Editorial

See corresponding article on page 141.

The future of phytochemical databases[1,2]

Regina G Ziegler

Observational epidemiologic studies around the world have shown that the risk of colorectal cancer is lower in populations with high intakes of vegetables, fruit, and grains and low intakes of meat and fat and that the risk changes after adoption of a different diet, but we still do not know why (1). In this issue of the Journal, Normén et al (2) report on the use of a carefully constructed database containing the plant sterol content of 292 food items to evaluate whether an increased intake of individual plant sterols, within the 2- to 3-fold range characteristic of the Dutch diet, is predictive of the reduced risk of colorectal cancer in the Netherlands Cohort Study on Diet and Cancer (2). The 5 most abundant plant sterols were measured: 3 unsaturated forms (β-sitosterol, campesterol, and stigmasterol) and 2 saturated forms (β-sitostanol and campestanol). The first, and possibly only other, epidemiologic study to address this hypothesis found that the predominantly vegetarian Seventh-day Adventists in California, who have lower rates of colorectal cancer than do other Californians, also consume more plant sterols (3). Plant sterols, because of their structural similarity to cholesterol, interfere with cholesterol absorption and possibly tumor promotion in the colon. In addition, colonic mucosal cell proliferation, a plausible biomarker of colon carcinogenesis, is decreased in animal models by plant sterols in a dose-dependent fashion (4).

The Netherlands Cohort Study on Diet and Cancer is one of several large prospective studies that collected detailed dietary information at baseline and has now accrued sufficient numbers of patients with common cancers to permit stable analyses. In this cohort of 120 852 participants, 48% men and 52% women aged 55–69 y at entry, 620 colon cancer and 344 rectal cancer cases were diagnosed in the 6.3 y of follow-up since 1986. On the basis of a case-cohort analysis, no clear reduction in risk of colon or rectal cancer was associated with consumption of any of the individual plant sterols in men or women, with or without adjustment for total energy intake. The only hint of an association was an increased risk of rectal cancer with consumption of the 2 less abundant saturated plant sterols. Thus, this study persuasively suggests that an increased intake of plant sterols, up to amounts typically achieved in the Netherlands (≈400 and 500 mg total plant sterols/d in women and men, respectively), may not protect against colorectal cancer. The detail and rigor that characterize the plant sterol database add to the credibility of the findings.

One of the most consistent findings of observational epidemiologic studies is that a reduced risk of many cancers is associated with plant-based diets high in vegetables and fruit (5). The search for which constituents of the plants provide the protection has prompted the development of databases for nonnutrient phytochemicals. The prototype phytochemical database in the United States was that developed by the National Cancer Institute and the US Department of Agriculture (USDA) for 5 individual carotenoids (6). Published analytic data that met explicit criteria were combined with new laboratory measurements to form a database for 120 vegetables and fruit; the carotenoid database is now regularly updated on the USDA Nutrient Data Laboratory Web site (7). Recently, a phytoestrogen database for 4 isoflavones, 1 coumestan, and 2 lignans was published for 112 foods (8); it complements the isoflavone database of the Nutrient Data Laboratory. To facilitate the development of a broadly inclusive database for flavonoids, a chromatographic system has been optimized to quantify 17 common flavonoids from all 5 major classes (9).

The hope is that these phytochemical databases will facilitate a more accurate measure of critical exposures and thus strengthen the associations with disease that epidemiologic studies detect. One promising aspect of most of these new databases is that each uses a single, state-of-the-art analytic method. However, certain simplifications are common to all of the databases. Variations in the composition of vegetables, fruit, grains, and legumes due to plant variety, geographic region, growing season, shipping, storage, and preparation are either not considered or averaged. In addition, for processed and prepared foods, industry choices about ingredients and procedures and differences between brands are rarely incorporated. Often a single typical recipe has to be assigned to a wide variety of dishes. Because of globalization of the food supply and the rising consumption of commercially prepared, ready-to-go, and restaurant foods, the development of accurate food-composition databases is becoming more difficult. More research is required to clarify the major sources of variation in intakes of phytochemicals and to assess the effect that obscuring this variation has on epidemiologic analyses.

Additional assumptions and simplifications need to be made when linking a phytochemical database to a food-frequency questionnaire developed for an epidemiologic study. To enhance participation and the quality of responses, food-frequency questionnaires usually ask questions about ≈150 commonly eaten, aggregated food items, such as whole-wheat bread, broccoli, orange juice, lasagna, cookies, and mayonnaise. Uncommon foods

[1] From the Epidemiology and Biostatistics Program, the Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD.

[2] Address reprint requests to RG Ziegler, National Cancer Institute, Executive Plaza South 8098, Bethesda, MD 20892-7246. E-mail: zieglerr@mail.nih.gov.

especially rich in a phytochemical of interest may not be included in the questionnaire because development of many food-frequency questionnaires predated the current phytochemical hypotheses. Portion size is often not explicitly asked about, and information on brand or variety is rarely collected. Any additional detail in the phytochemical database beyond that in the food-frequency questionnaire must be averaged according to a rubric or ignored. Food-frequency questionnaires lose accuracy in characterizing individual diets because they require study participants to recall and integrate intakes over 1 y or longer; the very nature of such questionnaires guarantees substantial imprecision and probable bias. It is conceivable that the imprecision and inaccuracy of food-frequency questionnaires overwhelms the detail in the phytochemical database.

