

*Review***Observational Epidemiologic Studies of Nutrition and Cancer: The Next Generation (with Better Observation)**

Arthur Schatzkin,¹ Amy F. Subar,² Steven Moore,¹ Yikyung Park,¹ Nancy Potischman,² Frances E. Thompson,² Michael Leitzmann,¹ Albert Hollenbeck,⁴ Kerry Grace Morrissey,⁵ and Victor Kipnis³

¹Division of Cancer Epidemiology and Genetics, ²Division of Cancer Control and Population Sciences, ³Division of Cancer Prevention, National Cancer Institute, Bethesda, Maryland; ⁴AARP, Washington, District of Columbia; and ⁵Westat, Rockville, Maryland

Abstract

It would be of enormous public health importance if diet and physical activity, both modifiable behavioral factors, were causally related to cancer. Nevertheless, the nutritional epidemiology of cancer remains problematic, in part because of persistent concerns that standard questionnaires measure diet and physical activity with too much error. We present a new strategy for addressing this measurement error problem. First, as background, we note that food frequency and physical activity questionnaires require respondents to report “typical” diet or activity over the previous year or longer. Multiple 24-hour recalls (24HR), based on reporting only the previous day’s behavior, offer potential cognitive advantages over the questionnaires, and biomarker evidence suggests the 24-hour dietary recall is more accurate than the food frequency questionnaire. The expense involved in administering multiple 24HRs in large epidemiologic studies,

however, has up to now been prohibitive. In that context, we suggest that Internet-based 24HRs, for both diet and physical activity, represent a practical and cost-effective approach for incorporating multiple recalls in large epidemiologic studies. We discuss (1) recent efforts to develop such Internet-based instruments and their accompanying software support systems; (2) ongoing studies to evaluate the feasibility of using these new instruments in cohort studies; (3) additional investigations to gauge the accuracy of the Internet-based recalls vis-à-vis standard instruments and biomarkers; and (4) new statistical approaches for combining the new instruments with standard assessment tools and biomarkers. The incorporation of Internet-based 24HRs into large epidemiologic studies may help advance our understanding of the nutritional determinants of cancer. (*Cancer Epidemiol Biomarkers Prev* 2009;18(4):1026–32)

There are several reasons why nutrition—conceived broadly to include diet, body size, and physical activity—should be causally linked to cancer at one or more anatomic sites. Pathophysiologic mechanisms accounting for nutrition-cancer relations are abundant and plausible (1). Animal studies have provided convincing evidence that calorie restriction and specific dietary interventions can modulate tumorigenesis (2, 3). Ecologic data—international cancer rate variation and related food consumption correlations (4), time trends (5), and migration patterns (6)—are consistent with nutritional determinants of malignant disease. Many epidemiologic studies with information on individual participants have associated various nutritional factors with cancer (1). Especially noteworthy is the accumulating evidence that adiposity over the life span is causally related to several malignancies (7).

Nevertheless, the nutrition and cancer field has been problematic, particularly with respect to the etiologic roles of dietary factors and physical activity. For a number of important foods, macronutrients, and micronutrients—total dietary fat and fat subtypes, dietary fiber and whole grains, fruits and vegetables, folate, lycopene, selenium, calcium, and vitamin D, to name just a few—the causal links to cancers remain unclear. Consistent, though rather modest, associations have been reported between physical activity and cancers of the colorectum and, to a lesser degree, breast; but for other cancer sites the evidence is ambiguous (7).

The inconsistency and uncertainty in the nutritional epidemiology of cancer can be interpreted in two ways: (1) important, public health-relevant, causal links for cancer are few, and many of the long-standing hypotheses are simply wrong; (2) many of these long-standing hypotheses are right, but methodologic difficulties have prevented us from generating the requisite evidence. The first interpretation is really one of exclusion: as long as methodologic problems prevent us from seeing the truth we cannot rule out that truth. This article focuses on these methodologic problems—one, in particular.

Received 11/26/08; revised 1/26/09; accepted 2/10/09; published OnlineFirst 3/31/09.

Requests for reprints: Arthur Schatzkin, 6120 Executive Boulevard, Suite 320, Rockville, MD 20886. Phone: 301-594-2931; Fax: 301-496-6829. E-mail: Schatza@mail.nih.gov

Copyright © 2009 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-08-1129

Methodologic Problems in Observational Epidemiologic Studies of Nutrition and Cancer

The earlier generation of case-control studies of nutrition and cancer were faced with potential recall, selection, and reverse-causation biases. These methodologic obstacles are essentially moot now that a newer generation of prospective cohort studies around the world has addressed the nutrition-cancer problem. Prospective cohort studies, however, still face two serious methodologic problems: (1) confounding by demographic, behavioral, and biologic characteristics associated with both nutritional exposures and cancer outcomes; (2) measurement error, the principal focus of this article.

