Methods for Pooling Results of Epidemiologic Studies

The Pooling Project of Prospective Studies of Diet and Cancer

Stephanie A. Smith-Warner1,2, Donna Spiegelman2,3, John Ritz2,3, Demetrius Albanes4, W. Lawrence Beeson6, Leslie Bernstein6, Franco Berrino7, Piet A. van den Brandt8, Julie E. Buring2,9, Eunyoung Cho10, Graham A. Colditz2,10,11, Aaron R. Folsom12, Jo L. Freudenheim13, Edward Giovannucci1,2,10, R. Alexandra Goldbohm14, Saxon Graham13, Lisa Harnack12, Pamela L. Horn-Ross15, Vittorio Krogh7, Michael F. Leitzmann16, Marjorie L. McCullough17, Anthony B. Miller18, Carmen Rodriguez17, Thomas E. Rohan19, Arthur Schatzkin16, Roy Shore20, Mikko Virtanen21, Walter C. Willett2,12,10, Alicja Wolk22, Anne Zeleniuch-Jacquotte20, Shumin M. Zhang2,9, and David J. Hunter1,2,10,11

1 Department of Nutrition, Harvard School of Public Health, Boston, MA.
2 Department of Epidemiology, Harvard School of Public Health, Boston, MA.
3 Department of Biostatistics, Harvard School of Public Health, Boston, MA.
4 Nutritional Epidemiology Branch, National Cancer Institute, Bethesda, MD.
5 Center for Health Research, School of Medicine, Loma Linda University, Loma Linda, CA.
6 Department of Preventive Medicine and USC/Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA.
7 Epidemiology Unit, National Cancer Institute, Milan, Italy.
8 Department of Epidemiology, Faculty of Health Sciences, Maastricht University, Maastricht, the Netherlands.
9 Division of Preventive Medicine, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA.
10 Channing Laboratory, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA.
11 Harvard Center for Cancer Prevention, Boston, MA.
12 Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN.
13 Department of Social and Preventive Medicine, University at Buffalo, State University of New York, Buffalo, NY.
14 Department of Epidemiology, TNO Quality of Life, Zeist, the Netherlands.
15 Northern California Cancer Center, Fremont, CA.
16 Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD.
17 Epidemiology and Surveillance Research, American Cancer Society, Atlanta, GA.
18 Department of Public Health Sciences, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada.
19 Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY.
20 Department of Environmental Medicine, School of Medicine, New York University, New York, NY.
21 Department of Epidemiology and Health Promotion, National Public Health Institute, Helsinki, Finland.
22 Division of Nutritional Epidemiology, National Institute of Environmental Medicine, Karolinska Institute, Stockholm, Sweden.

Received for publication March 22, 2005; accepted for publication December 21, 2005.

With the growing number of epidemiologic publications on the relation between dietary factors and cancer risk, pooled analyses that summarize results from multiple studies are becoming more common. Here, the authors describe the methods being used to summarize data on diet-cancer associations within the ongoing Pooling Project of Prospective Studies of Diet and Cancer, begun in 1991. In the Pooling Project, the primary data from prospective cohort studies meeting prespecified inclusion criteria are analyzed using standardized criteria for modeling of
exposure, confounding, and outcome variables. In addition to evaluating main exposure-disease associations, analyses are also conducted to evaluate whether exposure-disease associations are modified by other dietary and nondietary factors or vary among population subgroups or particular cancer subtypes. Study-specific relative risks are calculated using the Cox proportional hazards model and then pooled using a random- or mixed-effects model. The study-specific estimates are weighted by the inverse of their variances in forming summary estimates. Most of the methods used in the Pooling Project may be adapted for examining associations with dietary and nondietary factors in pooled analyses of case-control studies or case-control and cohort studies combined.

INCLUSION CRITERIA

The growing number of epidemiologic publications on the relation between diet and cancer risk has heightened the need for methods of summarizing results from multiple studies. These methods include qualitative reviews and quantitative summaries such as meta-analyses of the published literature and pooled analyses of the primary data (also called meta-analyses of individual data) (1). A general framework for conducting pooled analyses entails 1) formulating study inclusion criteria; 2) identifying all potential studies meeting these criteria; 3) obtaining each study’s primary data; 4) creating a standardized database; 5) estimating study-specific exposure-disease associations; 6) examining whether the study-specific results are heterogeneous; 7) calculating pooled estimates, if applicable; and 8) conducting sensitivity analyses to evaluate whether the estimates are robust (2). There are many advantages to reanalyzing the primary data from multiple studies rather than extracting the study-specific relative risks from published articles (1–5). In a pooled analysis, the modeling of the exposure, confounding, and outcome variables, the choice of which variables to control for, and the type of analysis conducted can be standardized, thereby removing potential sources of heterogeneity across studies. Because of larger sample sizes, pooled analyses also offer investigators the opportunity to examine uncommon exposures, rare diseases, and variation in associations among population subgroups with greater statistical power than is possible in individual studies.

The pooling of data from observational studies has become more common recently (6–13). Summary estimates have been calculated using a weighted average of the study-specific estimates (8, 9, 11) or by combining studies into a single data set for the analysis (6, 7, 10, 12, 13). In this paper, we describe the methods that are being used within the ongoing Pooling Project of Prospective Studies of Diet and Cancer (the Pooling Project), an international consortium of cohort studies with the goal of providing the best available summary of data on associations between diet and cancer (14–30). Most of these methods can also be adapted to examine associations in pooled analyses of case-control studies or both case-control and cohort studies combined.

