Epidemiologic Studies of Diet and Cancer: The Case for International Collaboration

James R. Hebert, MSPH, Sc.D.¹,²,³

¹ Professor, Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC
² and Director, Statewide Cancer Prevention & Control Program, Hollings Cancer Center, Charleston, SC

INTRODUCTION

Over 15 years ago we made a case for conducting studies focusing on diet and cancer in populations outside the "epidemiologic mainstream". At that time, few epidemiologic studies of diet and cancer had been conducted outside of the West. Most of these had been carried out in Japan, or included a Japanese component. Notable findings from those studies included reductions in gastric cancer that appeared after modern practices of refrigeration were adopted on a large-scale following World War I.

Comparative studies conducted in Japan and in various Japanese migrant populations within the United States have been extraordinarily useful in understanding the epidemiologic transition from low to high rates of certain epithelial cancers, particularly breast and prostate. Over the same period, Japan has experienced increases in the rates of many of these cancers, whose rates lag, but roughly parallel, changes in patterns of dietary intake.

The history of studying changes in patterns of cancer incidence and mortality in Japan and among Japanese migrants has provided a valuable backdrop against which to propose conducting a variety of studies in other populations. To provide more complete background regarding international collaboration, I would like to remind readers that our main argument for a change in philosophical perspective and research strategy revolves around these three issues:

1. diet assessment methods are inadequate to estimate true exposure with reasonably sufficient accuracy and precision, especially over long periods;
2. use of retrospective diet assessment methods in case-control study designs introduces information bias; and
3. sufficient within-study-group contrasts, in both relevant exposures and cancer-related outcomes, are often lacking.

WITHIN-STUDY-GROUP CONTRASTS

With an emphasis on understanding the process of carcinogenesis, and many years of experience and thought devoted to understanding methodologic issues, I would propose that we first focus on point 3., and consider point 2. the design link to the important matter of measurement technology (point 1.). Even before publication of our diet-cancer methodology paper, we had become convinced that the homogeneity in dietary exposures evident in Western populations severely limits the ability to identify and characterize true diet-cancer
relationships.\textsuperscript{17, 18} In retrospect, this appears especially important in light of the fact that exposures are compressed into regions of the distribution that lie above or below effective thresholds. A major area of scientific contention is the relationship between dietary fat and cancer outcomes, and both laboratory animal experiments and ecological studies in humans indicate that most Western populations are exposed at higher than threshold levels.\textsuperscript{19-22}

Although limitations exist in terms of the ability to control for confounding and to model effect modification, we and others have conducted a variety of ecological studies (i.e., cross-national comparisons) of diet and cancers of various sites.\textsuperscript{23-31} Our motivation is derived mainly from an appreciation that within-study-group contrasts in both cancer rates and diet are usually lacking within any population, no matter how diverse they may seem. For example, many tens of thousands of food items exist in the larger US food databases.\textsuperscript{32-34} However, nutrient intakes tend to be very narrowly distributed, as depicted for dietary fat in Figure 1.

Data for the world are those used in previous publications\textsuperscript{23-27} and derived from the Food and Agricultural Organization of the United Nations [and therefore represent country mean values for the 150 or so countries that were available for the time period of those analyses\textsuperscript{19}]. Data from the United States are derived from the time-appropriate National Health and Nutrition Examination Survey.\textsuperscript{36} Therefore, they are based on individual-level data derived from one 24-hour diet recall interview per examinee, with adjustments made for intra-person sources of variability.\textsuperscript{17} These data are relevant to cancer rates currently available (i.e., with about a twenty-year time lag). It also should be noted that newer data show similar relative differences.\textsuperscript{17, 30}

In the early 1990s, using data from one of the largest cancer case-control studies in the world, we conducted a simulation study in which we truncated the distribution of smoking to a degree analogous to that observed when comparing the distribution of dietary fat as a proportion of energy in the US versus the world.\textsuperscript{39} Essentially, this resulted in removing the lower 55% of the distribution of smoking. The consequence was that the apparent effect of tobacco exposure on lung cancer disappeared nearly completely. The obscuring of the effect of this dominant risk factor on lung cancer is shown for women in Figure 2. Findings were virtually identical in men (though the depiction of the relationship between tobacco smoking and lung cancer is more striking in women, where the relative risk is larger than in men). Our findings are remarkable in light of the fact that we assumed a simple linear relationship (i.e., we did not model a threshold). A threshold apparently does not exist in the relationship between tobacco and lung cancer. As noted above, however, it may very well exist in the relationship between fat and breast cancer. Based on available laboratory animal evidence, this threshold is most likely to lie between 25% and 30% of calories as fat,\textsuperscript{19-22} or nearly completely beneath the entire distribution of the proportion of calories consumed as fat in the United States.