Confounding. For a true causal relation between a nutritional exposure and cancer, the relative risk may well be only 1.5 to 2.0 (or even less). Confounding by one or more unknown or inadequately measured exposures can account for such a modest association. Confounding is a serious issue in nutritional epidemiology but is considered only briefly here. Because standard multivariate statistical techniques may not fully alleviate the problem, several investigators have proposed that sensitivity analyses be conducted to evaluate whether estimated relative risks are robust for potentially unknown confounders (8, 9). A major alternative design strategy for dealing with confounding in nutritional epidemiology is the randomized controlled trial: the randomization process renders the treatment groups similar with respect to both known and unknown confounders. More recently, Mendelian randomization—use of a genetic variant as a proxy for a nutritional exposure—has been proposed as a strategy for addressing confounding (10).

Measurement Error. Error in the measurement of exposures can bias epidemiologic findings. When only the main exposure is measured with error, that error generally leads to attenuation of associations between that exposure and a disease outcome and a loss of statistical power to detect exposure-disease associations (11). Despite the loss of power, a test of the association remains statistically valid. Things get more complicated in multivariate contexts. If, in addition to the main exposure, covariates (say, other dietary factors, including energy) are also measured with error, the association between the main exposure and disease may be attenuated or, in certain circumstances, even inflated (12). Moreover, statistical tests become invalid without a proper measurement error adjustment. When we are dealing with modest associations between nutritional exposures and malignant disease, error in the measurement of multiple nutritional factors can be a serious methodologic obstacle to observing and testing causal relations.

The literature on measurement error in dietary assessment is extensive and growing (13, 14). The instrument typically used in epidemiologic studies is the food frequency questionnaire (FFQ), which elicits information on frequency of consumption and usual portion size over the previous year for some 100 to 150 foods. It has long been acknowledged that the FFQ assesses dietary intake with error, both systematic and random (12-16). The argument in support of use of the

FFQ is that, despite such error, study participants are reasonably well ranked with respect to dietary consumption—especially after energy adjustment, which purportedly reduces error—such that meaningful comparisons across intake quantiles can be made.

Over the last decade, investigators have used biomarkers to evaluate dietary assessment instruments. Kaaks and colleagues made an important distinction between “recovery” and “concentration” biomarkers (17). Recovery biomarkers are based on precise and quantitative knowledge of the physiologic balance between intake (dietary protein, for example) and output (urinary nitrogen over 24 hours). In contrast, concentration biomarkers, including many blood nutrient levels, are based on the measurement of a specific compound’s concentration at a given point in time. One characteristic of concentration markers is that the quantitative relation between dietary intake and marker can vary substantially among individuals. This interindividual variation results from personal lifestyle characteristics (such as smoking and other dietary constituents) as well as physiologic factors (such as absorptive and metabolic function) that influence the “translation” of dietary intake into a biomarker value. Because recovery biomarkers are generally not subject to the influence of these personal characteristics, which may be correlated with true intake, they are especially useful for evaluating the measurement error structure of assessment instruments.

Studies using recovery biomarkers such as doubly labeled water (DLW, a measure of total energy expenditure and, in the context of energy balance, total energy intake) and urinary nitrogen (UN, a measure of protein intake), however, have suggested that measurement error with the FFQ can be large. For example, the OPEN (Observing Protein Energy Nutrition) study showed that, for absolute energy and protein intakes, relative risks of 2.0 could be attenuated down to as little as 1.05 with the FFQ (18). OPEN also showed that energy adjustment could help the situation, such that, for energy-adjusted protein (protein density) a true relative risk of 2.0 would be attenuated to 1.25; an improvement, certainly, but still demonstrating substantial FFQ error and subsequent relative risk attenuation. (A true energy-adjusted relative risk of, say, 1.4, would therefore be attenuated to 1.1). This undermines the “ranking” argument: with this much measurement error, even after energy adjustment, some individuals with true high intake will be ranked as low, and vice versa; moreover, there will be considerable misclassification across proximate categories (a person who truly belongs in quintile 4 is categorized in quintile 3 or 2), which further attenuates the observed relative risk. Recent reports from the Women’s Health Initiative (WHI), also based on use of recovery biomarkers, confirmed substantial reporting error associated with use of an FFQ (13).

These studies are limited in that the number of recovery biomarkers with which to compare FFQ intake is rather small—there are no recovery biomarkers for other nutrients of interest such as dietary fat, or fiber, or fruits and vegetables—though for nutritional factors such as total dietary fat or lean meat that are strongly correlated, respectively, with energy and protein intake, the DLW and UN data may give us some indication of the extent of error produced by the FFQ. In light of these recent biomarker study findings, and with a backdrop of

null findings and inconsistent results for many diet-and-cancer hypotheses, a lively debate on the ultimate utility of the FFQ, and whether it should be abandoned for newer instruments, has arisen (16, 19, 20).

