Biostatistics (2001), **2**, 4, *pp.* 365–381
Printed in Great Britain

Methodologic challenges in chronic disease population research

ROSS L. PRENTICE
Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, MP-1002, Seattle, WA 98109-1024, USA
rprentic@fhcrc.org

SUMMARY

Some recent important, and controversial, chronic disease population research settings are reviewed. These include studies of hormone replacement therapy and coronary heart disease; studies of dietary fat in relation to breast cancer; and studies of beta-carotene supplementation and lung cancer. In each case methodologic developments having a strong biostatistical component are identified as key to future progress. Some comments are also made on the need for an expanded disease prevention intervention development enterprise, and on the role that microarray genetic and genomic data may play in such development.

Keywords: Adherence bias; Benefits versus risk assessment; Measurement error; Microarray technology; Multivariate outcomes; Objective exposure measures.

1. INTRODUCTION

The ultimate goal of chronic disease population research is to identify practical approaches to reduce the incidence or mortality from chronic diseases in targeted populations. It is crucial that the maneuvers to be recommended, whether behavioral or pharmaceutical, convey benefits that exceed risks. This implies an obligation to assess the impact of such interventions or treatments on a spectrum of health-related outcomes, and to do so over a sufficient duration to allow an informative benefit-versus-risk assessment.

While randomized controlled trials are a natural gold standard for the assessment of preventive interventions they are attended by a number of challenges in this chronic disease population research setting, mainly related to the large sample sizes that may be needed, the difficulty in ensuring adequate adherence to assigned interventions over a possibly lengthy follow-up period, and the attendant costs. At a minimum these challenges imply the need for a careful intervention development and screening program, so that only interventions having major public health potential are tested in randomized controlled trials. Alternatively one could take the viewpoint that the obstacles just mentioned are sufficiently daunting that purely observational approaches should be relied on for preventive intervention identification.

Three controversial research settings where there has been much observational study, along with ongoing or completed intervention trials, will be reviewed in an attempt to cast light on the fundamental issue of need for randomized controlled trials in chronic disease population research. This review will bring out some more specialized research questions where methodology development is needed. Also, some comments will be given on the importance of a major interdisciplinary effort to identify promising chronic disease prevention hypotheses.

Table 1. *Coronary heart disease relative risk estimates in relation to the current use of hormone replacement therapy, and related methodologic issues* (*much observational HRT data corresponds to the use of estrogen alone whereas the HERS study compared a specific estrogen/progestin regimen to placebo*)

Meta-analysis (Stampfer and Colditz, 1991) of 31 observational studies		
Summary relative risk estimate (95% confidence interval):	0.56	(0.50, 0.61)
Randomized trial (HERS: Hulley *et al.*, 1998) among women with established coronary disease		
Summary relative risk estimate (95% confidence interval):	0.99	(0.80, 1.22)
Methodologic issues and needed research		
• Reliability of purely observational approaches		
— can prescription (confounding) bias be controlled, especially concerning difficult to measure characteristics (e.g. physical activity patterns, dietary patterns)		
— can potential adherence bias be quantified and controlled		
• Criteria for initiating intermediate outcome trials		
• Criteria for timely initiation of full-scale disease prevention trials		
• Methods for trial monitoring and for summary benefit versus risk assessment for interventions that target multiple organs		

2. LESSONS FROM STUDIES OF HORMONE REPLACEMENT THERAPY AND CORONARY HEART DISEASE

2.1 *Summary of available research*

As a first specific context consider the hypothesis that use of hormone replacement therapy among postmenopausal women may reduce the risk of coronary heart disease (CHD), a leading cause of morbidity and mortality in Western populations. Replacement hormones, initially in the form of estrogen alone, became popular in the late 1960s and 1970s, for the principal purpose of controlling menopausal symptoms. By the mid-1970s the use of these preparations had been associated with a 5–10-fold increase in uterine cancer, after which estrogen dosages were reduced, and preparations that combined estrogen with a progestin to protect the uterus became available. Since the mid-1980s replacement hormones have become very popular in the United States, and much epidemiologic study of the association of their use with chronic disease risk ensued. Of particular interest was the fact that a majority of more than 30 observational studies reported lower coronary heart disease risk among estrogen users. Meta-analyses of these studies (Stampfer and Colditz, 1991; Grady *et al.*, 1992) pointed to a 40–50% lower coronary disease risk among current estrogen users (Table 1).

Exogenous ovarian steroid hormones have multiple target tissues in addition to the heart, including other elements of the vascular system, the breast, the uterus and the bone, among others. Reports from observational studies suggest adverse associations with venous thrombolic disease, and breast cancer among long-term users, as well as favorable associations with fracture rates and with total mortality. To cite a recent example of this last important potential benefit, Grotstein *et al.* (1997) reporting data from the Nurses Health Study, write that 'after adjusting for confounding variables, current hormone users had a lower risk of death' with a relative risk estimate of 0.63 and corresponding 95% confidence interval of (0.56, 0.70). The authors went on to state that 'current hormone users with coronary risk factors had

the largest reduction in mortality' ($RR = 0.51$, 95% CI 0.45–0.57). If the causal language used in these statements is appropriate then hormone replacement therapy would be an extremely important approach to chronic disease prevention in the United States.

The body of observational data is still considerably less extensive for combined preparations than for estrogen alone, but there are reports to the effect that the addition of a progestin may blunt the putative cardioprotective effect of estrogen and may augment the putative beneficial effects on bone, and the putative adverse effects on the breast.

The observational data were sufficiently encouraging that the National Heart Lung and Blood Institute initiated the postmenopausal estrogen/progestin interventions (PEPI) trial (PEPI Trial Writing Group, 1995). This three-year, multicenter, randomized, double-blind, placebo-controlled trial compared the effect of four hormone replacement therapies to placebo in respect to selected heart disease risk factors, as well as risk factors for endometrial cancer and fractures. The trial included 875 healthy postmenopausal women aged 45–64. In 1995 this study reported (PEPI Trial Writing Group, 1995) that estrogen alone or in combination with a progestin improves lipoprotein profiles and lowers fibrinogen levels without detectable effects on post-challenge insulin or blood pressure. Estrogen alone gave the largest increase in high-density lipoprotein (HDL) cholesterol, a favorable risk factor change, while all preparations induced favorable reductions in low-density lipoprotein (LDL) cholesterol. A sufficiently high rate of endometrial hyperplasia arose on the estrogen-only preparation to preclude its long-term usefulness among women with a uterus.