To address the emerging interest in phytochemicals and other nonnutrient food constituents, is the specificity and immediacy of repeated 24-h dietary recalls or extended dietary records necessary in large prospective studies to address the emerging interest in phytochemicals and other nonnutrient food constituents? It is daunting to consider the imposition of these burdensome dietary assessment instruments on study participants. However, given the critical role of epidemiology in deciphering etiology and determining public policy, it is equally disturbing to question the validity of epidemiologic results.

As phytochemical databases proliferate and are incorporated into epidemiologic studies, analytic challenges will arise. Hundreds of structurally distinct phytochemicals exist in plant foods and may play a role in carcinogenesis (10, 11). Numerous mechanisms have been suggested by in vitro and animal research but have not yet been confirmed in humans (10, 11). Incorporating estimates of phytochemical intake into epidemiologic studies of various cancers will lead to multiple comparisons, many of which will be statistically significant by chance alone. In the absence of firm knowledge about the metabolism and function of phytochemicals in humans, it is unclear whether one should evaluate individual phytochemicals, sum the intakes of phytochemicals on the basis of their structure or activity, or focus on relative intakes and broad patterns of consumption. A conservative approach—one demanding replication in well-designed epidemiologic studies and biological plausibility—is essential.

Validation studies of food-frequency questionnaires generally compare estimated intakes with intakes obtained from repeated 24-h dietary recalls or extended dietary records. Most phytochemical databases are too new to have been calibrated in this manner. However, results of such validations are likely to be too optimistic because many simplifications and assumptions about the phytochemical content of foods are shared by both dietary assessment instruments and because an individual participant tends to show comparable biases in completing both instruments. Thus, validation studies using replicate meals, with one meal saved for laboratory analysis of phytochemical content, may need to be considered. Relying on the new database for individual carotenoids in foods (6, 7), several studies showed correlations of 0.2–0.5 between blood concentrations and dietary intakes of the 5 carotenoids (12, 13). However, it is not possible to distinguish whether these only moderate correlations reflect the influence of absorption and metabolism or inaccurate assessment of intake.

Phytochemical databases have multiple uses. When integrated into dietary data collected from a representative population, they can provide pertinent information on sources, ranges of intake, and intercorrelations. For example, incorporation of the plant sterol database into the baseline dietary information collected for the Netherlands cohort showed that in the Netherlands whole-wheat bread and high-fat margarine each provide more total plant sterols than all vegetables or all fruit (2). Also, the high correlations between phytosterol intake and polyunsaturated fat, dietary fiber, and vitamin E intakes (all correlation coefficients ≥0.8) indicated how difficult it is to disentangle the effect of plant sterols on colorectal cancer risk. In metabolic studies, phytochemical databases facilitate the development of reliable biomarkers for phytochemical intake and function. Given the enthusiasm for identifying the protective factors in a plant-based diet and the multiple mechanisms whereby phytochemicals might reduce cancer risk, phytochemical databases will proliferate. The challenge will be to use them judiciously.

REFERENCES

1. Byers T. Diet, colorectal adenomas, and colorectal cancer. N Engl J Med 2000;342:286–7.
2. Normén AL, Brants HAM, Voorrips LE, Andersson HA, van den Brandt PA, Goldbohm RAB. Plant sterol intakes and colorectal cancer risk in the Netherlands Cohort Study on Diet and Cancer. Am J Clin Nutr 2001;74:141–8.
3. Nair PP, Turjman N, Kessie G, et al. Diet, nutrition intake, and metabolism in populations at high and low risk for colon cancer. Dietary cholesterol, β-sitosterol, and stigmasterol. Am J Clin Nutr 1984;40:927–30.
4. Janezic SA, Rao AV. Dose-dependent effects of dietary phytosterol on epithelial cell proliferation of the murine colon. Food Chem Toxicol 1992;30:611–6.
5. World Cancer Research Fund. Food, nutrition and the prevention of cancer: a global perspective. Washington, DC: American Institute for Cancer Research, 1997.
6. Mangels AR, Holden JM, Beecher GR, Forman MR, Lanza E. Carotenoid content of fruits and vegetables: an evaluation of analytic data. J Am Diet Assoc 1993;93:284–96.
7. US Department of Agriculture. USDA nutrient database. Version current 8 March 2001. World Wide Web: http://www.nal.usda.gov/fnic/foodcomp (accessed 19 April 2001).
8. Horn-Ross PL, Barnes S, Lee M, et al. Assessing phytoestrogen exposure in epidemiologic studies: development of a database (United States). Cancer Causes Control 2000;11:289–98.
9. Merken HM, Beecher GR. Liquid chromatographic method for the separation and quantification of prominent flavonoid aglycones. J Chromatogr A 2000;897:177–84.
10. Steinmetz KA, Potter JD. Vegetables, fruit, and cancer. II. Mechanisms. Cancer Causes Control 1991;2:427–42.
11. Potter JD, Steinmetz K. Vegetables, fruit and phytoestrogens as preventive agents. IARC Sci Publ 1996;21:61–90.
12. Yong LC, Forman MR, Beecher GR, et al. Relationship between dietary intake and plasma concentrations of carotenoids in premenopausal women: application of the USDA-NCI carotenoid food-composition database. Am J Clin Nutr 1994;60:223–30.
13. Michaud DS, Giovannucci EL, Ascherio A, et al. Associations of plasma carotenoid concentrations and dietary intake of specific carotenoids in samples of two prospective cohort studies using a new carotenoid database. Cancer Epidemiol Biomarkers Prev 1998;7:283–90.