The simulation and its graphical depiction focus

Figure 1. Plot of Proportions of Total Calories as Fat – US individual vs. World mean values

Figure 2. The Effect of Truncating the Distribution of Cigarette Smoking on Detecting the Relationship between Smoking and Lung Cancer in Women (CPD = cigarettes/day)
attention on the role of exposure heterogeneity in detection of risk estimates. In neither instance do we consider variability in disease rates. Prostate is a cancer site that is of great public health and clinical concern. It is the most common cancer among men, rates vary from place to place, in many places rates are increasing over time, and there is much confusion over the etiology of the disease. In the United States, men of African origin have prostate cancer incidence rates that are about 50% higher than those of European American (EA) men. In South Carolina, the incidence disparity is higher, with rates in African Americans (AA) about 76% higher than those in EA. Mortality rate differences are even larger, with AA men having nearly triple the rate of EA men. A few other cancers evince even larger incidence differences in U.S. populations. For example, in South Carolina, African Americans have about six-fold higher incidence of squamous cell carcinoma of the esophagus than do European Americans. In international comparisons, there is generally much greater variation in disease rates. We note that in our ecological study of prostate cancer, mortality rates varied approximately 150-fold in the 59 countries used in those analyses. For other cancers, such as esophagus, differences may be up to 500-fold. So, issues of linearity and threshold notwithstanding, these larger rate differences may make it easier to detect relationships between putative risk factors and outcomes - even without overcoming known methodological limitations.

Much of the focus in epidemiology is on “outcome yield” (as indicated by high rates of disease) rather than “information yield” (as indicated by optimizing the likelihood of observing a relationship between the risk factor under study and the outcome - usually expressed as the relative risk or odds ratio). Hence, there has been a tendency to examine relationships between putative risk factor(s)-outcome relationships in populations where cancer rates tend to be very high. While this strategy seems sensible on its face, it actually may be more efficient to study a population with lower rates of cancer but higher potential relative risks for the dietary factors of interest. Recall the example depicted in Figure 2, where we concentrated only on a high-risk subpopulation. Although there might be fewer cases to study in lower-risk populations, the proportion of cases attributable to factors such as diet may be larger. Consequently, the association may be stronger and easier to detect in the face of inaccurate dietary information. This problem may, in fact, provide at least some of the explanation for the paradoxical affects observed in the ATBC trial of Finnish smokers.

**STUDY DESIGN AND DIET ASSESSMENT METHODS**

Study design options remain essentially the same as they were 15 or 20 years ago. However, analytic techniques and procedures have improved. These additional options could improve our capacity to exercise greater creativity. Regarding dietary assessment, the most striking improvements over the past 15 years include:

- the identification of specific biases in dietary self-report that appear to differ by culture, gender, and education;
- the information from which we can use to improve risk estimation in a variety of study situations;
- the development of structured assessment instruments that can be used in a variety of study populations that differ by culture and socioeconomic status, including education;
- methods to adjust for attenuation of effects due to measurable error;
- work showing that common errors may exist across dissimilar measurement instruments, such as 24-hour recall interviews and food frequency questionnaires (FFQ).

Among these developments in instrumentation are three FFQ that have been designed and tested for use in three distinct populations in India. There are many fewer FFQ than there are unique populations in which these kinds of instruments are used. We and others have emphasized the importance of questionnaire development in the populations studied. Rarely is this done adequately, if indeed at all. Each of the FFQ developed for use in India has proven to work well in context. In turn, they have been employed successfully in large-scale epidemiologic studies in the three parts in India for which they were designed. In the process of developing these instruments we have found that villagers in each of the three states (Gujarat, Kerala, and Andhra Pradesh) did a much better job of food portion size estimation than did much more highly educated subjects in Massachusetts. Additionally, we found that in comparing and contrasting data derived from these FFQ in Kerala and Gujarat that there was much greater inter-area variability than what is normally encountered in epidemiologic studies. So, for the most part, we have been able to make good progress in this aspect of the work. In addition, it must be noted that future work will need to be conducted to understand error structures in the data that appear to vary from place to place and over time.