The heart and estrogen/progestin replacement study (HERS) was initiated in 1993 as enthusiasm for the widespread use of hormone replacement therapy escalated in the United States. HERS was a secondary-prevention randomized trial among 2763 postmenopausal women with established coronary disease. It included only women with an intact uterus, who were randomized to 0.625 mg d^{-1} conjugated equine estrogen plus 2.5 mg d^{-1} medroxyprogesterone acetate, the most prevalent combined preparation in use in the USA, or placebo. Women assigned to the active preparation had on average a 10% higher HDL cholesterol and 11% lower LDL cholesterol, consistent with PEPI study results. However, when the study reached its planned termination (Hulley *et al.*, 1998) no difference was observed between the active and placebo groups in respect to CHD incidence. In fact, the estimated relative risk (Table 1) was 0.99 with a 95% confidence interval of 0.80–1.22 based on 172 events in the active arm and 176 in the placebo arm. The number of CHD deaths was 71 in the active and 58 in the placebo arm, giving an estimated RR (95% confidence interval) of 1.24 (0.87, 1.75). A significant excess of venous thromboembolic events was reported by HERS investigators prior to study termination. A post hoc examination of active versus placebo relative risks for CHD incidence as a function of time from randomization gave RR estimates (95% confidence intervals) of 1.52 (1.01, 2.29) in the first year from randomization, 1.00 (0.67, 1.49) in the second year, 0.87 (0.55, 1.37) in the third year, and 0.67 (0.43, 1.04) after the third year, based on an average 4.1 years of follow-up. These CHD data suggest a trend from early harm to later benefit as the duration of use of this combined preparation increases.

2.2 *Discussion following presentation of HERS findings*

The unexpected results from the HERS study have generated much discussion. A key aspect of such discussion is the issue of the reliability that attends observational studies and clinical trials of preventive interventions, such as hormone replacement therapy. It is instructive to consider some of the related editorials. Pettiti (1998) in a *JAMA* editorial that accompanied the HERS results (Hulley *et al.*, 1998), writes that 'these observations are a sobering reminder of the limitations of observational research, the incompleteness of current understanding of the mechanisms of vascular disease and the dangers of extrapolation.' Referring to CHD incidence rates among control group subjects who were compliant to their assigned pills versus those were not, in previous randomized trials of CHD Pettiti went on to

offer the interesting suggestion that 'compliance bias is large enough to explain entirely reductions in the relative risk of CHD between users and non-users of estrogen replacement therapy and hormone replacement therapy of the magnitude found in observational studies.' Concerning the suggested time trend in relative risk in HERS she argued that 'the lipid hypothesis has dominated thinking about CHD for at least four decades. There is a growing recognition that thrombotic phenomena play an important role in acute coronary syndromes.' Finally, as a more general observation concerning research strategy Pettiti argued that 'when an exposure can be assigned at random, it should be assigned randomly. Commitment to randomized trials as the standard of proof must be especially strong when the public health implications are so great.'

Whitehead and Stampfer (1998) mounted a defense of the role of observational studies in this area in the journal *Climacteric*. They wrote an editorial reminding readers that HERS was a trial of only one hormone replacement preparation, and that it was conducted only among women with established coronary disease. They went on to criticize the HERS study stating that its failure to achieve the planned period of observation (4.1 years versus 4.75 years average follow-up due to recruitment lags) was a 'major flaw' and noting 'other problems', most notably some difference between projected and realized pill taking adherence rates. Furthermore, based on CHD analyses among women with a prior history of cardiovascular disease, that were evidently stimulated by the HERS results, Whitehead and Stampfer wrote that 'Preliminary data from the Nurses Health Study support a pattern of early transient increase followed by a substantial decline in risk as the duration of therapy is extended. Thus the results of HERS do not contradict the directly relevant observational data.' They finish with a broader concern about randomized trials stating that 'unless prospective, randomized trials possess sufficient power then definitive conclusions cannot be drawn and more questions than answers will result', and they conclude that 'This, regrettably, is the HERS legacy.'

Utian (1998) questions whether the HERS findings should impact medical practice in an editorial for *Menopause Management* entitled 'HRT and the HERS findings—has the ground shifted.' He questions the HERS study design, particularly the exclusion of an estrogen-only comparison and writes 'Unfortunately HERS was poorly conceived and designed, taking an unnecessary gamble that has now come back to haunt all parties concerned, and confuse consumers and providers alike.' He went on to write that there are ' . . . already indications that progestins might attenuate some of the estrogen-induced cardiac-benefit effect. HERS would therefore have best included an estrogen-only arm, or have been designed as an estrogen versus placebo comparative study.' Gaining momentum, he explained the HERS time trend in CHD relative risk by stating that 'The initial negative effect is almost certainly due to the attenuation of estrogen-induced increase in coronary flow' by the progestin in the replacement hormone regimen studied in HERS.

2.3 *Methodologic research needs*

The hormone replacement therapy and CHD story raises important questions (Table 1) about the reliability that can be achieved using purely observational approaches. This is a context in which key aspects of the exposure of interest, including current use of replacement hormones, are likely recalled and reported fairly accurately, adding to the potential of observational studies. On the other hand the characteristics of women that might enhance the likelihood of their choosing, or of their being assigned to, hormone replacement therapy may be too subtle and too extensive to ever be well controlled in a cohort or case-control study. Hormone replacement therapy users are known to have a favorable CHD risk profile compared to non-users in respect to some standard risk factors, but one wonders also about differences for less standard risk factors, including lifetime endogenous hormone levels, difficult to measure dietary patterns, and physical activity patterns, that may confound observational studies. Equally important is the issue raised by Pettiti concerning possible differences in CHD risk between hormone replacement therapy users who

become long-term users of these preparations versus those who do not. Myriad biobehavioral factors could underlie any such difference with important implications for observational study bias, and with resulting implications for randomized trial analyses beyond intention-to-treat comparisons. Methodologic research and additional pertinent data collection is needed to meaningfully quantify the magnitude of confounding and adherence biases that may attend observational studies in this and other areas. For confounding, there is a need for improved methods to control for confounding by factors that are measured with considerable error. Concerning adherence bias there is a need to systematically report the risk of targeted diseases as a function of the extent of adherence to blinded, placebo regimens as trials are completed.

There are also a range of issues relating to intermediate outcome trials, like PEPI. First it is important that such trials are initiated early for treatments, such as hormone replacement therapy that have substantial public health implications, especially if the treatments in question are coming into widespread use. The PEPI study also demonstrates, once again (see, for example, Fleming *et al.*, 1994), that even when an attempt is made to include a range of pertinent intermediate outcomes, and the treatment effect is as expected on pertinent intermediate outcomes, it does not necessarily follow that the treatment will show the hypothesized effect on the clinical outcomes of greatest interest. In fact, the circumstances under which an intermediate outcome, or group of outcomes, can stand as a surrogate for a 'true' outcome in evaluating a treatment are quite restrictive (see, for example, Prentice, 1989). Another statistical challenge relates to the need for criteria that would either justify proceeding with a full-scale intervention trial with clinical outcomes, or that would argue against further intervention evaluation.

Other important questions arise in the design and analysis of intervention trials with disease outcomes. In a specific context, such as the evaluation of a hormone replacement therapy preparation, it is important to ask whether a trial of sufficient size and duration, with adequate adherence patterns can be conducted to provide a powerful hypothesis test, and a sufficiently comprehensive picture of benefits versus risks. Answering these questions will typically require a noteworthy pilot study, perhaps doubling as an intermediate outcomes trial.