COMPONENT STUDIES

Sixteen studies (32–46) are currently included in the Pooling Project (table 1). As we become aware of new studies meeting the inclusion criteria, the investigators from those studies are invited to join the Project. The Canadian National Breast Screening Study and the Netherlands Cohort Study are each analyzed as case-cohort studies (47), because the investigators in these two studies each selected a random sample of the cohort to provide the person-time data for the cohort and have processed questionnaires for only this random sample and the cases. We divide the person-time and numbers of cases compiled during follow-up of the Nurses’ Health Study into two segments to take advantage of the expanded food frequency questionnaire administered in 1986 as compared with 1980. In this paper, we refer to the follow-up period from 1980 to 1986 as “Nurses’ Health Study A”; the follow-up period beginning in 1986 is referred to as “Nurses’ Health Study B.” Following standard survival data analysis theory, blocks of person-time in different time periods are asymptotically uncorrelated, regardless of the extent to which they are derived from the same people (48, 49). Thus, pooling of the estimates from these two time periods produces estimates and standard errors which are as valid as those from a single time period.

Data collection

The investigators in each Pooling Project study send their primary data on select variables to the Harvard School of Public Health (Boston, Massachusetts). There we inspect

the data for completeness and resolve inconsistencies with the investigators of each study.

Each study used a food frequency questionnaire or diet history instrument that was designed and pretested in its specific study population or a similar population (P. L. Horn-Ross, unpublished data; V. Krogh, unpublished data; A. Wolk, unpublished data) (50–59) (table 1). Although the numbers of items included in the food frequency questionnaires varied over fivefold across the studies (table 1), the study-specific correlation coefficients comparing the food frequency questionnaire used in each cohort or a closely related instrument with multiple dietary records or 24-hour recalls generally exceed 0.40 for total fat, dietary fiber, and several micronutrients (P. L. Horn-Ross, unpublished data; V. Krogh, unpublished data; A. Wolk, unpublished data) (50–59) (table 2).

Information on nondietary risk factors was collected at baseline in each study using self-administered questionnaires. For measured covariates, the proportion of missing data for nondietary risk factors is generally low across studies (table 3). The exception is the Swedish Mammography Cohort, in which some covariate information was available for only one of the two counties in the study.

**Case ascertainment**

Incident cancer diagnoses are identified through follow-up questionnaires, with subsequent medical record review (37, 44, 46), linkage with cancer registries (32, 36, 39–42, 45), or both (33–35, 38, 43). In addition, investigators in some studies ascertain incident and/or fatal outcomes using mortality registries (32, 34, 35, 37–39, 41–46). Case ascertainment has generally been estimated to be greater than 90 percent in each study (table 1).

**STATISTICAL APPROACHES AND RATIONALE**

For each cohort, after applying the exclusion criteria used in that study, we further exclude participants who reported log$_e$-transformed energy intakes beyond three standard deviations from the study-specific log$_e$-transformed mean energy intake of the baseline population (or subcohort, for the case-cohort studies) or who reported a history of cancer (except nonmelanoma skin cancer) at baseline. Additional exclusion criteria may be applied for analyses of specific cancer sites. Because many cancers appear to have hormonal antecedents and because lifestyle factors may differ between women and men, studies including both women and men are split into two studies: a cohort of women and a cohort of men. This conservative approach, in which all estimates are calculated separately for women and men in those studies including both genders, allows for potential effect modification by sex for every determinant of the outcome.

Follow-up time is calculated for each participant from the date on which his/her baseline questionnaire was returned to the date of diagnosis of the specific cancer being examined, the date of death, the date on which the participant moves out of the study area (if applicable), or the end of follow-up, whichever comes first.

In our analyses, we create standardized categories for most confounding variables across studies. We create a missing-data indicator variable for missing responses for each measured confounder in a study, if applicable. As long as 1) the association between the confounding variable and the exposure of interest is weak, or the association between the confounding variable and the outcome is weak, or the confounding variable has little variability in the study and 2) the percentage of missing data within the study is low, the use of the missing-data indicator method is likely to improve efficiency without introducing appreciable bias in comparison with the complete case method (60, 61). As table 3 shows, the proportion of missing data for each covariate across studies is generally low, satisfying one of the conditions for valid use of the missing-data indicator method. In addition, potentially confounding factors generally have had moderate-to-weak associations with the cancer sites we have examined and have had low-to-moderate correlations with the dietary exposures that are of primary interest in the Pooling Project. Information on age, which is typically the strongest measured risk factor for cancer incidence, is never missing in the constituent studies.

