This review discusses research methodology in the relation between diet and disease. Medical research can be divided into two types: complex research (the detailed study of disease mechanisms using such methods as biochemistry and molecular genetics) and simple research (the investigation of the factors that cause or prevent disease using methods such as epidemiology, intervention trials, and analogous studies on animals). Although complex research has received the bulk of resources, the large majority of our information of practical value has come from simple research. This general principle is illustrated in the area of diet and disease by examples from various areas: selenium, carotenoids, and cancer; vitamin E, ω-3 fatty acids, and coronary heart disease; dietary fat and obesity; dietary sodium and hypertension; and alcohol and stroke. Discussion then turns to aspects of the design of cohort (prospective) studies. Because of problems of sample size and relative lack of diversity, previous studies often failed to give clear-cut results. Suggestions are made concerning the design of cohort studies, notably the use of much larger numbers of subjects and with greater diversity in their diets. The problem of confounding also is discussed. Lifestyle factors often cluster together but cohort studies may not have fully unraveled this. Nutrition 2002;18:343–347. ©Elsevier Science Inc. 2002

**KEY WORDS:** cohort studies, epidemiologic research design, epidemiology, research, research design

**INTRODUCTION**

Great progress has been made since the 1970s in elucidating the relation between diet and disease. In this review the various research strategies are discussed to draw conclusions as to the types of research that have been most successful in advancing our knowledge. I also discuss specific problems in the design of cohort (prospective) studies. In addition, attention is turned to the problem of confounding.

**SIMPLE AND COMPLEX RESEARCH**

The human body may be likened to a black box of immense complexity. Various factors enter the black box and the end result is health or disease. Ideally, one may wish for a proper understanding of the inner working of the black box. In practice, however, identifying the relation between incoming factors (diet, lifestyle, and the environment) and resulting health and disease is sufficient for an understanding of how to prevent disease. Based on this analogy medical research has been divided into two types: complex and simple.1–3

**Complex Research**

This type of research consists of studies of disease mechanisms. The large majority of studies in such areas as biochemistry, physiology, and molecular genetics are examples of complex research. This type of research appeals to the human sense of curiosity. However, disease mechanisms usually prove to be so complex that we enter an almost never-ending quest for clear answers. This type of research is also referred to as mechanistic or reductionist research.

**Simple Research**

This type of research comprises those types of study where the results directly indicate what factors cause or prevent disease. Of course, the research itself may not be simple. The main forms of simple research are epidemiologic studies (ecologic, case-control, and cohort studies), intervention trials in humans, and analogous studies on animals. Biochemical and physiological measurements may be included with simple research if they are clear indicators of disease risk. Examples of such parameters include blood levels of cholesterol, glucose, and insulin.

The words complex and simple refer to the degree of complexity in translating observations into practical knowledge that can be applied to problems of human health.

In the area of diet and disease, as much as in any area of medical research, an objective evaluation shows that simple research has generated most of the important information of practical value. The following examples illustrate the superior value of simple research.

**SELENIUM AND CANCER.** There is considerable evidence from population comparison data4 and animal studies5,6 that selenium is anticarcinogenic. Findings from case-control and cohort studies have also supported a role for selenium.4 An intervention study showed a halving in cancer mortality in subjects given selenium supplements.7 Thus, this evidence from simple research has been of great value in identifying a means of preventing much cancer cheaply and safely, namely by increasing dietary intake or using supplementation. Studies of the biochemistry of selenium (complex research) have found several possible mechanisms of action.4 However, the complexity involved is such that the findings are of no apparent practical value.

**CAROTENOIDS AND CANCER.** In 1981 Peto et al.8 proposed that β-carotene prevents cancer. This inspired a large research
effort to investigate the hypothesis. Possible mechanisms include antioxidant activity, increased metabolic detoxification, and anti-inflammatory action. At the same time many case-control and cohort studies found a strong inverse relationship between intake of β-carotene and risk of cancer. Based on that finding, several intervention trials were conducted. However, these generated no evidence that supplements of β-carotene prevent cancer. Attention is now turning to other carotenoids, such as lutein and lycopene, and to other components of fruit and vegetables.

Complex research (i.e., mechanism studies) has made little contribution to our understanding of the relation between carotenoids and cancer. Indeed, such research may well have shone more darkness than light on the subject because it created a false belief that β-carotene prevents cancer. On this occasion epidemiology also produced false clues. In retrospect, there was a general failure to appreciate that the protective association between β-carotene and cancer might easily be due to confounding by other components in fruit and vegetables. This is certainly an area where epidemiology shows its limitations. Perhaps the one piece of hard information to emerge from those studies is that it is dietary patterns, namely a relatively high intake of fruit and vegetables, that provides the surest route to the prevention of cancer. The one reliable way to make progress is by intervention trials.

