Pooling Prospective Studies to Investigate the Etiology of Second Cancers

Amanda Black1, Todd M. Gibson1, Meredith S. Shiels1, Yikyung Park1, Kim Robien3, Demetrius Albanes1, Stephanie J. Weinstein1, Laura E. Beane Freeman1, Gabriella Andreotti1, Mark P. Purdue1, Joseph F. Fraumeni1, Margaret A. Tucker1, Robert N. Hoover1, James R. Cerhan4, Anne Zeleniuch-Jacquotte5, Rochelle E. Curtis1, Joanne Elena2, Joshua N. Sampson1, Amy Berrington de Gonzalez1, and Lindsay M. Morton1

Abstract

Background: With over 13 million cancer survivors in the United States today, second cancers are of rapidly growing importance. However, data on nontreatment risk factors for second cancers are sparse. We explored the feasibility of pooling data from cohort studies of cancer incidence to investigate second cancer etiology.

Methods: We combined data from five prospective studies including more than 800,000 individuals. We compared study designs and populations; evaluated availability of and ability to harmonize risk factor data; compared incidence and survival for common first primary malignancies and incidence of second primary malignancies; and estimated sample size requirements.

Results: Overall, 96,513 incident, first primary malignancies were diagnosed during 1985 to 2009. Incidence rates and survival following the first primary varied among the cohorts, but most of the heterogeneity could be explained by characteristics of the study populations (age, sex, smoking, and screening rates). A total of 7,890 second primary cancers (excluding original primary site) were identified, yielding sufficient statistical power (≥80%) for detecting modest associations with risk of all second cancers among survivors of common first primary malignancies (e.g., colorectal cancer); however, there were insufficient events for studying survivors of rarer cancers or identifying risk factors for specific second cancers.

Conclusions: Pooling data from cohort studies to investigate nontreatment risk factors for second primary cancers seems feasible but there are important methodologic issues—some of which are barriers to specific research questions—that require special attention.

Impact: Increased understanding of nontreatment risk factors for second cancers will provide valuable prevention and surveillance information.

Introduction

The prevalence of cancer survivors in the United States has increased steadily in recent decades due to dramatic improvements in survival resulting from new treatments and early detection (1). With over 13 million survivors in the United States today, understanding the long-term health of cancer survivors is of vital public health importance. Second primary cancers (referred to hereafter as second cancers) are a leading cause of morbidity and mortality in this population, yet the etiology of most second cancers is largely unknown. A recent study estimated that fewer than 10% of second solid cancers among adults are due to radiotherapy, the main treatment modality known to increase cancer risk, suggesting that the majority result from other risk factors, such as lifestyle or genetics (2). Characterizing second cancer risks may identify patients for targeted cancer prevention efforts or increased surveillance and can provide key new insights into carcinogenesis.

The ideal study to investigate the etiology of second cancers would be very large and longitudinal with detailed data on treatment and exposures (e.g., cigarette smoking, obesity) before and after a first cancer diagnosis. Because it is seldom feasible to enroll a large healthy population and collect such information, evidence on second cancer etiology primarily comes from registry studies (typically large and representative but lack individual-level exposures; refs. 3–7), clinical trials (typically...
IWHS began during the 1980s, and the rest of the studies, and ATBC was conducted in Finland. ATBC and IWHS, NIH-AARP, and PLCO were all U.S.-based studies: the Agricultural Health Study (AHS; ref. 19), Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC; ref. 20), Iowa Women’s Health Study (IWHS; ref. 21), NIH-AARP (formerly American Association Study of Retired Persons) Diet and Health Study (22), or case–control studies that rely on medical record abstraction (12–16). Recently, attention has turned to leveraging existing epidemiologic cohorts by supplementing them with information on cancer treatment, survival, and second cancers (17, 18). Individual cohorts have only moderate numbers of survivors of specific cancers, and small numbers of second cancers among these survivors; however, combined, they comprise a very large number of individuals.

Pooling data from prospective cohort studies presents challenges, such as differences in study design (observational versus interventional study), populations (e.g., age, gender, and calendar period), methods for patient follow-up, methods and timing of exposure assessment, and the lack of clinical data on detailed screening or availability of treatment information. These challenges introduce both technical issues (e.g., data harmonization) and potential sources of bias (e.g., selection bias). We therefore combined data from five prospective epidemiologic studies to explore the feasibility of pooling data from cohort studies of cancer incidence to investigate second cancer etiology.