**Two-stage analysis**

Our analytic approach generally is a two-stage process. In the first step, we calculate study-specific relative risks using the Cox proportional hazards model (49), defined through the hazard function $h$ by

$$h_{jks}(t|u_{is},x_{is}) = h_{0jks}(t)\exp(\alpha_iu_{is} + \beta_s x_{is})$$

for $s = 1, \ldots, S$, where $s$ is the study number, $t$ is follow-up time, $u_{is}$ and $x_{is}$ are the study-specific confounding and exposure variables, respectively, for individual $i$ in study $s$, and $h_{0jks}(t)$ is the baseline incidence rate at age $j$ (in years), in calendar year $k$, and for time since entry into the study $t$. The estimated study-specific log relative risks for a one-unit increase in the exposures, $x_{is}$, are given by the $\beta_s$. The study-specific log relative risks for a one-unit increase in the confounding variables, $u_{is}$, are given by the $\alpha_i$. Stratifying jointly by age at baseline (years) and the year in which the baseline questionnaire was returned (indexed by $j$ and $k$, respectively) and treating follow-up time (in years) as the time metric in the Cox model is equivalent to treating age as the time metric in the Cox model and stratifying jointly on calendar time (in years) and duration of time in the study, with one exception: There is a difference in which two-way interactions are allowed. With our approach, no assumptions are made about the shape of the age or calendar-year incidence curves, each of which is fully adjusted for the other, and arbitrary two-way interactions of the joint dependency of the outcome on age and calendar time are allowed. Each case-cohort study is analyzed using EPICURE software (HiroSoft International Corporation, Seattle, Washington) (47, 62); each remaining study is analyzed using SAS PROC PHREG (SAS Institute, Inc., Cary, North Carolina) (63).

If case-control studies were included in our pooled analyses, the model for these studies would be similar to equation 1, except that we would stratify the participants by...
<table>
<thead>
<tr>
<th>Study</th>
<th>Study population</th>
<th>Location</th>
<th>Study dates</th>
<th>Baseline cohort size</th>
<th>Age (years) at baseline</th>
<th>Food frequency questionnaire/diet history instrument</th>
<th>Outcome ascertainment</th>
<th>Estimated case ascertainment rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adventist Health Study (33)</td>
<td>Non-Hispanic White men and women living in Seventh-Day Adventist households</td>
<td>California, United States</td>
<td>1976–1982</td>
<td>18,403</td>
<td>&gt;24</td>
<td>Frequency FQs/MRR; cancer registry; mortality registry</td>
<td>&gt;99</td>
<td></td>
</tr>
<tr>
<td>Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (34)</td>
<td>Male smokers who participated in a randomized double-blind placebo-controlled clinical trial of α-tocopherol and β-carotene supplement use</td>
<td>Southwestern Finland</td>
<td>1985 onward (ongoing)</td>
<td>0</td>
<td>26,987</td>
<td>Frequency and portion size FQs/MRR; cancer registry; mortality registry</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Breast Cancer Detection Demonstration Project Follow-up Cohort (35)</td>
<td>Subset of women participating in a breast cancer screening program in 1973–1980 who had been diagnosed with breast cancer or had undergone or been recommended to receive a breast biopsy, plus a random sample of the remaining women who had been screened</td>
<td>United States</td>
<td>1987 onward (ongoing)</td>
<td>41,987</td>
<td>0</td>
<td>Frequency and portion size FQs/MRR; cancer registry; mortality registry</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>California Teachers Study (45)</td>
<td>Active and retired female teachers and administrators participating in the California State Teachers Retirement System</td>
<td>California, United States</td>
<td>1995 onward (ongoing)</td>
<td>100,036</td>
<td>0</td>
<td>Frequency and portion size Cancer registry; mortality registry</td>
<td>&gt;97</td>
<td></td>
</tr>
<tr>
<td>Canadian National Breast Screening Study (36)</td>
<td>Women who participated in a multicenter randomized controlled trial of mammography screening for female breast cancer</td>
<td>Canada</td>
<td>1980 onward (ongoing)</td>
<td>56,837</td>
<td>0</td>
<td>Frequency and portion size Cancer registry; mortality registry</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Cancer Prevention Study II Nutrition Cohort (38)</td>
<td>Subset of men and women participating in Cancer Prevention Study II who completed a diet questionnaire in 1992</td>
<td>United States</td>
<td>1992 onward (ongoing)</td>
<td>74,053</td>
<td>66,090</td>