A heightened appreciation for the error involved in assessing physical activity has evolved in recent years. Physical activity has generally been assessed in epidemiologic studies via frequency-type questionnaires. Several different physical activity questionnaires have been used, but in general they inquire about typical activity over time with some attempt to differentiate between different levels of activity (sedentary versus light versus moderate versus vigorous). As with the FFQ, the cognitive focus is on an extended period of time ("typical" activity, or activity over the previous year, etc.) and requires complex mental calculations to come up with the amount of typical activity. Investigators are now using more objective measures of physical activity, including accelerometers, heart rate monitors, or combinations. Recent reports have shown that the amount of physical activity measured by accelerometer is substantially lower than that determined by self-report (21, 22). With regard to cancer, the consistent findings of protection for physical activity versus colorectal cancer might in fact reflect underestimates of the true protection conferred by physical activity; whereas the inconsistent findings for other sites may, in the context of substantial physical activity assessment error, be obscuring true associations, perhaps less strong than those for colorectal cancer, but important nonetheless.

New Generation of Studies, New Instruments

An underlying theme emerges here: we have stronger studies—prospective cohorts set up around the world—but assessment instruments that have substantial measurement error. Moreover, we do not have accurate reference measurements (with the exception of a few recovery biomarkers like DLW and UN for diet) to adjust for measurement error in FFQs or physical activity questionnaires. In this situation, the question immediately arises: can we develop, and incorporate in prospective studies, more accurate dietary and physical activity assessment instruments. And the corollaries to that question are, what approaches, in terms of study design and statistical techniques, can we take to evaluate such new instruments, combine them with other assessment tools (if and when appropriate), and adjust for measurement error.

Diet. Alternatives to the FFQ do exist, including dietary records (also called "diaries") and 24-hour dietary recalls (23).

Dietary Records. Study participants completing dietary records record, in real time, the types and amounts of foods consumed for one or more days. Typically these records have been on paper, but work is ongoing to determine whether various electronic devices, including cell phones, can be used for recording dietary data. Often some sort of tutorial, written or video, is provided to the participant to assist in filling out the records.

24-Hour Dietary Recalls. 24-Hour dietary recalls (24HR) are generally interviewer-administered, often without participants knowing in advance when they will be

contacted. The interviewer, in person or by phone, asks about all foods eaten and beverages drunk over the previous day. The U.S. Department of Agriculture has adopted the Automated Multiple Pass Method (AMPM) for administering the 24HR in the U.S. National Health and Nutrition Survey (NHANES) (24); five distinct "passes" are used to assist the respondent in remembering what was consumed throughout the day. Given individual day-to-day variation in intake, multiple 24-hour recalls are required to estimate a study participant's usual intake.

Records and recalls, which require an individual to either record intake over several days or remember what was eaten on the previous day, have potential cognitive advantages over the FFQ, which asks respondents to provide "typical" frequencies and portion sizes generally over the previous year and requires mental averaging over varying intakes and seasons. Like FFQs, however, both records and recalls also measure intake with systematic and random error. In fact, as OPEN has suggested, the likely correlation of errors in FFQs and recalls (or records) means that calibration substudies using recalls (or records) as "reference" instruments will underestimate the extent of error in the FFQ (18). Recent reports, based on a comparison of FFQs to records in epidemiologic studies of dietary fat and breast cancer, are compatible with the notion that records measure diet with less error than FFQs (25, 26).

For dietary records, the most common method currently used is a printed record given to the respondent to be filled in over a specified period of time. This can be done relatively inexpensively in a large cohort. Research is underway on the possibility of using cell phone digital camera technology, voice recognition, or other complementary electronic devices to collect real time dietary intake data that can be completely or partially coded for analysis of daily food and nutrient intakes. These techniques are being evaluated for accuracy; further development would also be necessary to show their feasibility for large-scale research. The coding of either paper and pencil or electronic real time diet record information, can be prohibitively expensive, \$100 or more per 4-day food record per participant. Such costs can be substantially defrayed if the coding is done on a nested case-control basis: for example, records could be coded for all breast cancer cases and an appropriately selected number of controls.

A more serious problem with records, however, is their propensity to influence behavior; that is, they are "reactive." Because participants know in advance that they will be filling out the diary over a given series of days, even if reporting is accurate, the participant may alter dietary habits for the duration of the diary completion period. Use of records is particularly prone to underreporting and therefore an underreporting of usual diet. In fact, diaries are often used for weight loss programs in which the intent is not to assess intake but to change it. The extent of this reactivity may vary from population to population, but little is known about such variation; although there is some evidence that reactivity would be nontrivial in U.S. populations (27), whether it is an issue in, say, the United Kingdom is not clear.

Dietary recalls are among the most trusted of dietary assessment methods and used worldwide in dietary surveillance. Because it measures past diet, it is not

reactive. The task of remembering the past 24 hours may be easier than forming the complex judgments needed for reporting on the past year in an FFQ. Respondents may have difficulty remembering foods consumed the previous day and therefore may overtly omit or add foods or report inaccurate portion sizes. Nevertheless, OPEN showed considerably less measurement error in the 24HR than the FFQ (28).