VITAMIN E AND CORONARY HEART DISEASE. Another important dietary antioxidant is vitamin E. It is effective in enhancing the resistance of low-density lipoprotein to oxidation. Based on that finding, it was believed that vitamin E supplements would prevent coronary heart disease (CHD). Indeed, even though the evidence lacks consistency, some epidemiologic studies have suggested an inverse association between vitamin E intake, especially from supplements, and risk of CHD. As a result of those findings, intervention trials were conducted. Subjects at high risk of CHD were given vitamin E or placebo. Although there was some evidence of a reduction in non-fatal CHD, vitamin E does not appear to prevent CHD death. Perhaps of most significance, the total mortality was actually increased slightly in groups given supplemental vitamin E.

Arguably, the clearest lesson that we learn from all this is that biochemical studies, in this case the effect of vitamin E on low-density lipoprotein oxidation, is an unreliable basis for predicting whether a substance will prevent disease. The inconsistency of the epidemiologic findings and the proneness of such data to generate misleading associations because of the problem of confounding should have made clinical investigators more cautious before large-scale intervention trials were launched.

ω-3 FATTY ACIDS AND CHD. Evidence from several case-control and cohort studies showed an inverse association between fish consumption and risk of CHD. As a high content of long-chain ω-3 fatty acids is the most pronounced nutritional difference between fish and other foods, this was the obvious food component responsible for the benefit. This provided a firm basis for intervention trials using fatty fish or fish oil. These in turn have provided solid evidence that long-chain ω-3 fatty acids are protective against CHD.

It is noteworthy that there have been few biochemical studies into how ω-3 fatty acids affect the cardiovascular disease process. The absence of such studies has clearly in no way hindered progress.

DIETARY FAT AND OBESITY. In 1998 there was a debate in the pages of the American Journal of Clinical Nutrition concerning the role of excess dietary fat in obesity. What is noteworthy about this debate is that the only data referred to came from simple research. The types of study mainly discussed were laboratory studies and intervention trials that investigated the interrelationships between dietary fat, energy intake, satiety, metabolic rate, and weight change. Also discussed were epidemiologic studies and animal studies (the effect of high-fat diets on weight gain in animals). In contrast, no reference was made to complex research such as the large literature on intermediary metabolism. Thus it appears that the debaters found little of value in that area.

DIETARY SODIUM AND HYPERTENSION. Another debate that took place in the American Journal of Clinical Nutrition concerned the relative contribution of dietary sodium to hypertension and the effectiveness of sodium restriction as a therapy. The types of study discussed were from simple research, namely population comparisons, intervention trials, and animal studies (the effect of high-salt diets on blood pressure in animals). There was a limited discussion of the findings from complex research (i.e., the mechanisms by which excess sodium affects the blood pressure), but there was little indication that these data were of particular importance in resolving the question. Rather, it appears that the mechanisms of action of salt on blood pressure were discussed mainly to help explain the findings from simple research. Indeed, it is common that, after a study based on simple research has pointed to a relation between a dietary factor and a disease (or a parameter closely associated with a disease), various mechanisms are then discussed to explain the findings.

ALCOHOL AND STROKE. It is well established that excessive alcohol intake tends to increase blood pressure. This fact strongly suggests that alcohol should be positively associated with the risk of stroke. This is indeed the case in heavy drinkers. However, in recent years, case-control studies and one cohort study have found that, as with CHD, light to moderate consumers of alcohol actually have a reduced risk of stroke. This unexpected observation stems entirely from simple research. The great many studies in such areas as the action of alcohol in the body and the pathophysiology of blood vessels made no apparent contribution to this serendipitous discovery.

LESSONS TO BE LEARNED

In the those examples we see a dearth of evidence that complex research has been of significant practical value in improving our understanding of the relation between diet and disease. Contrary examples can, of course, be pointed to. A number of new drugs may owe their origin to complex research. But, overall, an enormous research effort into the etiology of disease has contributed remarkably little to disease prevention. In contrast, although it is difficult to be exact, simple research has received well under half of the resources available for medical research but has generated the large majority of information of practical value.

Perhaps the area where there has been the greatest misdirection of resources is in cancer research. There has been an enormous research effort into the study of such complex areas as the metabolism of carcinogens and oncogenetics. Nevertheless, the bulk of our knowledge concerning how to prevent cancer comes from simple studies into the relation between lifestyle and environmental factors (mainly diet and smoking) and the risk of cancer.