<td>Frequency and portion size FQs/MRR; cancer registry; mortality registry</td>
<td>&gt;90</td>
<td></td>
</tr>
<tr>
<td>Health Professionals Follow-up Study (37)</td>
<td>Male dentists, optometrists, osteopathic physicians, podiatrists, pharmacists, and veterinarians</td>
<td>United States</td>
<td>1986 onward (ongoing)</td>
<td>0</td>
<td>47,673</td>
<td>Frequency of specified portions FQs/MRR; mortality registry</td>
<td>&gt;94</td>
<td></td>
</tr>
<tr>
<td>Iowa Women's Health Study (41)</td>
<td>Postmenopausal women selected randomly from the 1985 Department of Transportation’s driver’s license list in Iowa</td>
<td>Iowa, United States</td>
<td>1986 onward (ongoing)</td>
<td>34,603</td>
<td>0</td>
<td>Frequency of specified portions Cancer registry; mortality registry</td>
<td>98§</td>
<td></td>
</tr>
<tr>
<td>Netherlands Cohort Study (40)</td>
<td>Men and women from 204 municipal population registries throughout the Netherlands</td>
<td>The Netherlands</td>
<td>1986 onward (ongoing)</td>
<td>62,573</td>
<td>58,279</td>
<td>Frequency of specified portions Cancer registry; pathology database</td>
<td>&gt;95</td>
<td></td>
</tr>
<tr>
<td>Study Description</td>
<td>Participants</td>
<td>Location</td>
<td>Time Period</td>
<td>Frequency</td>
<td>Cancer Registry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------------------</td>
<td>---------------------------------------</td>
<td>-----------------------------------</td>
<td>---------------</td>
<td>-----------</td>
<td>----------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York State Cohort (42)</td>
<td>Male and female residents who had had the same address and telephone number for the previous 18 years</td>
<td>New York, United States</td>
<td>1980–1987</td>
<td>22,550</td>
<td>50–93 45 Past year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York University Women's Health Study (43)</td>
<td>Women visiting a breast screening clinic who had not used any hormonal medications or been pregnant in the previous 6 months</td>
<td>New York, United States</td>
<td>1985 onward (ongoing)</td>
<td>13,258</td>
<td>0 34–65 71 Past year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nurses' Health Study A (37)</td>
<td>Female registered nurses</td>
<td>United States</td>
<td>1980–1986</td>
<td>88,651</td>
<td>0 34–59 61 Past year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nurses' Health Study B (37)</td>
<td>Female registered nurses</td>
<td>United States</td>
<td>1986 onward (ongoing)</td>
<td>68,540</td>
<td>0 40–65 131 Past year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nurses' Health Study II (46)</td>
<td>Female registered nurses</td>
<td>United States</td>
<td>1991 onward (ongoing)</td>
<td>93,894</td>
<td>0 26–46 133 Past year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prospective Study on Hormones, Diet and Breast Cancer (39)</td>
<td>Female volunteers recruited from the general population using mass media advertising and from breast cancer prevention units</td>
<td>Varese Province, Italy</td>
<td>1987 onward (ongoing)</td>
<td>9,027</td>
<td>0 35–69 177 Past year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swedish Mammography Cohort (32)</td>
<td>Women who participated in a population-based mammography screening program</td>
<td>Västmanland and Uppsala counties, Sweden</td>
<td>1987 onward (ongoing)</td>
<td>61,463</td>
<td>0 40–74 67 Past 6 months</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women's Health Study (44)</td>
<td>Female health professionals who participated in a randomized, double-blind, placebo-controlled trial of low-dose aspirin, β-carotene, and vitamin E use</td>
<td>United States</td>
<td>1993 onward (ongoing)</td>
<td>38,384</td>
<td>0 ≥45 131 Past year</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The baseline cohort size corresponds to the number of participants in the Pooling Project database for the renal cell cancer analyses in the California Teachers Study (45) and for the colorectal cancer analyses in the remaining studies.
† FQs, follow-up questionnaires; MRR, medical record review.
‡ For California residents only.
§ For Iowa residents only.
¶ Cancer outcomes in the New York State Cohort (42) were identified through linkage with a cancer registry; thus, it is difficult to determine the follow-up rate in the cohort. When a subset of the cohort was followed intensively, loss to follow-up was not related to exposure.
### TABLE 2. Correlation coefficients (CCs) for nutrient intakes estimated using a food frequency questionnaire versus a comparison method for studies in the Pooling Project of Prospective Studies of Diet and Cancer, 1991–2004*

