><strong>0.40</strong></td>
<td>0.13</td>
<td>0.36</td>
</tr>
<tr>
<td>2-h Insulin × glucose</td>
<td><strong>0.83</strong></td>
<td>0.16</td>
<td>0.10</td>
<td>0.23</td>
</tr>
<tr>
<td>Fasting insulin/glucose</td>
<td><strong>0.85</strong></td>
<td>0.29</td>
<td>0.03</td>
<td>0.05</td>
</tr>
<tr>
<td>2-h Insulin/glucose</td>
<td><strong>0.87</strong></td>
<td>0.00</td>
<td>−0.06</td>
<td>−0.16</td>
</tr>
<tr>
<td>Percentage of variance</td>
<td>26.6</td>
<td>21.0</td>
<td>16.0</td>
<td>15.5</td>
</tr>
</tbody>
</table>

Factor loadings with an absolute value ≥0.4, which indicates that the variable can be considered a major constituent of the factor, are shown in bold. Each factor is named according to the variables which are its major constituents. The percentage of the variance represents the proportion of variance in the original 10 variables accounted for by each factor.
significantly in prediction of diabetes ($P = 0.26$), but both ROC curve areas were significantly greater than that for the blood pressure factor ($P < 0.01$ and $P = 0.03$, respectively). Area under the ROC curve for the multivariable score was significantly greater than that for any one of the factors ($P < 0.01$ for each).

At baseline 31% of participants had metabolic syndrome by WHO criteria and 31% by NCEP criteria. For prediction of diabetes, WHO criteria were more specific and more sensitive than NCEP criteria (Fig. 2). Age- and sex-standardized diabetes incidence rates according to metabolic syndrome by WHO or NCEP criteria are shown in Fig. 3. Those who met WHO criteria had a much higher incidence of diabetes than those who did not (IRR 3.58, 95% CI 2.56–5.00), whereas increased incidence in those who met NCEP criteria compared with those who did not was more modest (2.09, 1.49–2.92). WHO criteria require hyperglycemia or insulin resistance in addition to other metabolic abnormalities, whereas NCEP criteria treat the components more equally. To examine whether this requirement accounts for the stronger prediction of WHO criteria, incidence rates were further stratified according to presence of hyperglycemia or insulin resistance (Fig. 3, right panels). Individuals with hyperglycemia/insulin resistance had a much higher incidence of diabetes than those without it, regardless of whether WHO or NCEP criteria for metabolic syndrome were met. This suggests that the requirement for hyperglycemia or insulin resistance is the major reason for different predictive properties between WHO and NCEP criteria.

**DISCUSSION**

The observation that many risk factors for diabetes are strongly intercorrelated has led to the hypothesis that they have certain etiological factors in common (3–6). In particular, insulin resistance, obesity, central adiposity, hypertension, hypertriglyceridemia, and hypoalphalipidemia by WHO criteria and 31% by NCEP criteria. For prediction of diabetes, WHO criteria were more specific and more sensitive than NCEP criteria (Fig. 2). Age- and sex-standardized diabetes incidence rates according to metabolic syndrome by WHO or NCEP criteria are shown in Fig. 3. Those who met WHO criteria had a much higher incidence of diabetes than those who did not (IRR 3.58, 95% CI 2.56–5.00), whereas increased incidence in those who met NCEP criteria compared with those who did not was more modest (2.09, 1.49–2.92). WHO criteria require hyperglycemia or insulin resistance in addition to other metabolic abnormalities, whereas NCEP criteria treat the components more equally. To examine whether this requirement accounts for the stronger prediction of WHO criteria, incidence rates were further stratified according to presence of hyperglycemia or insulin resistance (Fig. 3, right panels). Individuals with hyperglycemia/insulin resistance had a much higher incidence of diabetes than those without it, regardless of whether WHO or NCEP criteria for metabolic syndrome were met. This suggests that the requirement for hyperglycemia or insulin resistance is the major reason for different predictive properties between WHO and NCEP criteria.
poproteinemia have all been described as risk factors for diabetes (1,3). Because many of these risk factors cluster together, it has been hypothesized that they may reflect a limited number of etiological metabolic abnormalities or perhaps even a single abnormality. The combination of obesity, insulin resistance, hypertension, and dyslipidemia has, thus, been categorized as a single syndrome, called variously “syndrome X,” “insulin resistance syndrome,” “metabolic syndrome,” or “dysmetabolic syndrome.”

However, the present factor analyses did not identify a single factor underlying the correlation structure in the variables included. Rather, it identified four factors, each of which is interpretable as representative of a proposed component of the metabolic syndrome: insulinemia, body size, blood pressure, and lipid metabolism. This suggests that the 10 variables analyzed are best described as reflecting these 4 independent physiological domains. Factor analysis is designed to reduce a large number of correlated variables to fewer factors, which reflect unique underlying phenomena and which extract much of the variation in the original variables. It is thus well suited to describing and quantifying components of metabolic syndrome. The technique of factor analysis is subject to a number of limitations, however. Most notably, results can be dependent on a number of somewhat arbitrary criteria, i.e., the threshold chosen for selection of number of factors retained, the method of rotation used, and the minimum factor loading chosen to designate a variable as a primary constituent of a factor. Robustness of a particular solution is best evaluated by whether it is consistent across different groups of individuals and different analytic procedures. The present analyses gave very similar results in diabetic and nondiabetic individuals, and these results were consistent using different procedures for factor rotation.