There is no reason to believe that the lessons from the past
several decades have been learned. Indeed, it is the seldom disputed dogma that studies of disease mechanisms—reductionism—is the basis for medical advance. The largest current focus of medical research is into the human genome and how an understanding of it can be translated into new cures for disease. The complexity of this challenge is so immense that one must be highly pessimistic that any significant success will be achieved, certainly not within the foreseeable future.

The research strategy that has proven most productive has been a combination of epidemiologic studies followed by intervention trials. Important findings have come from all forms of epidemiologic investigations, both population comparisons (ecologic studies) and studies within a population (case-control and cohort studies). Often valuable results have also come from simple research on animals, such as the study of the effect of dietary fat on weight and of selenium on cancer incidence in animals exposed to carcinogens.

Aside from intervention trials, which are still the undisputed gold standard, of the different types of simple research, the one that has proven of most value has been cohort studies. But there have been several instances of misleading findings. Reference was made to the case of β-carotene and risk of cancer and to vitamin E and risk of CHD. In the next section, I turn my attention to the design of cohort studies.

COHORT STUDIES

The Nurses’ Health Study commenced in 1976 and is based on detailed information collected from 121 700 American female nurses. Similarly, the Health Professionals. Follow-up Study, which started in 1986, is based on 51 500 American male health professionals. These studies have generated large amounts of valuable information. Despite costs running into several millions of dollars each year, few would disagree that they are excellent value for the money.

Even with the seemingly enormous size of these studies, many of the observations have not been statistically significant; this is because the sample size still poses limitations. This is illustrated by the following examples.

1. In the Health Professionals Follow-up Study, 812 new cases of prostate cancer developed in the first 6 y. Lycopene, a carotenoid found in tomatoes, showed a negative association with this disease (relative risk [RR] = 0.79), but the relation was only marginally significant.25 Whereas tomatoes and tomato sauce showed an inverse relationship, tomato juice did not. These results may be of considerable importance for the prevention of this fairly common cancer. Nevertheless, despite an apparent 21% fall in risk, the results were not clear-cut.

2. In a publication on the Nurses’ Health Study, findings were presented based on 442 cases of colon cancer developed during 14 y.26 A high intake of multivitamins containing folate was associated with a decreased risk. The RRs were 0.83 at years 5 to 9 and 0.80 at years 10 to 14; neither association was significant. However, the association became strong and significant after 15 y (RR = 0.25). These results produced uncertainty as to whether folate supplements given for fewer than 10 y lead to an appreciable reduction in the risk of colon cancer.

3. In another report from the Nurses’ Health Study, data were given based on 2956 cases of breast cancer that developed during 14 y.27 The surprising observation associated trans fatty acids with an 8% decreased risk but α-3 fatty acids from fish with a 9% increased risk. Both associations were marginally significant. Possibly these observations were due to confounding or represented only random “noise.” Alternately, they may be real; in that case they are potentially important, for the control of this important cancer and our understanding of the health and safety effects of various types of fat.

To gain a clearer picture, it is necessary to expand the sample sizes of the cited studies several-fold because to halve the confidence intervals the population size must increase by a factor of four. Studies with such sample sizes would help clarify the roles of major factors in minor diseases and of minor factors in major diseases. For example, we would better understand the role of diet in gallstones, arthritis, osteoporosis, eye disease, stroke, various less common cancers, Alzheimer’s disease, and the decline in memory and lung function with aging.

Studies of this size would also help clarify the role of antioxidants and phytochemicals in various diseases. The data would indicate which protective substances have the strongest (inverse) associations with disease. Moreover, the large sample size would facilitate dealing with the problem of confounding (see the PROBLEM OF CONFOUNDING).

Relative lack of diversity of diets has been a shortcoming of previous studies. For instance, in the Nurses’ Health Study, the mean fat intake in the lowest quintile was 32% of total energy,28 a level well above a low-fat diet. In the Health Professionals Follow-up Study, the median intake of vitamin C in the lowest quintile was 92 mg,29 which indicates that the entire sample had an adequate intake. In that study, the median intake of dietary fiber in the highest quintile was 28.9 g.30 a level still well short of a truly high-fiber intake.

The effect of a lack of diversity is that the resulting RR moves closer to unity.31 For example, if the RR of a high versus low intake of vitamin X for risk of cancer is 0.6 in a population with a wide range of intake, then the RR may be 0.8 where the range of intake is less. As a result, more subjects are required to achieve the same level of statistical significance. Thus, sample size and diversity are closely related: as the degree of diversity increases, the number of subjects required to detect a significant difference decreases.