<table>
<thead>
<tr>
<th>Study</th>
<th>Sex</th>
<th>No. of participants</th>
<th>Comparison method</th>
<th>Type of CC</th>
<th>Total fat</th>
<th>Saturated fat</th>
<th>Mono-unsaturated fat</th>
<th>Poly-unsaturated fat</th>
<th>Dietary fiber</th>
<th>Alcohol</th>
<th>Vitamin A†</th>
<th>Vitamin C†</th>
<th>Vitamin E†</th>
<th>Folate†</th>
<th>Calcium†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adventist Health Study (50)</td>
<td>Women</td>
<td>103</td>
<td>Five 24-hour recalls over 6 months</td>
<td>Spearman CCs‡</td>
<td>0.40</td>
<td>0.45</td>
<td>0.41</td>
<td>0.26</td>
<td>0.47‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Men</td>
<td>44</td>
<td>Five 24-hour recalls over 6 months</td>
<td>Spearman CCs‡</td>
<td>0.38</td>
<td>0.57</td>
<td>0.29</td>
<td>0.15</td>
<td>0.50‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (51)</td>
<td>Men</td>
<td>178</td>
<td>Twelve 2-day diet records over 6 months</td>
<td>Energy-adjusted, deattenuated Pearson CCs</td>
<td>0.75</td>
<td>0.79</td>
<td>0.68</td>
<td>0.85</td>
<td>0.82</td>
<td>0.68</td>
<td>0.71</td>
<td>0.82</td>
<td>0.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>California Teachers Study (unpublished)</td>
<td>Women</td>
<td>185</td>
<td>Four 24-hour recalls over 10 months</td>
<td>Energy-adjusted, deattenuated Pearson CCs</td>
<td>0.64</td>
<td>0.82</td>
<td>0.41</td>
<td>0.23</td>
<td>0.77</td>
<td>0.82</td>
<td>0.35‡</td>
<td>0.62‡</td>
<td>0.82‡</td>
<td>0.73‡</td>
<td>0.30‡</td>
</tr>
<tr>
<td>Canadian National Breast Screening Study (52)</td>
<td>Women</td>
<td>108</td>
<td>7-day diet records</td>
<td>Energy-adjusted Pearson CCs</td>
<td>0.44</td>
<td>0.61</td>
<td>0.43#</td>
<td>0.40‡</td>
<td>0.60</td>
<td>0.59</td>
<td>0.60</td>
<td>0.53</td>
<td>0.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer Prevention Study II Nutrition Cohort (53)</td>
<td>Women</td>
<td>188‡</td>
<td>Four 24-hour recalls over 1 year</td>
<td>Energy-adjusted, deattenuated Pearson CCs</td>
<td>0.66</td>
<td>0.66</td>
<td>0.58#</td>
<td>0.42‡</td>
<td>0.61</td>
<td>0.77</td>
<td>0.65</td>
<td>0.27</td>
<td>0.43</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Men</td>
<td>229‡</td>
<td>Four 24-hour recalls over 1 year</td>
<td>Energy-adjusted, deattenuated Pearson CCs</td>
<td>0.58</td>
<td>0.64</td>
<td>0.61#</td>
<td>0.48‡</td>
<td>0.64</td>
<td>0.82</td>
<td>0.65</td>
<td>0.23</td>
<td>0.51</td>
<td>0.57</td>
<td></td>
</tr>
<tr>
<td>Health Professionals Follow-up Study (54)</td>
<td>Women</td>
<td>127</td>
<td>Two 7-day diet records over 6 months</td>
<td>Energy-adjusted Pearson CCs</td>
<td>0.67</td>
<td>0.75</td>
<td>0.68</td>
<td>0.37</td>
<td>0.68</td>
<td>0.86‡</td>
<td>0.61</td>
<td>0.77</td>
<td>0.42</td>
<td>0.70</td>
<td>0.60</td>
</tr>
<tr>
<td>Iowa Women's Health Study (55)</td>
<td>Women</td>
<td>44</td>
<td>Five 24-hour recalls over 2 months</td>
<td>Energy-adjusted Pearson CCs</td>
<td>0.62</td>
<td>0.59</td>
<td>0.62</td>
<td>0.43</td>
<td>0.24‡</td>
<td>0.32</td>
<td>0.14</td>
<td>0.53</td>
<td>0.79</td>
<td>0.26</td>
<td>0.49</td>
</tr>
<tr>
<td>Netherlands Cohort Study (56)</td>
<td>Women and men</td>
<td>109</td>
<td>Three 3-day diet records over 1 year</td>
<td>Energy- and sex-adjusted Pearson CCs</td>
<td>0.53</td>
<td>0.58‡</td>
<td>0.80</td>
<td>0.79</td>
<td>0.86</td>
<td>0.76</td>
<td>0.58</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York State Cohort (unpublished)</td>
<td>Women</td>
<td>190</td>
<td>Simulated study</td>
<td>Energy-adjusted, deattenuated Pearson CCs</td>
<td>0.21</td>
<td>0.18</td>
<td>0.41</td>
<td>0.25</td>
<td>0.53</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Men (57)</td>
<td>127</td>
<td>Simulated study</td>
<td>Energy-adjusted Pearson CCs‡</td>
<td>0.57</td>
<td>0.60</td>
<td>0.61</td>
<td>0.22</td>
<td>0.65</td>
<td>0.39</td>
<td>0.76</td>
<td>0.46</td>
<td>0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nurses’ Health Study A (58)</td>
<td>Women</td>
<td>173</td>
<td>Four 7-day diet records over 1 year</td>
<td>Energy-adjusted Pearson CCs</td>
<td>0.53</td>
<td>0.59</td>
<td>0.48</td>
<td>0.48‡</td>
<td>0.90‡</td>
<td>0.36</td>
<td>0.36</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nurses’ Health Study B (59)</td>
<td>Women</td>
<td>191</td>
<td>Two 7-day diet records over 1 year</td>
<td>Energy-adjusted, deattenuated Pearson CCs</td>
<td>0.57</td>
<td>0.68</td>
<td>0.58</td>
<td>0.48</td>
<td>0.79</td>
<td>0.76</td>
<td>0.76</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
study-specific matching factors. Since the Cox model and the conditional logistic regression model produce algebraically identical log-(partial)-likelihood functions, SAS PROC PHREG could also be used for case-control studies to estimate study-specific odds ratios and their standard errors.