Interesting statistical issues arise also in the monitoring and reporting of trials in which the intervention is hypothesized to have a range of benefits and risks. One might propose to emphasize total mortality, for example, as an integrated summary of benefits and risks, in monitoring for early trial stoppage and for eventual trial interpretation. However, total mortality will typically be too insensitive for these purposes, in the primary prevention setting. Instead one might consider the monitoring of designated primary outcomes and anticipated adverse effects, along with some type of global index that combines outcomes that correspond to serious diseases that are plausibly affected by the intervention. For example, the ongoing Women's Health Initiative (WHI) clinical trial includes the study of the same estrogen/progestin preparation used in HERS among women with a uterus and estrogen alone among hysterectomized women, in a total of 27 348 postmenopausal women (WHI Study Group, 1998). Monitoring guidelines in this trial involve significance levels for the designated primary outcome—CHD incidence, the primary hypothesized adverse effect—breast cancer, and a global index consisting of the earliest these outcomes or stroke, pulmonary embolism, hip fractures or other cause mortality, and for the combined regimen the earliest of each of these outcomes and endometrial cancer incidence (Freedman *et al.*, 1996). There are a range of related analysis and reporting challenges involving, for example, the estimation of relative risk parameters that take account of such a complex monitoring plan, and especially the need for multivariate response data analysis methods that yield meaningful and comprehensive benefit-versus-risk summaries. Such analyses may be critical to the formulation of useful recommendations to women concerning the use of these preparations.

3. LESSONS FROM STUDIES OF DIETARY FAT CONSUMPTION AND BREAST CANCER AMONG POSTMENOPAUSAL WOMEN

3.1 *Summary of available data*

The hypothesis that a reduction in dietary fat in Western populations will lead to reductions in breast cancer risk has been promulgated for some decades. Its evaluation has been the goal of considerable research, and considerable controversy: see Greenwald (1999) and Hunter (1999) for recent examples. In effect, the testing of this hypothesis has become a prominent test case in a debate over the study designs and research strategies that are needed to obtain reliable information in the broader diet and chronic disease research area. On the one hand, some believe that reliable information on the fat and breast cancer relationship can be obtained from self-reported assessments of diet in the context of observational studies. This perspective led investigators involved in a recent pooled analysis (Hunter *et al.*, 1996) of cohort studies of fat consumption and breast cancer risk to conclude that 'In the context of the Western lifestyle, lowering the total intake of fat in midlife is unlikely to reduce the risk of breast cancer substantially.' On the other hand the National Institutes of Health-sponsored Womens Health Initiative (WHI Study Group, 1998) is about halfway to completion. This initiative includes a randomized controlled clinical trial, one component of which is assessing the impact of a 'low fat eating pattern' on the incidence of breast cancer, colorectal cancer, and secondarily CHD, among 48 837 postmenopausal women in the United States. The existence of this large, complex intervention trial reflects the view of others that a low fat eating pattern and breast cancer prevention hypothesis is worth testing, and that observational studies alone may not be able to provide sufficiently reliable information on the fat and breast cancer association, regardless of their size or duration.

Experimental studies in rodents dating back to the 1940s (Tannenbaum, 1942) pointed to the ability of a high fat diet to promote mammary tumors. A question ensued (Albanes, 1987) as to whether fat specifically, or energy more generally, explained such promotion. A meta-analysis (Freedman *et al.*, 1990) of data from rodent feeding studies confirmed a substantial role for calorie restriction in inhibiting mammary tumorigenesis, but also revealed an important role for fat reduction, beyond that attributable to the corresponding energy reduction.

More recently various groups (see, for example, Armstrong and Doll, 1975; Gray *et al.*, 1979; Prentice and Sheppard, 1990) have reported a positive relationship between national estimates of the per capita supply of fat and the corresponding lagged rates of breast cancer incidence or mortality, or a positive relationship between changes in per capita fat supply and corresponding changes in lagged breast cancer rates. For example, regression analyses presented in Prentice and Sheppard (1990) suggest that a 50% reduction in fat consumption in the United States could eventually be associated with a lower breast cancer risk by a factor of about 0.39 among women in the age range 55–69 and by a factor of about 0.53 among women in the 30–44 age range.

It is fortunate that the individual level data from most of the case-control and cohort studies of dietary fat and breast cancer studies have been assembled for standardized analyses. Howe *et al.* (1990) carried out a combined analysis of 12 case-control studies involving 4427 breast cancer cases and 6095 controls, about two-thirds of whom were postmenopausal. Table 2 shows their summary postmenopausal breast cancer relative risk estimates across the total fat intake categories defined by the quintiles of one of the Canadian case-control studies. A highly significant trend ($p = 0.0002$) was reported. The relative risk trend was much less pronounced among premenopausal women and was non-significant ($p = 0.21$). Table 2 also shows relative risks projected from the international correlational analyses. These projections (Prentice, 1996) are based on a regression of log breast cancer incidence in the age range 55–69 on log total fat, and assume a classical measurement model for food frequency and four-day food record assessments of log fat intake. The projections are actually onto the intake quintiles of the baseline food frequency data in the Women's Health Trial (Insull *et al.*, 1990) which differ little from those in the Canadian study

Table 2. *Relative risk estimates from pooled analyses of observational studies of fat intake and postmenopausal breast cancer risk, along with corresponding projections from international correlational analyses, and related methodologic issues* (*most of the case-control studies and all cohort studies used a food frequency assessment of diet; international data projections include an accommodation of food frequency measurement error*)

	Total fat consumption quintile				
	1[†]	2	3	4	5
Case-control studies (Howe *et al.*, 1990, Table 7)	1	1.20	1.24	1.24	1.46
	Trend test ($p = 0.0002$)				
International data projection[‡]	1	1.11	1.19	1.28	1.42
	Calorie-adjusted fat intake quintile				
Cohort studies (Hunter *et al.*, 1996, Table 2)	1	1.01	1.12	1.07	1.05
	Trent test ($p = 0.21$)				
International data projection[‡]	1	1.07	1.12	1.19	1.27

Methodologic issues and needed research

- Reliability of observational approaches given that the diet a complex mixture of highly correlated nutrients and foods, some components of which have limited range of variation
- More flexible measurement models that acknowledge possible systematic bias in dietary assessment needed
- Objective measures of nutrient intake on appropriate subsamples as a fundamental component of data analysis
- The development of a varied research program involving observational approaches (e.g. aggregate and analytic) having different potential biases
- A strengthened intervention development enterprise, based on small scale human feeding studies
- A commitment to intervention trials for testing the dietary hypotheses having the greatest public health potential

[†]Reference category.
[‡]Projected onto quintiles defined by food frequency data in Women's Health Trial (see Prentice, 1996).

used by Howe and colleagues. A close correspondence between the case-control and the international data estimates can be seen. Howe *et al.* (1990) commented further that there was no evidence that nonfat sources of energy were associated with breast cancer risk, upon allowing for total fat, in their analyses.

