An Application of Causal Statistics to Pesticide Data

In 1974 the Pacific Biomedical Research Center in Honolulu, as a part of a national EPA research program, had been collecting data on pesticide exposures and health consequences. For 11 years they had been trying to analyze and make sense of their data, with little success. They gave me a grant to disentangle the labyrinth of associations and extract the causal connections from their data. I developed a 24 variable model -- including pesticide exposures, blood pressures, DDT levels, DDE levels, etc. -- and was successful in inferring the causal relationships.

The head of the Pacific Biomedical Research Center, Dr. Howard Klemmer, was ecstatic with my technique and its findings. Shortly thereafter he sent me to Washington to talk to the statistician for the national pesticide research project and to suggest that all the regional centers use causal statistics to analyze their data. The statistician turned out to be the quintessential government bureaucrat. He received the data from the regional centers; through them into a classical statistics computer program; printed out masses of correlations; and past the paper along to higher-ups. He seemed to have no knowledge of nor concern for what the results were used for after they left his office and no understanding of the policy, decision making, or other advantages that might flow from causal conclusions. Further, he had no interest in finding out.

After a long discussion -- which I would have described as otherworldly and now, after a lifetime of dealing with such people, I would describe as typical -- he eventually said, in a way designed to end the conversation, "We like the way we have been doing the statistics." He knew how to do classical statistical analysis; it was easy; he didn't have to think much; and that was the way he liked it. I left in disbelief. I have since learned that it must be me who is in the other world, because this world is full of that kind of people.

As I think about it now, I can see that I was young and a little stupid about organizational dynamics. If this happened to me today, I would make an appointment with the people who the statistical printouts went to and if necessary their bosses. These people would likely be the decision makers

and they would more easily understand what type of information they needed to ground their decisions,... maybe.

But, on third thought, I would realize that most government workers in an oversight agency are moles. These moles are placed there by the industry being overseen, in this case the pesticide industry, through the industry's contributions to and consequential influence on elected politicians. The main objective of these moles is to protect the pesticide industry, not to protect people or the environment from pesticide damage. These moles are "controlled" by the pesticide industry because they previously worked for the pesticide industry and/or will, after departing government employment, somehow benefit from the pesticide industry. With this understanding, today I would still not have great hopes for a sympathetic hearing by the higher-ups, because Causal Statistics would generate quality information and knowledge, commodities potentially dangerous to the pesticide industry. Cynical? Who, me?

Anyway, the following is taken from a 1974 quarterly report from the Pacific Biomedical Research Center to the Environmental Protection Agency, reporting on the pesticide data collection they had been carrying out for the EPA for 11 years.

The small piece presented below is a brief summary of the research paper I wrote laying out the results of my causal analysis of their pesticide data.

Wow! I just looked over the summary of my study and realized why they had no interest in causal statistics. The analysis found that DDT causes hypertension, a finding potentially quite damaging to the pesticide industry. If this result were known, the industry might have to fight lots of lawsuits from people with hypertension or relatives of people who died from complications stemming from hypertension. Not only that, but a tech meet with the power of causal statistics could be dangerous to all kinds of vested interests.

Coming from a university environment, I felt that better information was and unmitigated good, pretty naïve of me. I should have listened to my ancestors (Chickasaw and Cherokee Indians). I shouldn't criticize people until I have walked a mile in their moccasins, figuratively at least. In my life I have learned more than once that sometimes people make apparently stupid decisions because they are stupid, but other times they make smart decisions

that only appear stupid, because I am too stupid to put myself in their heads and understand their situations and motivations.

Hey, I am a causal inference thinker and researcher. I make no claim to being a good psychologist or being even average at inter human skills and understandings. Yet there is no doubt that my limited inter human abilities and understandings have damage my ability to obtain research funding and to disseminate the research that I have done.

On the other hand that's a powerful argument as to why funding agencies and philanthropists should endeavor to search out creative researchers and not simply fund those with the information and skills to play the funding game.

SECTION IV: Data Management and Statistical Analysis

Presented herein are summary write-ups by independent researchers on work accomplished during the past quarter in two areas of statistical analysis. This project has gained much from the work of these two researchers. The causal statistics to follow may be of particular significance in the future analysis of a large bulk of data that requires much more extensive analysis than has so far been possible. Unfortunately, what is needed for the continuation of such analysis is more time and money than are currently available.

Causal Statistics: Presented

Presented is an outline of a research paper by Dr. Charles S. Portwood.