A major problem with the use of the 24HR in large epidemiologic studies, as the individual-level assessment tool, not as merely a reference tool on a small subset of participants, is the expense incurred in administering multiple recalls to a large cohort. Unlike the record, in which coding can be done ad hoc in a selected manner, the costs of the 24HR are in the initial administration of the instrument. If, for example, the cost of a single interviewer-administered computer-assisted 24HR is about \$100, and six recalls are to be administered to each of 500,000 participants in a cohort, the cost would exceed a quarter of a billion dollars just to do the dietary assessment! The administration costs may be lower in many countries relative to those in the United States, but the costs would still be prohibitive in many cohorts around the world.

Rationale for Development of Internet-Based Assessment Instruments (ASA24 and ACT24)

More and more people around the world are gaining access to and using the Internet. Internet usage has increased to almost 73% in the United States by June 2008, a growth of almost 130% from 2000 to 2008 (29). According to Nielsen Online, more than 90% of Internet users have broadband connections, high-speed Internet connections that allow faster data transfer needed for the new Internet-based assessment instruments. More than 96% of U.S. workers have broadband connections available in the workplace (30). The "baby boomer" generation of some 81 million men and women age 43 to 63 has 60 million online consumers; more than half of persons age 64 and older in the U.S. are online at least monthly and 33% have broadband (up from 28% a year earlier)" (31). This increasing Internet use provides an attractive option for nutritional assessment on a large scale in prospective cohorts.

Diet. Over the past four years, the National Cancer Institute (NCI) has been developing a new Internet-based 24-hour dietary recall, called ASA24 (Automated Self-Administered 24-HR) (32). ASA24 is based on the U.S. Department of Agriculture's (USDA) AMPM 24HR approach. The AMPM, in its series of "passes," begins with a quicklist of foods recalled from the previous 24-hour period and then successively probes the respondent for greater detail and specificity on such elements as added foods, portion size, and amount consumed. Adaptation of the AMPM was necessary given that ASA24 takes place in a self-administered computer-based environment.

ASA24 uses state-of-the-art automated computer technology, including a tutorial, graphic enhancements, animated characters to guide participants, and audio language/cues. The instrument is meal-based. Respondents select foods consumed the previous day from USDA's Food and Nutrient Database for Dietary Studies (FNDDS) (33) database, which includes approximately 8,000 foods. Respondents find foods by browsing

through food groups or by typing and searching. Information is collected about eating occasion, time of consumption, and details such as preparation methods and additions to food. To estimate amount consumed, respondents are presented with up to eight sequentially sized digital pictures. Pictures of beverage containers, a coffee mug or bottle of milk, for example, include a movable "slider" allowing the respondent to indicate what percent of beverage in the container was actually consumed. Respondents are given multiple opportunities to modify or edit their food lists.

For example, a respondent who ate Kellogg's All Bran cereal at breakfast with skim milk can report this either by typing the food into a search box and entering or by clicking the "cereal" food group that opens subgroups in which "ready-to-eat cereals" is selected, followed by selection of the specific food term, "Kellogg's All Bran." The program also contains up to eight photographs of foods to assist in portion size designations; for example, for All Bran cereal, eight photos are given, ranging from 113 g to more than 454 g. Respondents have the option of reporting more than 454 g by using a spin dial that increases from 454 g up to 5,670 g in 226-g increments. Modules that researchers can select as options for their specific research include location, with whom one ate, whether the TV was on for each meal, salt use, and dietary supplement use.

The software has the capacity to immediately compute nutrient and food group estimates for each recall day. Optional modules allow collection of data on dietary supplements, where meals are consumed, and where food is obtained. A module targeting meat preparation and consumption is also being added. After the first version is produced, NCI will proceed to translate ASA24 into Spanish.

Considerable qualitative testing, including use of focus groups, has gone into the design of ASA24. It is estimated that study participants will take about 30 to 40 minutes to complete ASA24 the first time. Use of the instrument seems to become easier and faster with successive completions. More information on ASA24 can be found at <http://riskfactor.cancer.gov/tools/instruments/asa24.html>.

French researchers are also developing a comprehensive 24HR program, called Nutrinet Santé, which has much in common with ASA24; this program uses both successive probes and photographs but is, of course, oriented toward the French rather than U.S. cuisine.⁶ Investigators in other countries are also exploring development and use of automated recall; efforts are now underway, for example, to adapt ASA24 to the U.K. context.

