The Senior Authors’ Response: Factor Analysis as a Tool for Evaluating Eating Patterns

Martha L. Slattery and Kenneth M. Boucher

While it is inherently reasonable that patterns of dietary intake may represent a more accurate picture of diet than isolated foods, most research into the relation between diet and disease has studied single nutrients and foods. As we stated in our paper on eating patterns and colon cancer (1) and as implied by Martinez et al. (2) in their evaluation of factor analysis, assessment of eating patterns is not straightforward. Martinez et al. have outlined several issues that should be taken into consideration when using factor analysis.

SUBJECTIVITY

Like most statistical analysis done in epidemiologic research, there are arbitrary decisions that must be made by the researcher when conducting either exploratory or confirmatory factor analysis. For factor analyses, these include: deciding which variables to use when constructing factors; deciding on a meaningful level at which a variable is believed to contribute to a factor; and deciding on labels for the factors. In our exploratory factor analyses of eating patterns and colon cancer (1), we first defined food groups based on our existing knowledge of diet and colon cancer. We then allowed the statistical methods to group the foods into factors; we believe that the factors identified describe eating patterns that are interpretable. To define food groups, we constructed what could be considered factors, or linear combinations of data. The 35 food groups we delineated were the basis of our eating patterns. Other investigators could have combined foods into different groups, thereby developing dietary variables that were different from those we constructed. Likewise, the final models used for factor analysis could vary by researcher.

The meaningful level of loading of a variable to a factor is also arbitrary. Although we used what could be considered a correlation of 0.2, often higher levels are used. We used this level because we believed that it contributed to the overall interpretation of the factor. Correlations of 0.2 are often found in the literature and are considered to be important, although at a low level of association. For instance, the association between physical activity and total energy intake has been reported at levels of 0.11 and 0.22 (3, 4). However, as Martinez et al. noted (2), variables that contribute less to a factor could contribute to more than one factor. We do not necessarily find this bothersome, since there are a multitude of eating patterns, many of which could potentially contain the same foods. It is also important to remember that all individuals have a loading on every factor; it is the level at which they are associated with the factor that varies.

Labels given to a factor are also arbitrary. Because of the arbitrary nature of labeling, it is important to show the actual factor loadings in the published results. This allows readers to determine for themselves what each factor represents.

INCONSISTENCIES

Perhaps the most important issue raised by Martinez et al. (2) is the utility of factor analysis in epidemiology. Are the identified factors reproducible? More importantly, will inferences made from factor analyses enable us to understand the inconsistencies that currently exist in the literature?

While consistency has been identified by epidemiologists as contributing to our belief in causal associations, there are reasons for inconsistencies. We believe that the eating patterns we identified which contributed to the most variation in the diet (the “Western” and “prudent” patterns (1)) are more likely to be identified in other populations. However, it is also reasonable to expect that eating patterns will differ between populations. The other patterns we identified accounted for less of the spread of the data in this population and might not contribute importantly in other populations. Likewise, dietary patterns not identified in this population may be meaningful patterns in other populations. For instance, in a population of vegetarians, the Western dietary pattern may be...
less important or not important at all. However, the absence of an association between colon cancer and Western diet in that population should not diminish its importance to colon cancer in this population, or make it any less believable. What is important will be our ability to interpret patterns generated from factor analysis, taking into consideration our knowledge of the populations being studied, and sorting out the relevant data that explain the differences between the populations.

OTHER USES OF FACTOR ANALYSIS IN EPIDEMIOLOGIC RESEARCH

Factor analysis is one method of reducing data. Another method that might be informative with regard to diet-disease relations is hierarchical modeling (5). The manner in which hierarchical regression would lend itself to identification of eating patterns is unclear; it could, however, be a useful statistical tool for delineating the relative importance of foods and nutrients in disease. Factor analysis may have utility in describing lifestyle patterns as well as dietary patterns. The underlying assumptions of this method are that foods eaten together can be characterized as part of a meaningful dietary pattern and that aspects of a person's life which occur together typify a lifestyle that is more epidemiologically meaningful than its individual components. However, we firmly believe that data reduction by any method can occur only after a thorough understanding of the individual data elements has been obtained.

SUMMARY

On the basis of our results (1), we believe that factor analysis has applications in epidemiology that will improve our understanding of many of the inconsistencies that exist in the literature. Wade Hampton Frost's characterization of epidemiology in 1936 (6) applies to many of our current attempts to understand disease, often through the use of relatively new methods. He described epidemiology as "something more than the total of its established facts":

It includes their orderly arrangement into chains of inference which extend more or less beyond the bounds of direct observation. Such of those chains as are well and truly laid guide investigation to the facts of the future; those that are ill made fetter progress. But it is not easy, when divergent theories are presented, to distinguish immediately between those which are sound and those which are merely plausible (6, p. 120).

The soundness of using factor analysis to identify eating and lifestyle patterns that give us insight into disease etiology will be better understood when more epidemiologists have begun to use this method.

REFERENCES