UNIVERSITY OF MARYLAND
SCHOOL OF MEDICINE
INSTITUTE OF INTERNATIONAL MEDICINE
BALTIMORE, MARYLAND 21201

DIVISION OF EPIDEMIOLOGY AND BIOSTATISTICS

March 25, 1971

Dr. Henry Blackburn
Universite de Geneve
Institut de medicine sociale
et preventive
1205 Geneve
le Rue de Candolle 12
Switzerland

Dear Henry:

Over the past few weeks you have asked some questions relating to multiple regression which I have not answered yet.

First, let me try again to explain how the "smoothed" relationship between mortality and cholesterol and triglyceride was obtained, as shown on some of the London slides. The situation is completely analogous to the one variable case. Suppose we are interested in the relationship between mortality and baseline cholesterol, for instance. One way to study this relationship is to group the patients by their baseline cholesterol values and then find the percent mortality within each group. Another way is to fit a straight line to the data, using the technique of simple linear regression. This latter method always gives a much smoother, though not necessarily more accurate, relationship than the former. By means of the fitted straight line, one can estimate the probability of death for any given cholesterol value. If we now study two variables, say baseline cholesterol and triglyceride, for their relationship to mortality, we can obtain in a similar way a smoothed relationship by fitting a two-dimensional "line" (that is, a plane) to the data using the technique of multiple linear regression. In this way, one can estimate the probability of death for any given combination of baseline cholesterol and triglyceride values.

Secondly, you asked how the cholesterol-triglyceride interaction could be close to zero when these two variables are generally highly correlated. Interaction and correlation are quite distinct statistical concepts. Cholesterol and triglyceride are indeed significantly correlated variables in the CDP population. But when we speak of interaction, we have reference to the joint effect of cholesterol and triglyceride on a third variable, in this case mortality. Suppose we fix triglyceride

at different values, each time computing the straight line relationship between cholesterol and mortality. If the slopes of these lines turn out to be the same for all values of triglyceride, we would say that there is no interaction between cholesterol and triglyceride. This is clearly seen in the example you sent me, a copy of which is enclosed. But if, on the other hand, cholesterol has no effect on mortality (zero slope) for low levels of triglyceride, but a sizable effect on mortality for high levels of triglyceride, we would say that there is a cholesterol-triglyceride interaction with respect to mortality.

Finally, a few words of explanation concerning the "T-values" shown on the London slides. They really should have been labeled "t-values" (after Student's "t") for they are nothing more than the regression coefficients (slopes) divided by the standard errors of the regression coefficients. The conversion of t-values to P-values is somewhat problemmatical. We are hesitant to associate a t-value of two with a P-value of .05 because of the large number of variables which we are looking at simultaneously and because of the way in which we selected these variables in the first place. There is no exact answer to the question of what t-value corresponds to a P-value of .05. Our own intuitive feelings are that perhaps things should not be called "statistically significant" until the t-value goes above three. It is our hope that through simulation studies we can shed some light on this problem during the next months.

Please let us know if you desire further clarification of these points. Perhaps we can get together when you are here next month and go over some of these things.

Sincerely yours,

Paul

Paul L. Canner, Ph.D.
Associate Professor

PLC:mmd
Enclosure

cc: Dr. Christian Klimt
Dr. William Krol
Dr. Thomas Landau
Dr. Curtis Meinert
Dr. Jeremiah Stamler
Dr. Suketami Tominaga