I. INTRODUCTION

Recent research has shown correlations between DDT and hypertension and between DDE and hypertension, but the researchers have been unable to determine whether or not these connections are causal or spurious. The first correlation could be explained by the fact that obesity is intercorrelated with both hypertension and DDT level. And, to complicate the situation, DDE and cholesterol are intercorrelated with each other and with the above three variables.

II. PURPOSE

It was the purpose of this research to determine--by way of a preliminary and rough analysis--whether or not causal connections exist between DDT and hypertension and between DDE and hypertension. In addition, it was the intent of the research to determine the strength of any causal connection found.

A person is considered to be hypertensive when his diastolic blood pressure is greater than 95 mm Hg or his systolic blood pressure is greater than 150 mm Hg or both. Since hypertension is a dichotomous variable and this type of variable has some statistical disadvantages, in the analysis we related DDT and DDE levels to both diastolic and systolic blood pressure, which are continuous variables.

III. THE STATISTICAL TECHNIQUE EMPLOYED

To infer causal connections the technique employed for data analysis was causal statistics, often referred to as causal modeling. Causal statistics is a mathematical inquiring system which enables empirical researchers to draw causal inferences from non-experimental, quasi-experimental, and imperfectly experimental data; based on the least restrictive assumptions possible.

An early form of causal statistics is Sewall Wright's path analysis.[1,2] Another method of causal inference is the Simon-Blalock technique.[3] The form used in the research reported here is mathematically similar to econometrics, but is causal in interpretation.

The best book in the field of causal statistics is a book of readings edited by Blalock, entitled Causal Models in the Social Sciences.[4] Causal statistics hold great potential benefit for biomedical and social sciences research; it simply needs more research and more application.

IV. PROCEDURE

In order to establish causal connections, one cannot only analyze the variables of interest; i.e., DDT, DDE, diastolic blood pressure, and systolic blood pressure. One must consider other intercorrelated variables, like obesity and cholesterol. But beyond that, one must add exogenous variables until the system is identified.[5] ("Identification" is a term specific to the mathematics of econometrics and causal statistics, but it represents a state locally necessary for any causal inference.)

In this study, in order to attain identification, a total of 24 variables were simultaneously analyzed. Figure 1 shows the causal model with which we began the analysis. This model was converted into linear causal mathematical notation and two data sets were employed to estimate two parallel sets of causal parameters, one for each data set. The data sets were the Lanai-Straub (n=480) and the Honolulu-Straub (n=877) data sets collected during 1968-70 and the estimation technique employed was two stage least squares.

---

1. Wright, Sewall: Correlation and Causation," *Journal of Agricultural Research*, Vol. 20 (January 1921), pp 557-85.
2. Wright, Sewall: "The Treatment of Reciprocal Interaction, With or Without Lag, in Path Analysis," *Biometrics*, Vol. 16 (September 1960), pp 423-45.
3. Blalock, H.M., Jr.: "Four-Variable Causal Models and Partial Correlations," *American Journal of Sociology*, Vol. 68 (1962), pp 182-94.
4. Blalock, Hubert M., Jr., ed.: *Causal Models in the Social Sciences*, Aldine-Atherton, Chicago, 1971.
5. Wonnacott, Ronald J. and Wonnacott, Thomas H.: *Econometrics*, John Wiley & Sons, New York, pp 172-89 and 343-56.