The second step consists of pooling the study-specific relative risks using a random-effects model (64–66) given by

$$\hat{\beta}_s = \beta + b_s + e_s,$$

(2)

where the $\hat{\beta}_s$ are the estimated study-specific exposure-disease effects, $\beta$ is the underlying common exposure-disease association, $b_s$ are the random between-studies effects, and $e_s$ are the within-study errors. Both $b_s$ and $e_s$ are assumed to be independent and asymptotically normally distributed with means of zero and variances of $\sigma^2_b$ and $\sigma^2_e$, respectively, and $\sigma^2_s = \text{Var}(\hat{\beta}_s)$. The study-specific exposure-disease effects are weighted by the inverse of their variances using

$$\hat{\beta} = \frac{1}{S} \sum_{s=1}^{S} w_s \hat{\beta}_s,$$

where

$$w_s = \left(\frac{\hat{\sigma}_B^2 + \hat{\sigma}_s^2}{\hat{\sigma}_B^2}\right)^{-1/2} = \sum_{t=1}^{S} \left(\frac{\hat{\sigma}_B^2 + \hat{\sigma}_t^2}{\hat{\sigma}_B^2}\right)^{-1}.$$

When the exposure variable is categorized into different levels, we calculate a pooled relative risk for each category separately.

We test for the statistical significance of between-studies heterogeneity among the study-specific exposure-disease estimates using the $Q$ test statistic given by

$$Q = \sum_{s=1}^{S} w_s^* (\hat{\beta}_s - \hat{\beta})^2,$$

(3)

where $w_s^* = \sqrt{\text{Var}(\hat{\beta}_s)}$. The $Q$ test statistic follows an approximate $\chi^2_{S-1}$ distribution (66, 67).

For the exposures of interest, we generally categorize participants into study-specific quantiles. Because the quantile approach does not take into account true differences in the distribution of population intakes across studies, we also create categories defined by identical absolute intake cutpoints across studies. Misclassification can also occur in the analyses based on identical absolute intake cutpoints, because reported intakes may differ across studies based on differences in the dietary assessment methods used. However, when possible we adjust our results for measurement error in the individual studies.

**Aggregated analysis**

We can also conduct analyses in which the data from all studies are combined into one data set (referred to as an aggregated analysis). A single exposure-disease effect is then calculated using the Cox proportional hazards model, including stratification by study, age at baseline, and the year in which the baseline questionnaire was returned.
Although combining the data from all studies is one way to take advantage of differences in the distributions of the exposure variable across studies, it assumes that the exposure was measured in comparable ways across studies. Because the distributions of dietary variables may differ across studies due to true differences in actual intake and due to differences in the dietary assessment methods used (and other study-specific sources of error), this assumption may not be reasonable, except for nutrients that come from a small number of food sources (e.g., alcohol). In addition, combining the studies into one data set assumes that there is no between-studies heterogeneity in the associations of the outcome with the exposure or any of the covariates. In the few instances where we have conducted both pooled and aggregated analyses, the results have been essentially identical (16, 25, 30). Nevertheless, because it is difficult to test the underlying assumptions, we have opted to use two-stage analyses as our primary analytic strategy.

Trend analysis

To test the significance of trends in disease risk over exposure categories, we conduct separate analyses in which participants are assigned the study-specific median value of their respective category (given by $\text{med}_{js}$ for $j = 1, \ldots, J$, where $J$ is the number of levels in which the exposure variable is categorized). For each study, we fit a Cox proportional hazards model with regression terms $\beta_s z_{is}$ for $s = 1, \ldots, S$, where $s$ is the study number and $z_{is}$ takes on the values $\text{med}_{js}$ corresponding to the category in which the individual’s exposure value falls. We then compute the pooled estimate for the regression coefficient for trend using a random-effects model (64–66). The pooled test for trend is a Wald test of the hypothesis $H_0: \beta = 0$. We test for the statistical significance of between-studies heterogeneity among the study-specific regression coefficients using the $Q$ test statistic (66, 67).

We also evaluate whether associations between dietary factors and cancer risk are linear by comparing nonparametric regression curves using restricted cubic splines with the linear model using the likelihood ratio test, and by visual inspection of the restricted cubic spline graphs (68, 69). For these analyses, the studies are combined into a single data set stratified by study.

Evaluation of heterogeneity of effects

An advantage of a pooled analysis is the ability to evaluate whether the exposure-disease association is modified by other risk factors. In these analyses, if the exposure-disease association is log-linear and the potential effect modifier is an ordinal or binary variable, we first compute estimates of the exposure-disease association and their standard errors for each study within each category of the

---

### TABLE 3. Prevalences of missing data for select nondietary factors across studies in the Pooling Project of Prospective Studies of Diet and Cancer, 1991–2004

<table>
<thead>
<tr>
<th>No. of studies in which the factor was measured*</th>
<th>% of missing data across studies (range)</th>
<th>Studies with &lt;5% missing data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>17</td>
<td>0–23</td>
</tr>
<tr>
<td>Education</td>
<td>17†</td>
<td>0–23</td>
</tr>
<tr>
<td>Body mass index</td>
<td>17</td>
<td>0–8</td>
</tr>
<tr>
<td>Smoking status</td>
<td>15</td>
<td>0–5</td>
</tr>
<tr>
<td>Physical activity</td>
<td>14</td>
<td>0–12</td>
</tr>
<tr>
<td>Multivitamin use</td>
<td>14</td>
<td>0–8</td>
</tr>
<tr>
<td>Age at menarche†,§</td>
<td>14</td>
<td>0–3</td>
</tr>
<tr>
<td>Parity‡</td>
<td>15</td>
<td>0–10</td>
</tr>
<tr>
<td>Menopausal status†,§</td>
<td>14</td>
<td>0–18</td>
</tr>
<tr>
<td>Oral contraceptive use†,§</td>
<td>13</td>
<td>0–21</td>
</tr>
<tr>
<td>Postmenopausal hormone use†,§</td>
<td>13</td>
<td>0–16</td>
</tr>
</tbody>
</table>