Materials and Methods

CoHorts and study population

Participants were enrolled in one of five prospective cohort studies: the Agricultural Health Study (AHS; ref. 19), Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC; ref. 20), Iowa Women’s Health Study (IWHS; ref. 21), NIH-AARP (formerly American Association of Retired Persons) Diet and Health Study (22), or Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO; ref. 23; Supplementary Table S1). Briefly, AHS, IWHS, and NIH-AARP were prospective observational cohort studies; ATBC was a randomized, double-blind, placebo-controlled primary prevention trial; and PLCO was a randomized controlled screening trial. AHS, IWHS, NIH-AARP, and PLCO were all U.S.-based studies, and ATBC was conducted in Finland. ATBC and IWHS began during the 1980s, and the rest of the studies began during the 1990s. AHS only included pesticide applicators and their spouses, ATBC included males who currently smoked, IWHS enrolled a population-based sample of postmenopausal women, whereas NIH-AARP and PLCO were general population samples. All studies enrolled adults ages ≥50 to 55 years or older, except AHS, which included participants as young as 18 years old.

For this analysis, individuals diagnosed with cancer and had at least 30 days of follow-up were defined as survivors. Eligibility criteria for this pooling study included having completed a baseline questionnaire, no report of cancer at baseline (other than nonmelanoma skin cancer), and having ≥30 days of follow-up. Within this cohort, we identified subcohorts of survivors of the most common first primary malignancies occurring among adults ages 50 to 79 years: breast, prostate, colon/rectum, lung, or bladder cancer, and non-Hodgkin lymphoma (NHL; 24). The institutional review boards from the NCI and all participating institutions approved the use of these data, and all participants provided informed consent.

Study variables and data harmonization

At study entry, participants from each cohort completed a study-specific self-administered baseline questionnaire, which inquired about general demographics and various potential cancer risk factors. For all cohorts, a subset of participants also completed at least one risk factor questionnaire during follow-up.

To compare the populations among the cohorts, we obtained individual-level risk factor data, including age, sex, race, height and weight, and history of cigarette smoking for all eligible participants (Table 1). We also determined the number of individuals who completed at least one follow-up questionnaire following a first primary cancer diagnosis and before a second primary cancer diagnosis to establish feasibility of using these data. Because of the potential for reverse causality (i.e., a second cancer causing changes in risk factors such as weight or pain relief medication use), we further estimated the number of cancer cases who completed a questionnaire at least 12 months before their second cancer diagnosis.

<table>
<thead>
<tr>
<th>Cohort</th>
<th>N</th>
<th>Median age at entry (y)</th>
<th>% Male</th>
<th>% White</th>
<th>% Ever smoker</th>
<th>% Overweight/obese</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Former</td>
<td>Current</td>
</tr>
<tr>
<td>AHS</td>
<td>84,920</td>
<td>45.2</td>
<td>62.7</td>
<td>94.4</td>
<td>24.2</td>
<td>14.2</td>
</tr>
<tr>
<td>ATBC</td>
<td>29,107</td>
<td>57.1</td>
<td>100.0</td>
<td>100.0</td>
<td>—</td>
<td>100.0</td>
</tr>
<tr>
<td>IWHS</td>
<td>37,976</td>
<td>61.9</td>
<td>0.0</td>
<td>97.8</td>
<td>19.0</td>
<td>14.6</td>
</tr>
<tr>
<td>NIH-AARP</td>
<td>513,356</td>
<td>62.5</td>
<td>60.6</td>
<td>91.1</td>
<td>49.1</td>
<td>12.1</td>
</tr>
<tr>
<td>PLCO</td>
<td>142,921</td>
<td>62.0</td>
<td>50.4</td>
<td>88.3</td>
<td>43.2</td>
<td>10.7</td>
</tr>
<tr>
<td>Pooled</td>
<td>808,280</td>
<td>61.4</td>
<td>57.6</td>
<td>91.6</td>
<td>42.2</td>
<td>15.3</td>
</tr>
</tbody>
</table>

NOTE: Includes participants with a valid baseline questionnaire; no history of previous cancer (as reported on baseline questionnaire); and at least 30 days of follow-up. Percentages are total of all (including missings or unknowns). Not all percentages add up to total due to rounding.
Outcome ascertainment

Cancer incidence was obtained by linkage to cancer registries in the geographic areas of the cohort participant populations, except in PLCO and ATBC (before 1999), in which cancer diagnoses were self-reported and subsequently verified by medical record abstraction (Supplementary Table S1). A previous validation study, comparing cancer registry data with self-reports and medical records in NIH-AARP study, estimated that linkage validly identified approximately 90% of all incident cancers among study participants (25).