To enhance diversity, it would be valuable to throw the net across much of Europe. Inclusion of economically poor countries such as Cyprus, Albania, Ukraine, and Romania would magnify contrasts in diet and lifestyle. Another advantage of recruiting across Europe is to facilitate recruitment. Other possible regions suitable for study are the Americas (North, South, and Caribbean) and the countries of East and Southeast Asia.

The ideal population for such a study is one that could be followed for at least 20 y, consists of reasonably responsible and cooperative people aged 40 to 50 y, and have a high degree of variability in diet and lifestyle. For instance, physicians may have less value than other people because they generally have low rates of smoking. Suitable groups might include teachers, civil servants, police, and nurses.

Three large cohort studies have been launched in recent years.32–34 In each case subjects were recruited from the general population. One is being done as part of the Women’s Health Initiative.32 It comprises 100 000 women aged 50 to 79 y who are being followed for 9 y. The statistical power is therefore comparable to that of Nurses’ Health Study. The other two studies are much larger. However, both have a focus on cancer and the design suggests that minimal effort will be made toward collection of data relevant to other diseases, especially non-fatal ones. The European Prospective Investigation Into Cancer and Nutrition study involves 484 000 men and women from nine countries of the European Union.33 The age range at baseline is 35 (women) or 40 (men) to between 60 and 74 y. The Multiethnic Cohort Study is being carried out on 215 250 men and women age 45 to 75 y at baseline.34 The subjects are from the ethnically diverse populations of Hawaii and Los Angeles. One limitation of this study is that blood samples were not collected.
The Problem of Confounding

The problem of confounding is a ubiquitous problem in cohort studies. Insufficient attention has been paid to the tendency for various lifestyle factors to cluster together; examples follow.

1. In the Nurses’ Health Study, women in the lowest quintile for folate intake had higher levels of smoking but lower levels of regular exercise compared with women in the middle quintile (36.9 versus 26.4% for smoking and 35.0 versus 47.1% for exercise, respectively).35

2. In the Health Professionals Follow-Up Study, men in the lowest quintile for intake of β-carotene and vitamins C and E took only half as much exercise but were 56% to 100% more likely to smoke compared with those in the top quintile.36

3. Consistent with the previous study, a survey of American adults associated low intake of fruit and vegetables with being sedentary, a heavy smoker, and a heavy drinker.37

4. A study in Wisconsin indicated that users of supplements tended to be female, better educated, to have a relatively low body mass index, to have never smoked, to be more likely to exercise, and to consume a more nutrient-dense diet.38

These data point to clustering of lifestyle factors. A healthy lifestyle cluster includes exercise, not smoking, and consuming a relatively high intake of fruit, vegetables, and various nutrients, but less animal fat. In theory, these factors should not present a problem in cohort studies because they can be corrected for in multivariate analysis. However, Marshall and Greenland39 drew attention to the problem of how measurement errors of strong confounders can lead to erroneous conclusions with regard to the relation between a weak risk factor and outcome. An example of this is the protective association between intake of β-carotene and cancer. It is possible that inaccurate estimations of smoking (a strong confounder) may have exaggerated the strength of the negative association between β-carotene and cancer.

Other factors may be associated with a healthy lifestyle and have generally been overlooked in previous studies. Such factors include avoidance of passive smoking, early diagnosis of disease combined with better treatment, and perhaps differences in psychological functioning. Socioeconomic status also has strong associations with health.40–42 Socioeconomic status includes economic status, job status, education, and area of residence. However, it is doubtful whether previous cohort studies have fully corrected for it. As a result of these various associations, cohort studies often have unrecognized errors of confounding. Margetts43 recently discussed the question of confounding in cohort studies.

Conclusion

This review is not a blueprint for fruitful research but may be a source of stimulating ideas that will lead to more insightful planning of research strategies. Most importantly, there should be a paradigm change in our priorities for nutrition research and, more generally, medical research. That means a reduced emphasis on “glamor” areas, most notably genomic medicine. There should instead be a higher priority toward such lines of research as how dietary patterns affect disease risk, the role of particular dietary components (such as phytochemicals), and the potential value of dietary supplements.

Cohort studies have played a vital role in unraveling the associations between diet and disease. However, in recent years there has been a growing appreciation of the various sources of error associated with this research tool. In particular, cohort studies need to be carried out on a truly large scale (200 000 to 400 000 subjects) and with wide diversity of diet and lifestyle. At the same time, we need a better understanding of how to minimize errors caused by confounding. This will entail improved accuracy of how various parameters are estimated and a better appreciation of the factors that may lead to confounding.

References

2. Temple NJ. Simplicity—the key to simple medical research. Med Hypotheses 1985;17:139

346 Temple Nutrition Volume 18, Number 4, 2002

(For an additional perspective, see Editorial Opinions)