* For this table, Nurses’ Health Study A (1980–1986) and Nurses’ Health Study B (1986–present) were counted as two separate studies (see Materials and Methods).
† All participants in the California Teachers Study, the Health Professionals Follow-up Study, the Nurses’ Health Study, and Nurses’ Health Study II were assumed to have received additional education after graduating from high school, because these populations were selected on the basis of their employment in occupations requiring a post-high-school education.
‡ Only cohort studies including women are included here. The prevalence of missing data was calculated only among the female participants.
§ For the Swedish Mammography Cohort, only the percentage of missing data for women living in Uppsala County is included, since these data were not collected for women living in Vastmanland County.
¶ Among postmenopausal women only.
potential effect modifier. The model uses the same format as equation 1, but
\[
h_{jkls}(t|u_{is}, x_{is}) = h_{0jkl}(t)\exp(\alpha_s u_{is} + \beta_s x_{is}),
\]
where \( l = 1, \ldots, L \) levels of the effect modifier, \( x_{is} \) is the study-specific exposure variable, and \( \alpha_s, \beta_s \) are the estimated study-specific log relative risks for a one-unit increase in the confounding variables, \( u_{is} \). The study-specific estimates \( \beta_s \) for each stratum are then pooled across studies and exponentiated to obtain the relative risk for each level of the potential effect modifier. For assessment of the statistical significance of the interaction, the Cox proportional hazards model is
\[
h_{jkls}(t|u_{is}, x_{is}, m_{is}) = h_{0jkl}(t)\exp(\alpha_s u_{is} + \xi_s m_{is} + \beta_s x_{is} + \gamma_s x_{is} m_{is}),
\]
where \( \gamma_s \) is the study-specific estimate for the cross-product term of the potential effect modifier variable \( (m_{is}) \) times the exposure variable \( (x_{is}) \) and \( \xi_s \) is the study-specific main effect of the effect modifier. The study-specific estimates \( \hat{\gamma}_s \) are then pooled across studies, and the \( p \) value corresponding to the test for interaction \( (H_0: \gamma = 0) \) is obtained from a Wald test based upon the pooled \( \hat{\gamma} \).

We use a mixed-effects meta-regression model (70) to test for effect modification when the exposure-disease association is nonlinear, when the potential effect modifier is a polytomous nominal variable, or when effect modification can be assessed only between studies. As an example, consider the test for effect modification by gender. The model here is a slightly modified version of equation 2:
\[
\hat{\beta}_s = \beta_0 + \beta_1 z_s + b_s + e_s
\]
for \( s = 1, \ldots, S \), where \( s \) is the study number, the \( \hat{\beta}_s \) are the estimated study-specific exposure-disease effects, \( \beta_0 \) is the log relative risk for the exposure in the reference level of the modifier (here, men), \( \beta_1 \) is the difference in the log relative risks between the reference level and each of the other levels (here, between genders), \( z_s = 1 \) if study \( s \) is carried out among women and \( z_s = 0 \) if it is carried out among men, \( b_s \) are the study-specific random effects, and \( e_s \) are the within-study sampling errors. The Wald test statistic based on the estimate \( \hat{\beta}_s \) and its standard error is used to test the null hypothesis \( (H_0: \beta_1 = 0) \) that there is no modification of the effect of exposure on the outcome by levels of the potential effect modifier (here, between genders).

Assessment of heterogeneity by outcome subtype

We can also evaluate whether associations differ by cancer subtype. For these analyses, we fit separate Cox proportional hazards models (equation 1) for each subtype. Occurrences of the cancer under study that are of a different subtype are censored at their date of diagnosis. The relative risks obtained for each subtype that are estimated in this way are asymptotically uncorrelated (71–73). In addition, because these estimates are asymptotically normally distributed with variances given by the square of their respective estimated standard errors, any linear combination of the different estimates is normally distributed, and it follows from the Cramer-Wald device (74) that the multivariate vector obtained by combining all of the competing risk estimates is multivariate normal. The corresponding variances are in the diagonal of the covariance matrix, and zeroes are in the off-diagonal. To test the null hypothesis that there is no difference in the pooled exposure-disease parameters among the subtypes, we use a contrast test (75). For example, to test whether the pooled exposure-disease parameters differed among three subtypes, we would use the test statistic \( Z^2 \) given by
\[
Z^2 = (C\hat{\beta})^T (C\hat{\Sigma}C^T)^{-1}(C\hat{\beta}),
\]
where \( C \) is a contrast matrix whose first and second rows are \( (1, -1, 0) \) and \( (1, 0, -1) \), \( \hat{\beta} \) is the vector of the pooled estimates of the exposure-disease association for the different subtypes, and \( \hat{\Sigma} \) is its estimated covariance matrix. The \( Z^2 \) statistic in this example has an approximate \( \chi^2 \) distribution with 2 df (defined by the number of different subtypes minus 1) (75). These methods can also be used to construct tests for heterogeneity of effects between any set of cancers or other outcomes.