Hunter *et al.* (1996) provided a pooled analysis of seven cohort studies including 4980 breast cancer cases out of 337 819 women. Summary relative risk estimates were presented across quintiles of calorie-adjusted fat intake (residual from regression of logarithm of fat on logarithm of calories), based on food frequency dietary assessments, and are shown in Table 2. The breast cancer risk did not show a significant trend ($p = 0.21$) across intake categories. Table 2 also shows corresponding relative risk projections from international correlational analyses, again with measurement error acknowledged using a classical measurement model. The projection categories (Prentice, 1996) are based on baseline food frequency percent energy from fat from the Women's Health Trial, which are likely to differ little from the calorie-adjusted fat categories used in the pooled cohort analyses. The projections are based on international breast cancer rates for the age range 55–69, so that these could be reduced somewhat to acknowledge the inclusion of premenopausal women in the pooled cohort analyses. Note the difference between observed and projected relative risks in the upper categories of calorie-adjusted fat intake.

3.2 *Debate concerning the interpretation of these data*

The relative risk estimates and projections shown in Table 2 apparently provide the basis for the controversy over the fat and breast cancer hypothesis. Population scientists are accustomed to viewing cohort studies as the most reliable of the observational study strategies so that it can be argued, as in Hunter *et al.* (1996) and Hunter (1999), that the cohort study lack of association noted in Table 2, the possibility of selection or recall biases in the case-control studies (Friedenreich *et al.*, 1991; Giovannucci *et al.*, 1993), a lack of ability to control confounding in international correlational analysis, and a lack of relevance of animal feeding experiments indicate that dietary fat reduction is unlikely to meaningfully reduce breast cancer among middle-aged and older women.

On the other hand one can regard these other data sources as suggestive of an important fat and breast cancer relationship and can question the extent to which the cohort analyses provide contrary evidence. Certainly the relative risk trend projected from international correlational analyses (Table 2) is modest enough to be well within the range that may not be reliably detected in observational studies, regardless of their size.

More specifically, certain nutrient intakes in Western populations may not be highly variable, in spite of the variety of foods available. For example, the vast majority of Americans likely currently have dietary percent energy from fat between 27 and 40, resulting in modest relative risk gradients within study populations, even if the nutrient exposures are important risk factors. Also the diet is a complex mixture of foods and nutrients including many highly correlated elements. Even if the nutrient exposures of interest could be measured exactly it may be a formidable task to estimate the relationship between a specific nutrient, such as fat, and breast cancer risk, while accommodating other dietary factors. Along the same lines, dietary patterns may relate in a complicated fashion to other breast cancer risk factors including anthropometric measures, reproductive factors and physical activity patterns. Hence, there is potential for confounding (or over-control) which may be unavoidable for such difficult to measure exposures as physical activity patterns. Furthermore, key measurement properties of existing dietary self-report instruments are unknown, even in the populations where these instruments have been applied.

These limitations of cohort and case-control studies combine to imply an uncertain interpretation of the data given in Table 2. Of the issues just listed, measurement error in dietary assessment is the least well understood and perhaps the most crucial to the interpretation of the analytic epidemiologic studies. Often in the reporting of these studies the authors will write words to the effect that analyses were conducted that correct for measurement error in dietary assessment (see, for example, Hunter *et al.*, 1996, p. 361) and that this correction did not alter study findings. It is important to understand the assumptions that underly such measurement error adjustments.

Available measurement error methods require a nutrient consumption assessment that estimates the quantity of interest aside from an additive error that must be independent of the targeted quantity and of other study subject characteristics. To be specific, let Z represent a woman's average daily total fat intake (or percent energy from fat) over a time period pertinent to breast cancer risk (e.g. preceding decade). One needs to assume that an estimate (X_1), where

$$X_1 = Z + e_1 \tag{3.1}$$

is available, at least on a subsample of the study population. The error term e_1 is assumed to be independent of the 'true' intake Z, so that women with high fat consumption are no more likely to underestimate their fat intake than are women consuming less fat. Also, the measurement error e_1 is typically assumed to be independent of such study subject characteristics or exposures as ethnicity, age, physical activity patterns or body mass. These assumptions are typically applied to data from food records or recalls collected on a small subsample of study subjects. Self-report data using a less expensive and less comprehensive food frequency questionnaire are typically obtained on all study subjects. The most flexible measurement

model for the food frequency estimate of fat that has been used in these studies supposes that the food frequency estimate X_2 can be written as

$$X_2 = Z^* + e_2 \tag{3.2}$$

where Z^* is a variable positively correlated with Z and where the error term e_2 is independent of Z^* and independent of other study subject characteristics and, importantly, is independent of the error term e_1. These assumptions allow X_1 to be used to 'calibrate' the food frequency measure in that X_2 can be replaced by an estimate of the expectation of Z given X_2 in relative risk analysis. These regression calibration methods (see, for example, Carroll *et al.*, 1995) work well when the measurement error is not too large and the assumptions listed above hold. In the diet and chronic disease area, however, the measurement errors are not small, as can be seen by observing the modest correlations between nutrient intake estimates from multiple self-report instruments at a single point in time (e.g. about 0.3 for total fat or 0.5 for calorie adjusted fat), or between repeat applications of the same instrument at different time points (see, for example, Willett *et al.*, 1985). Furthermore, there are reasons to strongly suspect that the assumptions underlying (3.1) and (3.2) do not hold for self-reported fat intake. The absence of an accepted biomarker of fat intake precludes direct study of the measurement errors in (3.1) and (3.2). However, doubly labeled water studies of short-term energy expenditure indicate energy consumption to be substantially under-reported, in the range 25–50%, among obese women with little or no under-reporting among slim women (Heitmann and Lissner, 1995), making it probable that fat calories are also greatly and differentially underreported by obese women at least in certain populations. Also, protein expenditure, as assessed by urinary nitrogen, appears to be underreported to a lesser extent than total energy (Heitmann and Lissner, 1995), making it plausible that percent energy from fat is also substantially underreported by obese women in their sudy population. These data suggest that energy or fat intakes from food records (or recalls) may not conform to the assumptions underlying (3.1), and hence may not serve as a suitable 'anchor' for calibrating the corresponding food frequency data. Another implication of these data is that measurement errors from food record (3.1) and food frequency (3.2) assessments of fat intake, or calorie-adjusted fat intake, may well be positively correlated. For example, an obese woman who underestimates her per capita fat consumption by, say, 50% on food records likely also systematically underestimates her fat consumption on a food frequency questionnaire, in which case even the modest correlations (e.g. 0.3) previously noted between food record and food frequency fat consumption may be due in whole or in part to correlated measurement error, rather than due to an ability of these dietary instruments to measure actual fat consumption. Prentice (1996) carried out some exercises under which the food record and food frequency estimates of total fat, or percent energy from fat, were allowed to have correlated measurement errors. The international data projected relative risks were very sensitive to such correlation in that even the modest trends shown in Table 2 could readily be further reduced or otherwise distorted. This type of exercise implies that it is necessary to have rather firm information on the measurement characteristics of the food frequency questionnaires in the context of the cohort or case-control studies in order to determine whether or not the relative risk trends shown in Table 2 differ appreciably from expectation under the generating hypothesis. In other words, one can take the point of view that cohort and case-control studies to date, are sufficiently limited that the hypothesis of breast cancer association with fat, or percent of energy from fat, has yet to be tested, let alone the hypothesis of fat reduction and breast cancer prevention. This state of affairs is a testimony to the real methodologic difficulties (Table 2) that pervade this research area.