Physical Activity. During the past 40 years, a multitude of physical activity questionnaires have been developed for use in epidemiologic studies (34). Despite these extensive development efforts, physical activity questionnaires, when compared to reference measures such as DLW, have not generally been found to measure physical activity with a high degree of accuracy (35, 36). This inaccuracy in self-reported physical activity likely reflects the cognitive challenges embedded in responding

⁶ S. Herceberg, personal communication.

to such questionnaires. For example, many physical activity questionnaires ask study participants to report their usual daily activity during the past year, as well as the average number of hours spent per day in sedentary, light, moderate, or vigorous activities, a task requiring complicated mental averaging across weekends, weekdays, and seasons, as well as varying activity types. In addition, few questionnaires ask for time spent according to specific activities, a practice that would likely improve accuracy of recall.

Unlike the situation for diet, investigators have paid scant attention to the possibility of using the 24-hour recall approach for physical activity (37, 38). As with the dietary recall, a physical activity recall has the potential cognitive virtues of a respondent's having to remember only the previous day's activities, not having to average over a more extended period of time. If a sufficient number of such physical activity recalls are completed to account for variability in activity (especially on weekdays as opposed to weekends), the multiple recall approach may provide more accurate data than current questionnaires.

NCI has been developing such a physical activity recall, ACT24. This instrument asks a study participant to go through the previous 24-hour period and list all activities for the entire day. Visually, the screen is divided into two portions. The major part of the screen displays the list of common activities, such as sleeping, eating, bathing, commuting, walking, working, exercising, TV watching; participants select from this list. The minor part of the screen shows a calendar, similar in appearance to an appointment calendar, where activities appear after each entry is completed.

Participants are initially prompted to enter the activity that he/she was doing at midnight on the day of interest. For each activity, participants are asked to enter a start time (the option of "before midnight" is also included) and end time, rounded to the nearest 15 minutes. After entering the activity, ACT24 probes the participants in greater detail about the activity. For example, the probes for time spent at work include queries about time spent sitting, standing, and walking at work, as well as time spent carrying weight and the amount of weight carried. When the respondent completes the entry for an activity, the corresponding time period is blocked off in the calendar portion of the screen. This provides the participant with a way to track entries and helps set activities in the context of that particular day, potentially facilitating recall of further activities. Participants can also edit previously entered activities through the calendar directly, by dragging and dropping or by manually entering a new start or end time. After completion of an activity entry, the participant is then prompted to enter the activity for the next time slot. Participants can then enter an activity for that time or alternatively for a different time by selecting a start time of their choosing, thus giving some flexibility in the way that entries are made. Qualitative testing suggests that ACT24 will require 20 to 30 minutes for completion the first time.

Research Management Systems

Whereas all field studies require support systems for administration and processing of data, a study using an Internet-based delivery system requires additional fea-

tures. For explicitly incorporating ASA24 and ACT24 into a large cohort study such as AARP, a Website Study Management System (WSMS) has been developed. Additional Internet-based instruments can be included for WSMS administration, and assignment and completion rules can be specially tailored for a given cohort.

From the perspective of participant interface functions, WSMS (1) guides a participant through the consent process, captures specified demographic and contact information, and allows the participant to self-register by creating a personal user account; (2) allows a respondent to securely and seamlessly access the ASA24 and ACT24 stand-alone applications based on protocol-specified rules; (3) prompts the participant on a periodic basis, via email, to complete ASA24 or ACT24 for the previous 24-hour period; (4) links with an outbound calling system to generate reminder phone calls and text messages to cell phone if the participants opt-in for this feature. The WSMS can, for example, prompt a participant to alternate completing ASA24 and ACT24 every month over a one-year period, resulting in six completions of each instrument during that year. WSMS will mix the assignment of weekend days and weekdays in the appropriate proportions. The WSMS can be configured on the basis of protocol rules so a participant can exit the instrument and return at a later time, with the proviso that completion take place within the calendar day following the recalled day; those going past this 24-hour period will be rescheduled according to protocol rules and asked to complete the recall at another time.

WSMS also has specific research administration functions: it (1) receives and manages the ASA24 and ACT24 tracking data, such as completion status; and (2) provides response rates and summary statistics.

Incorporating the New Assessment Tools in Epidemiologic Studies

NCI investigators are initiating a study to evaluate the feasibility of using ASA24 and ACT24, with the WSMS, in the existing National Institutes of Health (NIH)-AARP Diet and Health Study cohort of more than half a million men and women and among other AARP members (39). The aims of this feasibility study are to (1) evaluate the response rate for the new programs; (2) determine the distribution of dietary intakes and physical activity among respondents (establishing that the ranges of dietary consumption and physical activity allow for meaningful comparisons, making sure, for example, that respondents are not all "vegetarian marathoners"); and (3) identify technical problems arising with ASA24 and ACT24 as well as the WSMS. The feasibility study will also include online versions of standard questionnaires. The feasibility of using the new programs and management system in other existing cohorts and potential study populations is also being explored.

Statistical Approaches for Combining New and Standard Self-Report Instruments

Unlike other studies (40, 41) that have used 24HRs or dietary records as "reference" instruments to evaluate and adjust for measurement error, now the former

reference instruments, the automated Internet-based recalls, will provide the individual-level data. If ASA24 or other new Internet-based dietary recalls serve as the primary individual-level data sources in epidemiologic studies, then the information on nutrients, foods, food groups, and even dietary patterns and indexes (42) derived from the recall data can be analyzed in direct relation to cancer end points.