The matter of validating dietary assessment instruments has fallen in the domains of both epidemiology and
Most such validations focus on a relative criterion measure or inter-method reliability, rather than on a true criterion. In fact, it is interesting to note that epidemiologists rarely concern themselves with either content or construct validity. Self-report of diet is essentially a psychological test. Therefore, the psychosocial context in which the questions are asked may exert an important influence on the cognitive process of recall. So, the domain of interest does include the psychosocial relationships that individuals have with food, and not just the nutrient contents of those foods.

Typically, the dietary self-reports required in epidemiologic studies entail the standard sequence in articulating response.

- Question comprehension
- Information retrieval
- Estimation/Judgment
- Response formulation

Stimulus-irrelevant characteristics (e.g., in a diet study these would be cultural and personality factors that are shaped by socioeconomic status and measured by indices of acculturation and response sets) may exert more of an influence as the duration of the behavioral exposure increases. In epidemiology we are almost always interested in long-term exposure. We use structured questionnaires such as the FFQ to obtain these data; so, the stage is set for a variety of response set biases. Given the obsession with dieting in the West, especially among women, it is plausible that dietary information would be reported with greater accuracy in a place like India. Indeed, our validation studies indicate that this may be the case. Because long time frames are preferred to short, time-delimited “snapshots” in cancer research, issues around inaccuracy in the use of structured questionnaires are of great importance. Therefore, being able to use an FFQ more effectively would argue for conducting studies in India, even if everything else were held equal.

THE CURRENT SITUATION

Our studies in India, to which we alluded, have focused on oral precancerous lesions about which the study subject was unaware. This has allowed us to overcome one of the major methodological barriers that we had identified in 1988; i.e., the possibility of exposure assessment biased in light of known disease status or by a symptom of disease. Besides our studies, a number of other studies have been conducted in South Asia. Predominantly, these have focused on high-incidence cancers. We are in a phase where opportunities call us to reconsider a focus on information yield as opposed to outcome yield. Cancers that were once uncommon in less-developed countries are now becoming more common. Parts, though rarely all, of populations are going through a remarkable demographic and nutritional transition. So, epidemiology can document such changes and concentrate on identifying their cause(s); in a context where attributable fraction (that fraction of risk attributable to a specific dietary risk factor or set of such factors) may be higher than in populations that have already made the transition (such as North America or Europe).

EPIDEMIOLOGIC CLUES, “ECOLOGICAL INVERSIONS,” AND FUTURE POSSIBILITIES

Ecological Inversions are examples of relationships that change or “invert” when the observations move between larger and smaller sampling frames, and often over time. As such, they provide potentially important etiologic clues. They also have the ability to confound conventional ways of looking at disease-risk factor relationships.

Before proceeding with this discussion it will be important to review certain underlying concepts:

- Within populations, after age, socioeconomic status (SES) is the single most important “risk factor” for cancer overall.
- Cancers rates follow distinct SES gradients. For example, in countries of affluence such as the United States, rates of hygiene-related cancers such as stomach, lung, and esophagus are inversely related to SES. In contrast, hormone-sensitive cancers, such as prostate and breast, are directly associated with affluence internationally but may evince very different patterns within populations.
- About ½ of racial differences in cancer incidence and mortality in the U.S. can be “explained” by SES. Usually education does the best job, but income also may be important.

These observations raise some interesting questions such as:

- What does SES buy?
- Does this vary from place to place and across defined subgroups within a specific place?
- Does it change over time?
These are not just academic questions. Especially for countries going through demographic transitions, including many in Asia, Africa, and Latin America, there are important implications in terms of cancer. Certain population subgroups may be especially vulnerable to changes in diet.\textsuperscript{109-112}

In affluent countries, especially the US, recruitment of SES-disadvantaged subjects into studies is often a problem. By contrast, in India we have had remarkable success, especially among the most disadvantaged.\textsuperscript{66, 71, 72} Not only is it important to ask: why?; it is equally important to understand how this might affect the inferences we draw regarding disease processes.

