

II-7.

Jerome Cornfield: Joint Dependence of Risk of Coronary Heart Disease on Serum Cholesterol and Systolic Blood Pressure: a Discriminant Function Analysis. Federation Proceedings 1962; 2:58-61.

The following paper by Jerome Cornfield helped introduce to epidemiology the concept of using multivariate models to overcome the limits of "thinness of data" encountered in ordinary stratified analysis. The model employed (given in equation 5 of the article) is now commonly known as the multiple logistic model, and has become one of the primary tools of epidemiologic analysis.

The method used by Cornfield to derive the model and estimate its parameters was introduced in an earlier paper by Cornfield, Gordon, and Smith [1961], and is based on an assumption of multivariate normality of the risk factors within disease groups. This is an unrealistic assumption in general, and this method was soon rendered obsolete by the development of estimation methods based on less restrictive assumptions (e.g., Walker and Duncan [1967]). Nevertheless, it is instructive to work through Cornfield's derivation as an illustration of the connection between disease-specific risk factor distributions and multivariate risk functions.

More importantly, Cornfield's discussion of the *interpretation* of the fitted model's parameters is still valid, and his careful exercise in translating the parameter estimates into relative risk estimates remains a paradigm for proper use of modeling results. My only criticism here is that he did not mention the possibility that the very large relative risk estimated for the joint effect of cholesterol and blood pressure may have been an artifact of employing the logistic model, as the model can force relative risk estimates to be spuriously large when comparing risk at extremes of risk factor values or combinations. This problem can be circumvented by comparing risks within more typical values of the risk factors, and by checking results against those obtained from stratified analysis. Apparently, the problem and its solution were not well recognized before the 1970's, but are discussed by Gordon in the next article.

References:

- Cornfield J, Gordon T, Smith WW. Quantal response curves for experimentally uncontrolled variables. Bull Int Statist Inst 1961; 38:97-115.
- Walker SH, Duncan DB. Estimation of the probability of an event as a function of several independent variables. Biometrics 1967; 54:167-179.