In addition, the new digital environment will allow us to consider designs that include additional types of (web-based) dietary instruments and explore analytic strategies that combine data from two or more instruments. A FFQ could give frequency information for infrequently consumed foods of importance. Our research indicates that additional information on frequency of consumption is especially helpful in allowing more precise estimates of intake of episodically consumed foods (fish, for example). One recent approach takes the FFQ data, along with other demographic and lifestyle characteristics, as covariates in calibrating a 24HR-reported food group and relating it to health end points (43).

Similar arguments can be made for analysis of physical activity data in relation to cancer and other end points. First, individual-level data from ACT24 can be related in multivariable regression to the end points. Then, given that standard physical activity questionnaires can be administered online along with ACT24, it should be possible to combine the ACT24 data with other physical activity data in ways that parallel the approaches taken for diet.

Evaluation Studies

We have argued that the use of the new Internet-based tools will give us more accurate data on diet and physical activity, and that we may get closer still to the truth with various combination approaches. We need hard data, however, to back up this assertion. In particular, we need studies that show how the new instruments compare with standard ones as well as objective reference biomarkers. In addition, such studies may allow us to adjust for measurement error within epidemiologic studies; but again, we need to acknowledge the current limitations in the availability of dietary recovery biomarkers and objective measures of physical activity.

NCI investigators are planning a study that compares mean nutrient and food group values from ASA24 and a standardized interviewer-administered 24HR. Twenty-five percent will complete ASA24 twice, 50% will complete ASA24 once and AMPM once (with the order randomized), and 25% will complete the AMPM twice; this procedure provides information on both within- and between-person comparisons. In addition, a smaller study is planned ($n = 60$) in which NCI investigators would unobtrusively observe and record food intakes for participants who would be randomly assigned to complete one of two types of 24-hour recalls, ASA24 or the standard interviewer-administered recall. The investigators will then compare accuracy of the two methods.

The feasibility study described above will permit comparisons of ASA24 to a standard FFQ (44), ACT24 to a standard physical activity questionnaire, and may also include a comparison of ACT24 with accelerometer data.

With the incorporation of the new web-based instruments into the AARP population, we hope to integrate a

biomarker-based calibration study (OPEN2), using DLW, UN, possibly K+, and urinary sugars, analogous to the free-standing OPEN study; physical activity monitors could also be incorporated. Such a study would have two primary objectives. First, it would allow investigators to evaluate the accuracy of the new instruments in relation to objective biomarkers: energy from ASA24 versus DLW, protein from ASA24 versus UN, energy density from ASA24 versus DLW and UN, physical activity from ACT24 versus DLW, accelerometer. Second, it would provide quantitative data on the measurement error structure of the new instruments. Such information could be used to adjust observed relative risks for measurement error. This adjustment, however, could be strictly carried out only for those diet-disease associations that involve the factors reflected in the handful of available recovery markers, namely protein and energy intake, and possibly intakes of potassium and certain sugars. It is possible, though, to carry out sensitivity analyses by assuming a similar or somewhat different measurement error structure for other nutrients or foods of interest and thereby make some estimates of cancer risk for such dietary factors. Moreover, because participants in this study will complete one or more FFQs in addition to the multiple recalls, the measurement error structure for the FFQ can also be determined in this study population. Therefore, it may be possible to evaluate the measurement error structure for the combined tools, that is, multiple recalls plus FFQ, and use these data to adjust observed relative risks for measurement error derived from both instruments combined. This is an area of ongoing research, with the potential for extension to physical activity assessment as well.

Conclusions—On to the Next Generation

The incorporation into prospective cohort studies of new web-based assessment tools, 24-hour recalls for both diet and physical activity administered multiple times over the course of, say, a year, promises a new generation of epidemiologic studies. This new generation of studies can be diverse and international, with the appropriate adaptation of the web-based tools to various regional and ethnic cuisines and cultures. Internet access will only grow around the world and this growth will occur in most sociodemographic groups. Moreover, the digital nature of the new web-based studies allows for far greater flexibility in assessment tool modification over time, at remarkably little expense compared to that required in updating hard-copy questionnaires.