After decades of work, we know that SES and racial differences exist on a range of lifestyle factors, and these influence cancer stage at diagnosis, treatment options, and outcome.\textsuperscript{93-102, 106-108, 113-117} Where on the causal pathway (even to death) do these factors operate? It is important to note that some cancers reverse gradients (e.g., prostate) while others (e.g., breast) do not.\textsuperscript{28, 41, 42, 46} Why?

Before moving on to more complicated discussion of ecological inversions in relation to disease causation, we would like to present two simple examples that illustrate this concept and should condition our thinking about the relationship between socioeconomic factors and biological determinants of health. These are “Poverty Chic” and “Faux Wealth.” “Poverty Chic” is the phenomenon by which the affluent in affluent societies adopt the behaviors of the poor in poor countries. By contrast, “Faux Wealth” describes patterns of behavior of the poor in rich countries that are more typical of the rich in poor countries (or how the rich everywhere once behaved). Concomitantly, the poor in rich countries express the highest rates of conditions that, throughout human history, have been associated with wealth and therefore were typically very rare (e.g., certain cancers – but NOT all types, Type II diabetes). We now see patterns of health conditions of the poorer segments in rich countries that are similar to the more affluent segments in poor countries. “Poverty Chic” and its converse “Faux Wealth” scramble the meaning of SES, a ubiquitous determinant of human behavior and health, perhaps setting the stage for some of the paradoxical relationships we see in cancer epidemiology.

A concept central to epidemiology is confounding. The Word “confounding” derives from the Middle English confounden, the Old French confondre, and ultimately from Latin confundere - To pour or mingle together, confuse. Currently, it means:

1. To mingle and blend so that forms or natures cannot be distinguished; to throw into disorder; confuse.
2. To mistake for one another; to cause to become confused; to bewilder.

If the meaning of SES is scrambled, then is it (i.e., must it be) a confounder (in an epidemiologic sense)? Also, if SES is universally important, how can it function so inconsistently (i.e., so variably) from place to place and over time? The matter of potential confounding is important in light of the fact that the answer to questions regarding the role of diet in cancer varies so markedly across sampling frames and study types.

A major positive effect of understanding differences in cancer rates in relation to exposures that modify the effect of strong primary risk factors is illustrated by contrasts in lung cancer rates internationally. Originally, our motivation to examine causes of lung cancer other than tobacco was based on the observation that Japan has much lower lung cancer rates than one would expect based on tobacco exposure alone. This motivated a search for other factors that might explain discrepancies in lung cancer rates across many countries of the world.\textsuperscript{69} There, we showed that dietary fat was strongly associated with lung cancer mortality, suggesting effect modification of tobacco smoke by dietary fat (though we could not test effect modification in such a study). Results are consistent with laboratory animal studies showing that fat can alter metabolism of benzo[a]pyrene to a much more carcinogenic epoxide form.\textsuperscript{118}

More recently, we were asked to comment on Black-White differences in lung cancer rates in the United States.\textsuperscript{119} This illustrates that, rather than accept that 90% or so of all variability in lung cancer incidence is explained by tobacco, it is important to be guided by empirical facts that may be inconsistent with the dominant view. Likewise, our studies in India showed that dietary factors may be important even in situations where a dominant risk factor such as tobacco is operative.\textsuperscript{66, 71, 72}

Breast and prostate cancer are both hormone-sensitive cancers that have similar patterns of occurrence and identifiable, mainly dietary, risk factors in ecological studies (e.g., the correlation across countries is about 0.90 to 0.95, $R^2=0.8-0.9$).\textsuperscript{26, 27} Studies in animal models tend to corroborate the results of ecological studies, showing that the most important identifiable risk factors are those mainly dietary in nature. However, these cancers evince very different within-country patterns of occurrence. If the international pattern persisted in the United States, one would expect to see relatively low rates of both cancers...
in South Carolina, which is a relatively poor state. Indeed, BrCA incidence in South Carolina is ≤ 30th percentile of rates by state. However, PrCA incidence in South Carolina is > 95th percentile of rates by state. Incidence rates of BrCA, a cancer which is acknowledged to be influenced by hormonal factors that are, in turn, influenced by reproductive and related factors, tends not to invert. By contrast, PrCA appears not to be influenced by reproductive factors, and it does invert.