Measurement error correction

As with most exposures, measurement of dietary variables is not free from error. Measurement error in dietary data derives from normal within-person variation in intakes over time (76) and from errors associated with self-reports (77). Therefore, the relative risks will be biased, usually towards the null, but can be biased in either direction when there is also error in measuring confounding variables (78). One can use the validation data from each study to regress the “gold standard” (or an unbiased estimate of the gold standard, an “alloyed” gold standard (79)) on the error-prone measurement and confounding variables to obtain a correction factor. This correction factor can then be used to calibrate the uncorrected estimates of the exposure effect of interest obtained from logistic and Cox regression models (77, 79, 80). If the errors in the alloyed gold standard are correlated with the errors in the usual measure of dietary intake, the regression calibration method for measurement error correction will remove some, but not all, of the bias in the effect estimate (81). However, it appears that energy adjustment removes much of the bias in this method due to correlated errors for at least some dietary variables (e.g., protein) (82, 83). To remove the remaining bias, an additional method of assessment of intake is needed, such as a biomarker (81).

In the measurement error correction analyses, for each study, the true intake of the particular nutrient being evaluated or an unbiased estimate of the true intake (e.g., intakes calculated from several dietary records or 24-hour recalls) is regressed on the surrogate measurement of that nutrient (calculated from the food frequency questionnaire) to obtain the coefficient \( \hat{\lambda}_s \) and its estimated standard error. We then derive the corrected estimate of the log relative risk as \( \hat{\beta}_s / \hat{\lambda}_s \), where \( \hat{\beta}_s \) is the uncorrected estimated effect in each study from a logistic regression or Cox proportional hazards model.

regression analysis. The standard error of $\hat{b}_s / \hat{\lambda}_s$ is derived using the delta method (84). One can simultaneously correct for the error in several covariates in all point estimates and their standard errors using a multivariate extension of measurement error correction (79, 85). The corrected coefficient estimates are then pooled into a summary estimate. If a study has poor validity of nutrient measurements, its variance will be large, and the study will thus have little weight when the study-specific results are pooled. In addition, under the required assumption that the dietary records and 24-hour recalls provide an unbiased estimate of nutrient intake (even if subject to random error), this approach calibrates the estimated relative risks to a common unit of measurement across studies, thereby adjusting for systematic errors due to differences in the food frequency questionnaires used in the various studies.

STRENGTHS AND LIMITATIONS

The Pooling Project of Prospective Studies of Diet and Cancer provides a large collection of data in which multiple diet-and-cancer hypotheses can be examined with greater statistical power than is available in any one study. Each study included in the Pooling Project is a prospective cohort study in which diet was assessed prior to development of disease, thereby limiting recall and selection biases. In the Pooling Project, we standardize the modeling of the exposure and confounding variables to remove potential sources of noncomparability and heterogeneity that occur in the published literature. We are able to examine associations over a wide range of intakes with greater precision than in the individual studies, because of the larger sample size and the different diets consumed across the populations. In addition, we can evaluate whether associations are modified by other factors and whether associations differ among cancer subtypes. Because inclusion of an individual study in a particular analysis is not dependent on whether those investigators have published findings on that association, publication bias does not affect our pooled analyses—as opposed to meta-analyses of the published literature, for which approximately half of the results may have some indication of publication bias (86). Finally, results from these pooled analyses may assist epidemiologists and other health professionals in synthesizing the vast amount of published data on specific diet-cancer associations.

A limitation of the Pooling Project is that it was planned retrospectively. Thus, there are differences in how the included studies were designed and implemented. First, the studies comprise populations from different geographic regions with different age ranges and education levels. However, these differences in study population characteristics may be considered a strength, particularly if the results are consistent across studies. Second, the dietary assessment methods used vary across studies, which may lead to artifactual differences in estimated intakes across studies, in addition to any true between-population differences in intakes. However, it is also possible that validity is enhanced by the use of study-specific questionnaires, since they may be tailored for use in each component study. Some heterogeneity of assessment instruments cannot be avoided, even in prospectively planned pooled studies—if, for instance, the language spoken and the foods consumed differ between populations. Another limitation of the Pooling Project is that only current diet at baseline was measured in most of the studies; thus, we cannot examine the effects of dietary changes occurring during follow-up or assess associations with diet at younger ages. There may be differential control for confounding across studies because the nondietary variables that were measured varied across studies, although many important potential confounders were measured in most studies. In addition, by standardizing which confounding variables are included in the multivariate models and their categorization, we have minimized between-studies heterogeneity resulting from how potentially confounding variables were modeled. A final restriction is our inability to examine effect modification by race and ethnicity, because the Pooling Project currently includes studies from only North America and Europe and a predominantly Caucasian population; however, as studies from other regions and with persons of different ethnicities become eligible to join the Pooling Project, the ethnic composition of the Pooling Project will expand.

Despite these limitations and restrictions, the data compiled in the Pooling Project are a valuable resource for prospectively investigating associations between diet and cancer, particularly for population subgroups, less common cancers, and specific cancer subtypes. In our analyses, we use standardized criteria to define each variable in order to reduce potential sources of between-studies heterogeneity. We then evaluate whether associations are consistent across different study populations. Finally, the methods that we use in the Pooling Project may be modified to pool data from both case-control and cohort studies to examine associations between dietary and nondietary risk factors and disease.

ACKNOWLEDGMENTS

This research was funded by National Institutes of Health grants CA55075 and CA78548. The work was performed at the Harvard School of Public Health (Boston, Massachusetts).

Conflict of interest: none declared.

REFERENCES