Because the cohorts used different versions of the International Classification of Diseases for Oncology (ICDO), we coded all cancers according to the Surveillance, Epidemiology, and End Results (SEER) Program Incidence Site Recode based on ICDO-3 (26, 27). Data were collected on the occurrence of all primary cancers (regardless of sequence) except for PLCO, which only confirmed the first occurrence of cancer in a specific topographical site (i.e., multiple primaries in the same organ were not recorded). To avoid misclassification of a local recurrence or metastatic lesion as a second primary cancer, we applied the SEER 2007 Multiple Primary and Histology Coding Rules to all reports of cancer (26). We further restricted our analyses to second primary malignancies diagnosed at a site other than the site of the original primary cancer.

Deaths in each cohort were ascertained through regular linkage to nationwide registry data.

Statistical analyses

To assess the validity of pooling populations of cancer survivors from different studies to study second cancer risk factors, we conducted three analyses. First, to evaluate representativeness of the populations of cancer survivors for each cohort, we compared the observed number of cancer diagnoses with that expected in the general population from which the cohort was drawn using a standardized incidence ratio (SIR; observed/expected; Supplementary Table S2). Expected numbers of cancers in each cohort were calculated by multiplying age-, sex-, race-, and calendar year–specific person-years at risk within the cohort by corresponding population-specific incidence rates. Second, to evaluate differences in cancer incidence rates between studies, we compared first primary cancer incidence across cohorts by calculating cohort-specific incidence rates standardized to the pooled cohort of cancer-free individuals at baseline based on age (5-year groups), sex, and race (white, non-white).

Third, because survival influences the likelihood of developing a second cancer, we explored potential differences in survival experiences of the cohorts after a first primary cancer diagnosis. We computed the 5-year relative survival, comparing the observed survival of each group of patients with that expected in the United States based on national life tables from the general population, matched by age, sex, race, and calendar year. For comparison, we also used data from 9 SEER registries to determine the relative survival of patients diagnosed with each first primary malignancy in the United States during 1985 to 2009 (28).

Because the occurrence of second cancers in the participating cohorts has not previously been reported, but is critical for assessing the feasibility of studying second cancer etiology in this setting, we determined the cumulative incidence of second cancers (excluding same-site primaries) diagnosed following each of the six most common primary sites in the pooled cohort. We also estimated the number of second primary cancer cases necessary to achieve 80% power at differing relative risks (RR) from 1.2 to 2.0, assuming 10% prevalence of second primary malignancies and binary risk factor prevalences of 10%, 20%, and 50%.

Because most prospective cohort studies were designed to evaluate first primary cancer etiology, they typically have either limited or no data on cancer treatment. However, PLCO collected information on initial treatment by medical record abstraction for cancers of the prostate, lung, and colorectum. For each of these cancer sites, we used the PLCO data to compare initial treatment by stage (and grade for prostate cancer) to assess the feasibility of using stage as a proxy for treatment.

Analyses were conducted using SAS, version 9.3 (SAS Institute Inc.) and SEER*Stat, version 8.0.2 (Surveillance Research Program, National Cancer Institute seer.cancer.gov/seerstat).

Results

Characteristics of the study cohorts

A total of 808,280 individuals from the five prospective studies were included in this study (Table 1). Overall, more than half (58%) of the participants were male. The median age at baseline was 61 years, and 92% were white. Fifteen percent of the pooled study population reported that they were current cigarette smokers, and 42% reported that they had smoked previously. Almost two thirds (63%) of the participants had a body mass index (BMI) ≥25 kg/m², with 41% categorized as overweight (25–29.9 kg/m²) and 21% as obese (≥30 kg/m²). The distribution of demographic characteristics among the cohorts was consistent with known differences in study design and population. For example, ATBC included Finnish male smokers, IWHS recruited only postmenopausal females, and PLCO had a slightly lower proportion of whites than the other cohorts as dedicated efforts were made to recruit minorities into this trial. NIH-AARP comprised 64% of the total population and, thus, had a strong influence on the pooled baseline characteristics.