Similar factor analyses have been conducted for a number of other populations (as reviewed by Meigs [39]). Despite the subjective nature of factor analysis and differences as to which variables are included, results of these studies have shown some consistency. Many of them, like the present study, have identified a four-factor solution with factors representing obesity, insulin-glucose metabolism, lipid metabolism, and blood pressure, with some minor variation in composition of individual factors (7–12). Others have identified a three-factor solution that combines some of these entities; commonly, obesity is combined with lipid and/or insulin variables (13–22). A few studies have identified a two-factor solution that further combines variables (23–26). The majority of studies have identified blood pressure as a factor distinct from the others (7–12,14–18,26). Most of the studies that have not identified a separate blood pressure factor (13,20–25) have analyzed only a single blood pressure variable (20–24). Thus, the results of most factor analyses, including the present study, suggest that relationships among the variables typically proposed as constituting the metabolic syndrome are best explained as resulting from multiple physiological processes and that the attempt to reduce these to a single entity will result in a substantial loss of information about these metabolic processes.

Although factor scores are derived to be uncorrelated with one another, this does not necessarily mean that underlying genetic or environmental determinants of these scores, or of their constituent variables, are uncorrelated. With respect to potential genetic determinants, some family-based studies have suggested substantial genetic correlations between certain variables constituting the metabolic syndrome (40,41), but others have not (42). Among Pimas, genetic linkage analyses of factor scores derived from the present study did not reveal any genomic regions with strong evidence for linkage to multiple factors (43); similar findings were seen in Mexican-Americans (18). This suggests that the major genetic determinants of these factors are separate.

Factor analysis is not designed to identify combinations of variables that are the strongest predictors of diabetes; this information is optimally derived from multivariable prediction models. However, epidemiological analyses that use factor scores can yield useful insights into relationships between various metabolic processes and disease incidence. Although most of the individual variables are risk factors for type 2 diabetes in various populations (1,3), results of analyses of individual risk factors are potentially confounded because of the correlations among these variables. Risk estimates associated with individual variables can be difficult to interpret, when derived from multivariable models that include several correlated variables, because of problems with multicollinearity (although predictive properties of such models tend to be robust [44]). Because factor analysis reduces a large number of correlated variables to fewer uncorrelated factors, it can help to circumvent these problems. In some studies, this approach has been used to examine relationships of metabolic syndrome components with incidence of cardiovascular disease (10,12,23). There are limited data relating metabolic syndrome factors to incidence of type 2 diabetes; one study in Finland identified an insulinemia/obesity/blood pressure/triglycerides factor that was associated with incidence of diabetes (22). The present analyses show that, among the Pimas, hyperinsulinemia and obesity are strong risk factors for type 2 diabetes. The combination of hypertriglyceridemia and low HDL cholesterol is a more modest, but still significant, risk factor, whereas high blood pressure is only weakly, if at all, associated with diabetes incidence. Furthermore, in ROC analyses the insulinemia, body size, and lipids factors were all significantly more strongly predictive of diabetes than the blood pressure factor. These findings suggest that the different physiological processes associated with various components of the metabolic syndrome contain unique information about diabetes risk.

The variable diabetes risk associated with each of these factors is reflected in differences in the extent to which different proposed criteria for metabolic syndrome predict diabetes incidence. WHO criteria were much more strongly associated with diabetes incidence than NCEP criteria. This difference was due to the greater weight given to hyperglycemia/insulin resistance in the WHO criteria than with the NCEP criteria, which essentially give the individual components equal weight. This further illustrates the importance of considering each of the metabolic syndrome components individually, at least with respect to diabetes risk.

The present findings have potential implications for
research and clinical practice. If the abnormalities constituting the metabolic syndrome are the result of largely independent physiological processes, then attempts to study a global syndrome phenotype may be counterproductive. Epidemiological comparisons of metabolic syndrome prevalence across diverse populations, for example, may well reflect very different entities, with substantial heterogeneity in risk for diabetes, or other health outcomes. It may be possible to address these problems by defining the syndrome so as to give each component a different weight, as per the multivariable prediction score in the present analyses. However, such an approach may require a different syndrome definition depending on whether diabetes or cardiovascular risk is the focus of investigation and, as effects of different components may vary among different populations, depending on the population in which it is applied. Furthermore, clinical treatment and prevention strategies that focus solely on the metabolic syndrome may prove suboptimal compared with treatment of the individual components.

ACKNOWLEDGMENTS
This work was presented in part at the 59th Scientific Sessions of the American Diabetes Association, San Diego, CA, June 19–22, 1999.

The authors thank the members of the Gila River Indian Community for participation in these studies and staff of the Diabetes and Arthritis Epidemiology Section for assistance. Additional thanks are due to Dr. David Mott for supervision of insulin assays and to Dr. Jonathan Krakoff for advice.

REFERENCES