3.3 *Methodologic research needs*

In view of the crucial role that measurement properties of the dietary assessment tools used in observational studies play in the interpretation of observed relative risk trends, and in the calculation of

measurement error-adjusted relative risk trends, and the likelihood that self-report assessments of nutrient intakes will fail to adhere, even approximately, to the measurement error assumptions that attend (3.1), it is natural and important to consider objective measures of nutrient intakes in (3.1) in future observational studies. For example, the American Association of Retired Persons cohort study being coordinated by the National Cancer Institute is incorporating a sizeable substudy that includes doubly labeled water assessments of energy expenditure and urinary nitrogen assessments of protein expenditure. Similarly, our group in Seattle is currently piloting a study of objective measures for a broader range of nutrients as a precursor to an intended larger objective measures substudy within the WHI. The concept behind these efforts is that the biological assessments of nutrient consumption would play a fundamental role in the analysis and reporting of association studies. Specifically, the available self-report measure X_2 in (3.2) would be replaced by an estimate of Z given X_2 using regression calibration equations. In that the measurement properties of the self-report data in applications to a specific population may depend on various study subject characteristics, possibly including body mass, ethnicity and age, it may be necessary for the objective measures substudy to be fairly large, perhaps on the order of 1000–2000 women, to be able to calibrate in a sufficiently precise and comprehensive fashion. The development of objective measures of the consumption of a broader range of nutrients, and the further study of the methods for the design and conduct of objective measure substudies is an important research goal toward strengthening the reliability and interpretability of analytic epidemiologic studies of diet and chronic diseases.

Students of epidemiology are typically taught a hierarchy of study reliability ranging from ecologic studies, case-control studies, cohort studies and intervention trials. In a circumstance of difficult to measure, highly correlated exposures and difficult to measure confounding factors, however, such a hierarchy may not always hold. In particular, well-conducted ecologic studies in which exposures and potential confounding factors are surveyed among moderate-sized random samples in each group of a multipopulation aggregate study are largely immune to the noise aspect of measurement error (Prentice and Sheppard, 1995; Sheppard and Prentice, 1995) and may be less sensitive to systematic aspects of measurement error as well, in part due to the ability to incorporate an unusually broad range of exposures. It is not clear how successfully 'between population' confounding can be controlled in such settings, but the challenges and uncertainties of observational studies in the diet and disease area argue for a varied research program, including research designs having differing sources of potential bias. Furthermore, methodologic work toward defining the comparative reliability of various observational approaches as a function of the distribution and measurement properties of key exposure and confounding variables should have a high priority in the chronic disease population research agenda.

Chronic disease prevention hypotheses having substantial public health implications should be put to the test in randomized controlled intervention trials whenever practical. In the case of dietary fat and breast cancer there are four pertinent intervention trials currently underway. None have reported breast cancer occurrence or recurrence data as yet. Two of these aim to prevent breast cancer recurrence or new primary breast cancers among women with early or intermediate stage breast cancer. The Women's Intervention Nutrition Study (Chlebowski *et al.*, 1993) targets a major reduction in dietary fat, while the Women's Healthy Eating and Lifestyle Study (Pierce *et al.*, 1995) emphasizes a plant-based diet high in vegetables and fruit. These studies are both in the course of enrolling about 2500–3000 women. The other two intervention trials focus on the reduction in breast cancer occurrence among women without a personal history of breast cancer. The Canadian trial of Boyd *et al.* (1997) is studying the impact of a low fat eating pattern on breast cancer incidence among about 9500 pre and postmenopausal women at elevated risk for breast cancer on the basis of mammographic dysplasia. The Womens Health Initiative (WHI Study Group, 1998), previously mentioned, is studying the effect of a low fat eating pattern on the occurrence of breast cancer and selected other diseases among 48 837 postmenopausal women in the United States.

Such intervention trials have a number of desirable features. First, by virtue of the randomized assignment to intervention or control status baseline risk factors, whether recognized or not, are

statistically independent of intervention assignment, eliminating the problem of confounding by pre-randomization factors. Second, the comparison of breast cancer rates between intervention and control women does not involve individual dietary assessment. Thirdly, intervention trials, by introducing a dietary change, have potential to directly identify practical preventive maneuvers. On the other hand intervention trials tend to be expensive and logistically difficult, and the maintenance of dietary adherence for a sufficient period of time to be able to ascertain clinical outcomes that may be affected beneficially or detrimentally with sufficient precision can be a formidable task. Furthermore, intervention trials by their very nature test a specific intervention or intervention program, so that the effects of specific elements of the intervention may be unclear. For example, the dietary intervention in the WHI Clinical Trial aims to induce a major reduction in percent energy from fat and total fat consumption among intervention women, but consistent with intervention goals such women also aim to increase the consumption of fruit and vegetables and grains. Analyses to relate these specific changes to any disease risk reductions may need to rely on individual self-reports of diet, very much like the association studies reviewed in the preceding section, unless objective measures of nutrient consumption are ascertained and appropriate measurement models are applied.

4. Lessons from studies of beta-carotene consumption and the prevention of lung cancer

4.1 *Summary of available studies*

Many observational studies have reported a lower risk of cancer among persons whose self-reported diet includes relatively large amounts of foods rich in beta-carotene and other carotenoids, and relatively high intakes of fruit and vegetables more generally. Review papers of these associations have noted that cohort and case-control study findings are particularly consistent (Table 3) for lung cancer (Ziegler *et al.*, 1996; Steinmetz and Potter, 1996).

It is particularly noteworthy that there have been at least ten prospective studies using blood levels of beta-carotene. These studies report a lower risk of lung cancer, heart disease, other cancer and all-cause mortality among persons having relatively high blood concentrations of beta-carotene. For example, Greenberg *et al.* (1996) report an estimated total mortality relative risk of 0.62, with a 95% confidence interval of (0.44, 0.87) for persons in the highest versus lowest quartile of blood beta-carotene, based on an average 8.2 year cohort follow-up period.

The consistency and magnitude of the beta-carotene relative risk trends stimulated three large lung cancer prevention intervention trials in the 1980s (Table 3). The ATBC Trial (Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group, 1994) randomized 29 133 male Finnish smokers to supplemental beta-carotene (20 mg d^{-1}) or placebo, and to alpha-tocopherol (vitamin E) versus placebo in a factorial design. After 5–8 years of follow-up, the estimated lung cancer relative risk (95% confidence interval) for beta-carotene supplementation was 1.18 (1.03, 1.36), based on 876 incident lung cancers during the planned follow-up period. There was no significant differences for other cancers, while the total mortality relative risk was 1.08 (1.01, 1.16).