We may find that relative risks for some nutrition-cancer hypotheses are indeed strengthened or even revealed for the first time (25, 26). This finding would clearly be of great public health importance. For some hypotheses, however, we might find that the multiple recall-based relative risks are comparable to those derived from FFQs and physical activity questionnaires (PAQs). This result is likely only to strengthen the credibility of the initial FFQ- and PAQ-based findings. Either way, the new generation of studies will advance our understanding of nutrition and cancer causation.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

- World Cancer Research Fund. Diet, cancer, physical activity, and cancer. Washington (DC): American Institute for Cancer Research; 2007.
- Tannenbaum A. The genesis and growth of tumors. III. Effects of a high-fat diet. *Cancer Res* 1942;2:468–75.
- Bjorkhem-Bergman L, Torndal UB, Eken S, et al. Selenium prevents tumor development in a rat model for chemical carcinogenesis. *Carcinogenesis* 2005;26:125–31. PubMed doi:10.1093/carcin/bgh290.
- Carroll KK, Khor HT. Dietary fat in relation to tumorigenesis. *Prog Biochem Pharmacol* 1975;10:308–53. PubMed.
- You WC, Jin F, Devesa S, et al. Rapid increase in colorectal cancer rates in urban Shanghai, 1972–97, in relation to dietary changes. *J Cancer Epidemiol Prev* 2002;7:143–6. PubMed.
- Kolonel LN, Wilkens LR. Migrant studies. In: Schottenfeld D, Fraumeni JF, editors. *Cancer epidemiology and prevention*. New York: Oxford University Press; 2006. p. 189–201.
- Working IARC. Group on the Evaluation of Cancer-Preventive Strategies. Weight control and physical activity. Lyon (France): IARC; 2001.
- Greenland S. The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. *J Am Stat Assoc* 2003;98:47–54. doi:10.1198/01621450338861905.
- McCandless LC, Gustafson P, Levy AR. A sensitivity analysis using information about measured confounders yielded improved uncertainty assessments for unmeasured confounding. *J Clin Epidemiol* 2008;61:247–55. PubMed doi:10.1016/j.jclinepi.2007.05.006.
- Schatzkin A, Abnet C, Cross A, et al. Mendelian randomization: how it can and cannot help confirm causal relations between nutrition and cancer. *Cancer Prev Res* 2009;2:104–13.
- Freudenheim JL, Marshall JR. The problem of profound mismeasurement and the power of epidemiological studies of diet and cancer. *Nutr Cancer* 1988;11:243–50. PubMed.
- Schatzkin A, Kipnis V. Could exposure assessment problems give us wrong answers to nutrition and cancer questions? *J Natl Cancer Inst* 2004;96:1564–5. PubMed.
- Neuhouser ML, Tinker L, Shaw PA, et al. Use of recovery biomarkers to calibrate nutrient consumption self-reports in the Women's Health Initiative. *Am J Epidemiol* 2008;167:1247–59. PubMed doi:10.1093/aje/kwn026.
- Rosner B, Michels KB, Chen YH, Day NE. Measurement error correction for nutritional exposures with correlated measurement error: use of the method of triads in a longitudinal setting. *Stat Med* 2008;27:3466–89. PubMed doi:10.1002/sim.3238.
- Willett WC. *Nutritional epidemiology*. 2nd ed. New York: Oxford University Press; 1998.
- Kristal AR, Peters U, Potter JD. Is it time to abandon the food frequency questionnaire. *Cancer Epidemiol Biomarkers Prev* 2005;14:2826–8. PubMed doi:10.1158/1055-9965.EPI-12-ED1.
- Kaaks R, Ferrari P, Ciampi A, Plummer M, Riboli E. Uses and limitations of statistical accounting for random error correlations, in the validation of dietary questionnaire assessments. *Public Health Nutr* 2002;5:969–76. PubMed doi:10.1079/PHN2002380.
- Kipnis V, Subar AF, Midthune D, et al. The structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol* 2003;158:14–21. PubMed doi:10.1093/aje/kwg091.
- Willett WC, Hu FB. Not the time to abandon the food frequency questionnaire: point. *Cancer Epidemiol Biomarkers Prev* 2006;15:1757–8. PubMed doi:10.1158/1055-9965.EPI-06-0388.
- Freedman LS, Schatzkin A, Thiebaut AC, et al. Abandon neither the food frequency questionnaire nor the dietary fat-breast cancer hypothesis. *Cancer Epidemiol Biomarkers Prev* 2007;16:1321–2. PubMed doi:10.1158/1055-9965.EPI-07-0179.
- Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc* 2008;40:181–8. PubMed.
- Hagströmer M, Oja P, Sjostrom M. Physical activity and inactivity in an adult population assessed by accelerometry. *Med Sci Sports Exerc* 2007;39:1502–8. PubMed doi:10.1249/mss.0b013e3180a76de5.
- Thompson F, Subar A. Dietary assessment methodology. In: Coulston Am, Boushey CJ, editors. *Nutrition in the prevention and treatment of disease*. Burlington (MA): Elsevier Academic Press; 2008. p. 3–39.
- Raper N, Perloff B, Ingwersen L, Steinfeldt L, Anand J. An overview of USDA's dietary intake system. *J Food Compos Anal* 2004;17:545–55. doi:10.1016/j.jfca.2004.02.013.