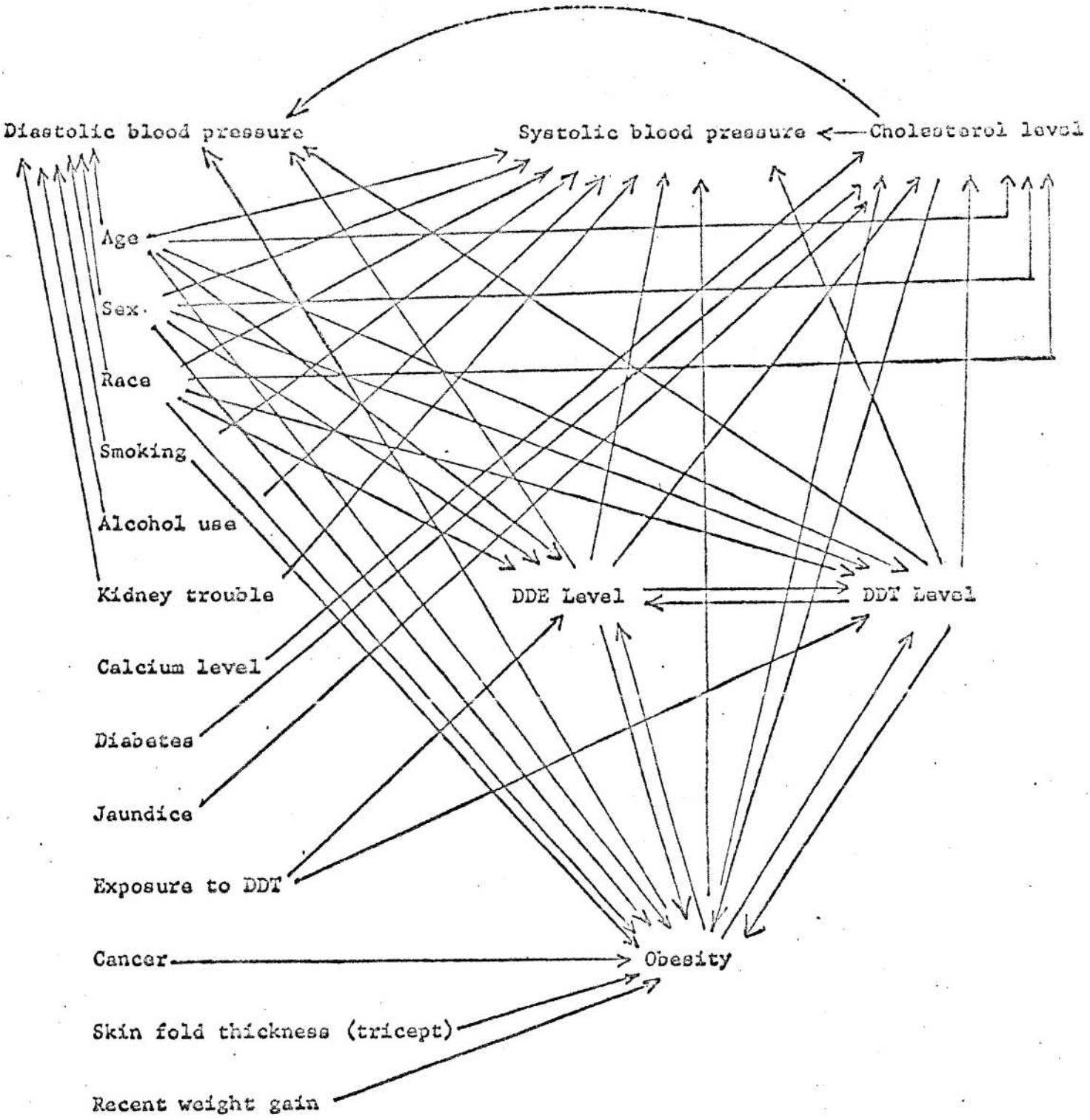
Diastolic blood pressure
Systolic blood pressure
Cholesterol level
Age
Sex
Race
Smoking
Alcohol use
Kidney trouble
Calcium level
Diabetes
Jaundice
Exposure to DDT
Cancer
Skin fold thickness (tricept)
Recent weight gain
DDE Level
DDT Level
Obesity

FIGURE 1

V. RESULTS AND CONCLUSIONS

The results of the analysis indicated that DDT is a significant cause of hypertension and that DDE is not. For the Honolulu data the causal parameters, for DDT causing both diastolic and systolic blood pressure, were positive and significant at the .05 level. For the Lanai data, DDT was found to be a significant (α = .05) positive cause of systolic blood pressure. In the Lanai data, the parameter for DDT causing diastolic blood pressure was positive, but significant at a lower level (α = .15). The causal parameters (and their t values) indicating the causal connections between DDE and blood pressure--both diastolic and systolic--were all very close to zero.

The average of the causal parameters for DDT causing diastolic blood pressure was 0.44, indicating that a one part per billion increase in DDT will--on the average--cause a .44 increase in diastolic blood pressure (measured in mm Hg). The average of the parameters found, for DDT causing systolic blood pressure, was 0.83 and its meaning is analogous.

Many other causal connections were found. For example, cholesterol, obesity, and age were determined to be significant causes of hypertension. The list of all the causal connections found is too lengthy to be presented and discussed here. But, from the conclusions above, one can begin to appreciate the great potential of causal statistics.

VI. RECOMMENDATION FOR FURTHER WORK

The analysis reported in this paper is a linear causal analysis. A more thorough analysis of the data--using more complex functional forms, interaction terms, threshold effects, etc.--is needed.

Also, it would be desirable to perform similar anlyses on additional samples and for an expanded number of variables.