This unexpected finding of an elevated lung cancer relative risk was followed by the results from the Carotene and Retinol Efficacy Trial (Omenn *et al.*, 1996). This US trial randomized 18 314 smokers, former smokers and asbestos-exposed persons to a combination of 30 mg d^{-1} beta-carotene plus 25 000 international units retinol per day versus placebo. This trial stopped about two years early after four years average follow-up and reported beta-carotene/retinol relative risks of 1.28 (1.04, 1.57) for lung cancer based on 388 incident cases, 1.26 (0.99, 1.61) for cardiovascular disease mortality, 1.17 (1.03, 1.33) for total mortality, with no significant differences for other cancers.

Table 3. *Lung cancer relative risk estimates in relation to beta-carotene consumption or supplementation, and related methodologic issues* (*average beta-carotene consumption in the United States is about* 3 mg d^{-1}*, considerably lower than the beta-carotene dosage provided by supplements, which themselves have differing bioavailability, in the intervention trials*)

Observational studies	
Carotenoid and/or fruit and vegetable intake associated with lower lung cancer risk in 8 of 8 prospective studies and 18 of 20 retrospective studies reviewed (Ziegler *et al.*, 1996).	
Randomized trials of beta-carotene supplementation	
Lung Cancer Relative Risk Estimate (95% confidence interval)	
ATBC Trial (Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group, 1994):	1.18 (1.03, 1.36)
Carotene and Retinol Efficacy Trial† (Omenn *et al.*, 1996):	1.28 (1.04, 1.57)
Physicians Health Study (Hennekens *et al.*, 1996):	0.93 (0.69, 1.26)
Methodologic issues and research needs	
• Reliability and interpretation of observational studies	
— beta-carotene consumption likely highly correlated with that for other micronutrients	
— joint measurement error modeling required for dietary measures of nutrient intakes and confounding factors	
— ability to ascertain and control for distinguishing characteristics of high fruit and vegetable consumers and beta-carotene supplement users	
• Need for quantitative methods to assess unexpected clinical trial results	
• Developmental process needed to justify a full-scale intervention trial	

†Intervention included both beta-carotene and retinol.

In comparison, the Physicians Health Study (Hennekens *et al.*, 1996) conducted among 22 071 male US physicians, aged 40–84, of whom only 11% were smokers and 39% were former smokers at baseline, were randomized to 50 mg of beta-carotene on alternate days in a factorial arrangement with aspirin supplementation. After a 12 year average follow-up period, there was no difference between the beta-carotene supplemental and placebo group in respect to the incidence of lung cancer, other cancers or cardiovascular disease or in total mortality. For example, the lung cancer comparison involved 82 incidence cases in the beta-carotene group versus 88 in the placebo group.

For completeness, the results of an esophageal and stomach cancer prevention trial in Linxian, China (Blot *et al.*, 1994), an area where the diet has been poor in several micronutrients, can be briefly mentioned. This trial reported a non-significant reduction in lung cancer mortality, with relative risk estimate of 0.55 (0.26, 1.14) among persons randomized to a combination of beta-carotene, alpha-tocopherol and selenium. This observation, in a cohort having a 30% smoking prevalence, was based on only 31 lung cancer deaths.

4.2 *Interpretation of available studies and needed research*

The lack of support from intervention trials for a protective effect of beta-carotene on lung cancer raises questions (Table 3) about the reliability and specificity of the observational studies leading up to these trials. As with the fat and breast cancer example, it is important to remember that the human diet is a complex mixture of foods and nutrients with many highly correlated elements. Furthermore, measurement error in dietary self-reports, and also in blood nutrient concentrations, may combine with

these high correlations to much reduce the reliability of observational study associations. Intuitively one would need precise information on the joint measurement error properties of beta-carotene and other nutrients in any attempt to distinguish a protective role for beta-carotene as compared to other carotenoids, or other correlated nutrients. The availability of an objective measure of nutrient consumption, (blood beta-carotene is affected by factors other than dietary intake) would represent an important step in an appropriate accommodation of measurement error, as was discussed in the preceding section. But such a marker would not be sufficient to handle confounding by other nutrients. Rather, objective markers adhering to (3.1) for both primary exposure and confounding factors in appropriate subsamples, would be needed to begin addressing the confounding issues. As mentioned previously, much work remains to be done on substudy design issues, and on the development of flexible data analysis methods. Confounding by non-dietary characteristics that distinguish frequent fruit and vegetable consumers and dietary supplement users from other persons is also an important issue for the interpretation of observational studies of micronutrients, such as beta-carotene, and disease risk. Some such characteristics, for example those related to physical activity patterns, may also be quite difficult to measure and accommodate in data analysis.

An explanation for the observed elevation of lung cancer among persons supplemented with beta-carotene in the ATBC trial, and an elevation in lung cancer among persons supplemented with beta-carotene and retinol in the CARET trial is still lacking. A large number of mechanistic possibilities have been proposed including the inhibition of absorption of other nutrients among persons consuming large amounts of beta-carotene, and the possibility of a prooxidant effect of large doses of beta carotene among persons having lungs damaged by cigarette smoking or asbestos exposure. Mayne (1996) provides an overview of related mechanistic research for lung cancer and other diseases. She also notes some marked differences in bioavailability of the beta-carotene preparations used in the various trials as a contributor to differences among trial results.

An additional issue raised by the beta-carotene and lung cancer experience concerns the methods for assessing unexpected clinical trial results. If the ATBC and CARET trials, with appropriate account of the sequential aspects of monitoring, had shown a reduced lung cancer risk then a very strong case could be made for the value of beta-carotene supplementation. One can ask, however, whether a stronger standard of proof should be required in order to declare harm, in view of the totality of evidence that motivated the hypothesis test. This topic is closely related to the establishment of asymmetric stopping boundaries in trial monitoring. The lack of consensus on this topic is evident from the very different viewpoints held by the lead investigators in the beta-carotene trials as to how the existing body of data is to be interpreted. A combined analysis of pertinent trial data may help resolve these differences, and is being planned. The beta-carotene supplementation trials, along with other prevention trials that have failed to provide evidence in support of primary hypotheses, also raise important questions concerning the developmental process needed to justify full-scale intervention trials. This topic will be discussed further below.