- Bingham SA, Luben R, Welch A, Wareham N, Khaw KT, Day N. Are imprecise methods obscuring a relation between fat and breast cancer? *Lancet* 2003;362:212–4. PubMed doi:10.1016/S0140-6736(03)13913-X.
- Freedman LS, Potischman N, Kipnis V, et al. A comparison of two dietary instruments for evaluating the fat-breast cancer relationship. *Int J Epidemiol* 2006;35:1011–21. PubMed doi:10.1093/ije/dyl085.
- Rebro SM, Patterson RE, Kristal AR, Cheney CK. The effect of keeping food records on eating patterns. *J Am Diet Assoc* 1998;98:1163–5. PubMed doi:10.1016/S0002-8223(98)00269-7.
- Schatzkin A, Kipnis V, Carroll RJ, et al. A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study: results from the biomarker-based OPEN study. *Int J Epidemiol* 2003;32:1054–62. PubMed doi:10.1093/ije/dyg264.
- Internetworldstats.com. Copyright. Miniwatts Marketing Group; 2008. Available from: <http://internetworldstats.com/>.
- Nielsen Online. The Nielsen Company; July 2008. Available from: <http://www.nielsen-online.com/>.
- Golvin CS, Schadler T. The state of consumers and technology: benchmark 2008. Cambridge (MA): Forrester Research, Inc.; 2008.
- National Cancer Institute. Automated Self-administered 24-hour Dietary Recall (ASA24). Bethesda (MD): National Cancer Institute; 2009. Available from: <http://riskfactor.cancer.gov/tools/instruments/asa24.html>.
- USDA Food and Nutrient Database for Dietary Studies, 1.0. Beltsville (MD): Agricultural Research Service, Food Surveys Research Group; 2004.
- Pereira MA, FitzerGerald SJ, Gregg EW, et al. A collection of Physical Activity Questionnaires for health-related research. *Med Sci Sports Exerc* 1997;29:S1–205. PubMed.
- Fuller Z, Horgan G, O'Reilly LM, et al. Comparing different measures of energy expenditure in human subjects resident in a metabolic facility. *Eur J Clin Nutr* 2008;62:560–9. PubMed doi:10.1038/sj.ejcn.1602739.
- Conway JM, Seale JL, Jacobs DR Jr, Irwin ML, Ainsworth BE. Comparison of energy expenditure estimates from doubly labeled water, a physical activity questionnaire, and physical activity records. *Am J Clin Nutr* 2002;75:519–25. PubMed.
- Adams SA, Matthews CE, Ebbeling CB, et al. The effect of social desirability and social approval on self-reports of physical activity. *Am J Epidemiol* 2005;161:389–98.
- Matthews CE, Freedson PS, Hebert JR, et al. Comparing physical activity assessment methods in the seasonal variation of blood cholesterol study. *Med Sci Sports Exerc* 2000;32:976–84. PubMed doi:10.1097/00005768-200005000-00015.
- Schatzkin A, Subar AF, Thompson F, et al. Design and serendipity in establishing a large cohort with wide dietary intake distributions: the National Institutes of Health-AARP Diet and Health Study. *Am J Epidemiol* 2001;154:1119–25. PubMed doi:10.1093/aje/154.12.1119.
- Hunter DJ, Spiegelman D, Adami HO, et al. Cohort studies of fat intake and the risk of breast cancer: a pooled analysis. *N Engl J Med* 1996;334:356–61. PubMed doi:10.1056/NEJM199602083340603.
- Thiebaut ACM, Kipnis V, Chang S-C, et al. Dietary fat and postmenopausal invasive breast cancer in the National Institutes of Health-AARP Diet and Health Study cohort. *J Natl Cancer Inst* 2007;99:451–62. PubMed doi:10.1093/jnci/djk094.
- Slattery ML. Defining dietary consumption: is the sum greater than its parts? *Am J Clin Nutr* 2008;88:14–5. PubMed.
- Kipnis V, Midthune D, Buckman DW, et al. Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes. *Biometrics*. In press 2009.
- National Cancer Institute. Diet History Questionnaire. Bethesda (MD): National Cancer Institute; 2009. Available from: <http://riskfactor.cancer.gov/DHQ/>.

Cancer Epidemiology, Biomarkers & Prevention

Observational Epidemiologic Studies of Nutrition and Cancer: The Next Generation (with Better Observation)

Arthur Schatzkin, Amy F. Subar, Steven Moore, et al.

Cancer Epidemiol Biomarkers Prev 2009;18:1026-1032.

Updated version Access the most recent version of this article at:
<http://cebp.aacrjournals.org/content/18/4/1026>

Cited articles This article cites 32 articles, 7 of which you can access for free at:
<http://cebp.aacrjournals.org/content/18/4/1026.full#ref-list-1>

Citing articles This article has been cited by 10 HighWire-hosted articles. Access the articles at:
<http://cebp.aacrjournals.org/content/18/4/1026.full#related-urls>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and
Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications
Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, use this link
<http://cebp.aacrjournals.org/content/18/4/1026>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's
(CCC)
Rightslink site.