The fact that patterns of PrCA are so different within, as opposed to across, populations leads us to conclude that important risk factors are operating differently within the US. Although reproductive/hormonal factors seem to play an important role in breast carcinogenesis, half or more of the variability still remains unexplained. As noted above, in South Carolina, which may have relatively low BrCA incidence, but high mortality, 28, 41, 42, 46

Esophagus is another cancer site for which much confusion exists. As noted above, in South Carolina squamous cell carcinoma rates are about six times higher in African-Americans than in European-Americans, yet smoking and drinking rates in African-Americans are actually lower. Clearly, there must be other important factors that are not observable in conventional epidemiologic studies that must be causing these large rate differences.

It is indeed interesting to note that conventional analytic studies in epidemiology rarely identify factors associated with risk ratios greater than 3.00 or less than 0.33, even though we see incidence and mortality rates that are literally orders of magnitude (up to 500-fold for esophageal cancer) different across populations. As a matter of historical reference recall that until international studies were brought to bear on our understanding of coronary heart disease, the role of diet was not appreciated (as the relationship was confounded by physical activity).

Following are methodologic problems that could result in an “inversion.” While any one of them might be insufficient to create distortion in estimation of risk, the combination of two or more could do so.

- Homogeneity in the underlying exposure of biological interest (e.g., nutrient intake)
- Heterogeneity in the reported exposure (e.g., food intake)
- Effect modification of responses to structured questionnaires by psychological traits of the individual and societal norms
- Biological effect modification that is not uniform and consistent across two, or more, risk factors
- Incomplete control of biases due to study design (including selection of the study base population)

As with most studies, the overall goal of investigations in cancer epidemiology is to optimize accuracy and precision so as to maximize the signal: noise ratio. This, in turn, depends on a variety of design choices. These design choices and the success in implementing them depend on the world view and knowledge base of the scientists designing studies, and the perceptions of study participants. By expanding our view of what is desirable (or even possible) these inversions can become our allies in the search for the underlying causes of cancer.

Careful design and execution of inter-regional, including cross-national, studies of diet and cancer offer improved scope for making significant progress in this field. By finding ways to combine and compare distinct populations (especially ones in which apparent risk relationships tend to invert) we may be able to exploit potentially large contrasts in risk factors and disease outcome. In order to reap the potential benefit of observing large relative risks we will have to confront the methodologic challenges of this approach, including advancements in measurement methodologies and analytic approaches to improve our ability to control confounding and to estimate effect modification, including across different levels of aggregation. Developments over the past 15 years include two large-scale trials—the EPIC study, involving 10 European countries, and the Hawaii/Los Angeles Multiethic Cohort Study involving various ethnic groups in the US. Having already experienced all but the last stage of the nutrition transition, it remains to be seen if variations in these populations are sufficient to allow for qualitative improvements in risk estimation. There are virtually no inter-regional studies of diet and cancer in populations undergoing demographic, dietary, and epidemiologic transitions of the type currently underway in South Asia and other parts of the developing world. The challenge will be to design and execute studies in situations that promise the greatest information yield; for example, comparing states within India and the US across which risk relationships appear to invert.
Acknowledgements

The author thanks Drs. Donald R. Miller and Andrew B. Lawson for their helpful comments on this manuscript, and Mr. Andrew N. Smith for his assistance with graphical portrayals of the data presented in Figure 2. We would like to acknowledge funding of the South Carolina Cancer Disparities Community Network (SCCDCN) through grant number U01 CA14601-01 from the National Cancer Institute (Community Networks Program).

REFERENCES

33. NDS. Version 2.7 TheNutrition Coordinating Center, University of Minnesota, Minneapolis.
34. The Nutrition Data System for Research (NDS-R Version 4.0, 1993). In: Minneapolis, Minnesota, Developed by the Nutrition Coordinating Center (NCC), University of Minnesota.; 2001.