5. Conclusion: priority topics for methodology development

There have been a number of important recent advances in primary prevention research. For example, recent randomized controlled trials have demonstrated the ability of 'statin' family drugs to prevent coronary events among men and women having elevated cholesterol levels (Shepherd *et al.*, 1995); the ability of tamoxifen to reduce breast cancer occurrence among women at elevated risk for this disease (Fisher *et al.*, 1998); and the ability of alendronate to reduce fracture occurrence among women having a low bone density (Cummings *et al.*, 1998). These pharmaceutical approaches to disease prevention were each tested first for secondary prevention among persons who had already experienced the targeted diseases, prior to testing among ostensibly healthy persons. Chronic disease prevention by means of

nutritional interventions is also beginning to experience some success. These include randomized trial results of hip fracture prevention among elderly French women taking calcium and vitamin D supplements (Chapuy *et al.*, 1992), a reported reduction in myocardial infarction incidence among persons randomized to a high dose of vitamin E (Stephens *et al.*, 1996); and a recent report of reduced colon polyp recurrence, a colon cancer precursor, among persons randomized to calcium and vitamin D (Baron *et al.*, 1999). However, few behavioral interventions, arguably the ultimate product of a chronic disease population research enterprise, have been tested for their ability to prevent chronic diseases. The recently Polyp Prevention Trial Schatzkin *et al*., (2000), and the ongoing dietary modification component of the WHI Study Group (1998) provide exceptions.

The preceding examples allow the identification of some priority topics for methodology development to facilitate chronic disease population research. Methodologic work is needed to clarify and define the complementary and overlapping role that ecologic studies, analytic epidemiologic studies and randomized trials can fulfill in developing and testing hypotheses: for example, in the diet, physical activity, and chronic disease research area. In this area, as well as in the study of chemopreventive agents, additional research is needed to clarify the importance of such non-standard biases as adherence bias in observational studies. Innovative approaches, involving both necessary data and appropriate statistical models and methods, are needed for the accommodation of such biases. Such methods are especially needed for the accommodation of exposure measurement error in the study of difficult to measure exposures, such as those related to diet or physical activity. Equally important, there is a need for improved methods to communicate study reliability to both scientific and lay audiences. For example, in the diet and chronic disease area the public is burdened with such a cacophony of conflicting messages that no message, or even the wrong message, may turn out to be communicated. For example, in spite of the controversial and worrisome clinical trial data described in the preceding section, a health focus article in the December 6, 1999 of *Newsweek* prominently displayed the assertion that alpha and beta-carotene help prevent lung cancer!

There are also a number of methodologic research needs related to randomized controlled trials in the chronic disease population research area. These include the development of methods to effectively measure and monitor benefit versus risk when dealing with interventions plausibly affecting a range of organs and body systems. Closely related to this is the need for methods to combine clinical outcome data to meaningfully compare interventions. Perhaps the greatest need, however, is for a strengthened intervention development enterprise. In addition to various types of observational studies, such an enterprise is likely to rely heavily on small-scale human intervention trials with intermediate biochemical and genomic outcomes. For example, the rapid development of functional genomics (see, for example, Brown and Hartwell, 1998) using microarray technologies may add much specificity to the screening of preventive interventions: for example, in the context of small-scale human feeding trials, or other small scale trials. These functional genomic techniques allow one to observe the simultaneous effect of an intervention on the expression of a large number of human genes, or expressed sequence tags, throughout the human genome and eventually will be able to provide insight into a broad range of potential intervention effects. The statistical and bioinformatic aspects of analysing and interpreting such voluminous microarray data are at an early stage of development, but are widely recognized as critical to the impact of genomic techniques. A closely related aspect is the need for methods to combine high-dimensional response data in a manner that allows a meaningful screen of interventions based on such short-term response data. Genetic and genomic techniques also enter the chronic disease population research area in the identification of disease-related genes, which among other uses may be able to lead to smaller and more efficient prevention trials, and in the elucidation of gene environment interactions which are expected to fulfill a substantial role in chronic disease development. It is encouraging that scientists from a broad range of disciplines are becoming interested and involved in chronic disease population research, and disease prevention research in particular. Biostatisticians have an excellent opportunity to be very active participants in this exciting research enterprise.

ACKNOWLEDGEMENTS

This work was supported by Public Health Service grant CA53996 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

REFERENCES

ALBANES, D. (1987). Total calories, body weight, and tumor incidence in mice. *Cancer Research* **47**, 1987–1992.

ALPHA-TOCOPHEROL, BETA CAROTENE CANCER PREVENTION STUDY GROUP (1994). The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers among male smokers. *New England Journal of Medicine* **330**, 1029–1035.

ARMSTRONG, B. AND DOLL, R. (1975). Environmental factors and cancer incidence and mortality in different countries with special reference to dietary practices. *International Journal of Cancer* **15**, 617–631.

BARON, J. A., BEACH, M., MANDEL, J. S. *et al.* (1999). Calcium supplements for the prevention of colorectal adenomas. *New England Journal of Medicine* **340**, 101–107.

BLOT, W. J., LI, J. Y., TAYLOR, P. R. AND LI, B. (1994). Lung cancer and vitamin supplementation. *New England Journal of Medicine* **331**, 614.

BOYD, N. F., GREENBERG, C., LOCKWOOD, G. *et al.* (1997). Effects at two years of a low fat, high carbohydrate diet on radiological features of the breast: results from a randomized trial. *Journal of the National Cancer Institute* **89**, 488–496.

BROWN, P. O. AND HARTWELL, L. (1998). Genomics and human disease—variations on variation. *Nature Genetics* **18**, 91–93.

CARROLL, R. J., RUPPERT, D. AND STEFANSKI, L. A. (1995). *Measurement Error in Nonlinear Models*. New York: Chapman and Hall.

CHAPUY, M. C., ARLOT, M. E., DEBOEUF, F. *et al.* (1992). Vitamin D_3 and calcium to prevent hip fractures in elderly women. *New England Journal of Medicine* **327**, 1637–1642.

CHLEBOWSKI, R. T., BLACKBURN, G. L., BUZZARD, I. M. *et al.* (1993). Adherence to a dietary fat intake reduction program in postmenopausal women receiving therapy for early breast cancer. The Women's Intervention Nutrition Study. *Journal of Clinical Oncology* **11**, 2072–2080.

CUMMINGS, S. R., BLACK, D. M., THOMPSON, D. E. *et al.* (1998). Effect of alendronate on risk of fracture in women with low bone density but without vertebral fractures. *Journal of the American Medical Association* **280**, 2077–2082.

FISHER, B., CONSTANTINO, F. P., WICKERHAM, D. L. *et al.* (1998). Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 study. *Journal of the National Cancer Institute* **90**, 1371–1388.

FLEMING, T. R., PRENTICE, R. L., PEPE, M. AND GLIDDEN, D. (1994). Surrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and AIDS research. *Statistics in Medicine* **13**, 955–968.

FREEDMAN, L., CLIFFORD, C. AND MESSINA, M. (1990). Analysis of dietary, calories, body weight and the development of mammary tumors in rats and mice: a review. *Cancer Research* **50**, 5710–5719.

FREEDMAN, L., ANDERSON, G. L., KIPNIS, V., PRENTICE, R., WANG, C. Y., ROSSOUW, J., WITTES, J. AND DEMETS, D. (1996). Approaches to monitoring the results of long-term disease prevention trials. *Controlled Clinical Trials* **17**, 509–525.

FRIEDENREICH, C. M., HOWE, G. R. AND MILLER, A. B. (1991). The effect of recall bias on the association of calorie-providing nutrients and breast cancer. *Epidemiology* **2**, 424–429.

GIOVANNUCCI, E., STAMPFER, M. J., COLDITZ, G. A. *et al.* (1993). A comparison of prospective and retrospective assessments of diet in the study of breast. *American Journal of Epidemiology* **137**, 502–511.

GRADY, D., RUBIN, S. M., PETTITI, D. B. *et al.* (1992). Hormone therapy to prevent disease and prolong life in postmenopausal women. *Annals of Internal Medicine* **117**, 1016–1036.

GRAY, G. E., PIKE, M. C. AND HENDERSON, B. E. (1979). Breast cancer incidence and mortality rates in different countries in relation to known risk factors and dietary practices. *British Journal of Cancer* **39**, 1–7.

GREENBERG, E. R., BARON, J. A., KARAKAS, M. R. *et al.* (1996). Mortality associated with low plasma concentration of beta carotene and the effect of oral supplementation. *Journal of the American Medical Association* **275**, 699–703.

GREENWALD, P. (1999). Role of dietary fat in the causation of breast cancer: point. *Cancer Epidemiology, Biomarkers and Prevention* **8**, 3–7.

GROTSTEIN, F., STAMPFER, M. F., COLDITZ, G. A. *et al.* (1997). Postmenopausal therapy and mortality. *New England Journal of Medicine* **336**, 1769–1775.

HEITMANN, B. L. AND LISSNER, L. (1995). Dietary underreporting by obese individuals—is it specific or non-specific. *British Medical Journal* **311**, 986–989.

HENNEKENS, C. H., BURING, J. E., MANSON, J. E. *et al.* (1996). Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *New England Journal of Medicine* **334**, 1150–1155.

HOWE, G. R., HIROHATA, T., HISLOP, T. G. *et al.* (1990). Dietary factors and risk of breast cancer: combined analysis of 12 case-control studies. *Journal of the National Cancer Institute* **82**, 561–569.

HULLEY, S., GRADY, D., BUSH, T. *et al.* (1998). Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *Journal of the American Medical Assocation* **280**, 605–613.

HUNTER, D. J., SPIEGELMAN, D., ADAMI, H. O. *et al.* (1996). Cohort studies of fat intake and the risk of breastcancer—a pooled analysis. *New England Journal of Medicine* **334**, 356–361.

HUNTER, D. J. (1999). Role of dietary fat in the causation of breast cancer: counterpoint. *Cancer Epidemiology, Biomarkers and Prevention* **8**, 9–13.

INSULL, W., HENDERSON, M. M., PRENTICE, R. L. *et al.* (1990). Results of a randomized feasibility study of a low-fat diet. *Archives of Internal Medicine* **150**, 421–427.

MAYNE, S. T. (1996). Beta carotene, carotenoids, and disease prevention in humans. *FASEB Journal* **10**, 690–701.

OMENN, G. S., GOODMAN, G. E., THORNQUIST, M. D. *et al.* (1996). Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. *New England Journal of Medicine* **334**, 1150–1155.

PEPI TRIAL WRITING GROUP (1995). Effects of estrogen or estrogen/progestin regimens on heart disease risk factors in postmenopausal women: the Postmenopausal Estrogen/Progestin Interventions (PEPI) Trial. *Journal of the American Medical Association* **273**, 199–208.

PETTITI, D. B. (1998). Hormone replacement therapy and heart disease prevention. Experimentation trumps observation (editorial). *Journal of the American Medical Association* **280**, 650–651.

PIERCE, J. P., PARKER, B. AND WRIGHT, F. (1995). A feasibility study for a randomized controlled trial of a plant-based diet on breast cancer recurrence. *Proceedings of the American Society of Clinical Oncology* **14**, 160 (abstr).

PRENTICE, R. L. (1989). Surrogate endpoints in clinical trials: discussion, definition and operational criteria. *Statistics in Medicine* **8**, 431–440.

PRENTICE, R. L. AND SHEPPARD, L. (1990). Dietary fat and cancer: consistency of the epidemiologic data and disease prevention that may follow from a practical reduction in fat consumption. *Cancer Causes and Control* **1**, 81–97.

PRENTICE, R. L. AND SHEPPARD, L. (1995). Aggregate data studies of disease risk factors. *Biometrika* **82**, 113–125.

PRENTICE, R. L. (1996). Measurement error and results from analytic epidemiology: dietary fat and breast cancer. *Journal of the National Cancer Institute* **88**, 1738–1747.

SCHATZIN, A., LANZA, E., CORLE, D. *et al.* (2000). Lack of effect on a low-fat, high-fibre diet on the recurrence of coloreatal adenomas. *New England Journal of Medicine* **342**, 1149–1155.

SHEPHERD, J., COBBE, S. M., FORD, I. *et al.* (1995). Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. *New England Journal of Medicine* **333**, 1301–1307.

SHEPPARD, L. AND PRENTICE, R. L. (1995). On the reliability and precision of within- and between-population estimates of relative risk parameters. *Biometrics* **51**, 853–863.

STAMPFER, M. J. AND COLDITZ, G. A. (1991). Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. *Preventive Medicine* **20**, 47–63.

STEINMETZ, K. A. AND POTTER, J. D. (1996). Vegetables, fruit and cancer prevention: a review. *Journal of the American Dietetic Association* **96**, 1027–1039.

STEPHENS, N. G., PARSONS, A., SCHOFIELD, P. M., KELLY, F., CHEESEMAN, K. AND MICHINSON, M. J. (1996). Randomized controlled trial of vitamin E in patients with coronary disease: Cambridge Heart Antioxidant Study (CHAOS). *The Lancet* **347**, 781–786.

TANNENBAUM, A. (1942). Genesis and growth of tumors III. Effects of a high fat diet. *Cancer Research* **2**, 468–475.

UTIAN, W. H. (1998). HRT and HERS findings—has the ground shiften (editorial). *Menopause Management*, p. 5.

WHITEHEAD, M. AND STAMPFER, M. (1998). HERS—a missed opportunity (editorial). *Climactoric* **1**, 170–171.

WOMEN'S HEALTH INITIATIVE STUDY GROUP (1998). Design of the women's health initiative clinical trial and observational study. *Controlled Clinical Trials* **19**, 61–109.

WILLETT, W. C., SAMPSON, L., STAMPFER, M. J. *et al.* (1985). Reproducibility and validity of a semiquantitative food frequency questionnaire. *American Journal of Epidemiology* **122**, 51–65.

ZIEGLER, R. G., MAYNE, S. T. AND SWANSON, C. A. (1996). Nutrition and lung cancer. *Cancer Causes and Control* **7**, 157–177.

[*Received 27 December, 1999; revised 7 April, 2000; accepted for publication 12 April, 2000*]