Characteristics of cohort participants diagnosed with cancer

Among the 808,280 individuals, 136,848 (16.9%) were diagnosed with a first primary cancer during follow-up. Supplementary Figure S1 presents the baseline characteristics, overall and by cohort, for individuals diagnosed with one of the six most common malignancies, including
female breast cancer (N = 17,775), prostate cancer (N = 36,540), lung cancer (N = 17,373), colorectal cancer (N = 13,058), bladder cancer (N = 6,806), and NHL (N = 4,961). Comparisons among the cohorts within each cancer group primarily reflect the differences in the study populations. For example, patients in AHS were diagnosed with their first primary cancer at a younger age on average compared with the other studies because the study included younger individuals, whereas IWHS recruited older women and had the longest follow-up time, resulting in the highest median age at first cancer diagnosis. In contrast, comparisons across the survivor populations by type of first primary malignancy generally reflect differences in the etiology of the malignancies evaluated. For example, patients diagnosed with lung or bladder cancer were most likely to smoke, and patients diagnosed with colorectal cancer were slightly more likely to be overweight or obese.

Incidence of first primary malignancies
To assess how cancer incidence within each cohort compared with that in the general population from which the cohort was drawn, we computed population-specific SIRs for each of the selected first primary malignancies (Supplementary Table S2). Each of the cohorts generally had a higher incidence than expected of breast (except AHS) and prostate cancers. In contrast, incidence of lung cancer was lower than expected in all cohorts except ATBC, and incidence of colorectal cancer was lower than expected in all cohorts except IWHS and NIH-AARP.

To compare cancer incidence among the cohorts, we computed age-, sex-, and race-standardized incidence rates using the pooled cohort of cancer-free individuals as the referent (Supplementary Table S3). The standardized incidence rates varied most dramatically (>7-fold) for lung cancer, from 5.1 cases/10,000 person-years in IWHS to 37.6 in ATBC, and for bladder cancer, from 1.2 in IWHS to 10.5 in ATBC. Incidence of NHL varied approximately 3-fold among the cohorts, from 1.9 in ATBC to 6.3 in AHS. Modest variation (<2-fold) in incidence was observed among the cohorts for breast, prostate, and colorectal cancers.

Survival following first primary malignancy diagnosis
Regardless of first primary malignancy type, 5-year relative survival was consistently poorest for ATBC and best for PLCO (Fig. 1). Participants from cohorts other than ATBC generally had better survival than observed in patients with cancer with first primary malignancies reported to the SEER9 registry population. The greatest variability in 5-year relative survival was observed for colorectal cancer, which ranged from 51% in ATBC to 80% in both NIH-AARP and PLCO, and for NHL, which ranged from 39% in ATBC to 72% to 75% in AHS, NIH-AARP, and PLCO. The 5-year relative survival was very high for breast and prostate cancers, with little variation among the cohorts.

Second primary malignancies
Among the 96,513 individuals diagnosed with one of the six first primary malignancies in our analysis, 7,890 (8.2%) were diagnosed with a second primary malignancy at a site other than the site of their original primary cancer (Table 2). Cumulative incidence rates of second cancers (excluding same site) are presented in Fig. 2. Overall, cumulative incidence of second cancers was highest in those diagnosed with bladder cancer (21.7% at 10 years) and lowest in those diagnosed with lung cancer (4.9% at 10 years). Across cancer sites, cumulative incidence of second cancers was almost universally lowest in AHS.

Approximately 10% (837 of 7,890) of second primary cancer cases completed a follow-up questionnaire before their second diagnosis, of which 574 occurred ≥12 months before their second cancer diagnosis (Table 2). Figure 3 illustrates the number of second primary cancer cases necessary to achieve 80% power for a range of risk factor prevalences and relative risks. For example, to achieve adequate power to detect an RR of 1.8 with a risk factor prevalence of 10%, approximately 250 second cancer cases are needed; approximately 750 cases are needed to detect an RR of 1.4, under the same assumptions.

Cancer stage as a proxy for treatment
Because of the challenge of obtaining treatment data in prospective cohort studies, we evaluated the relation of initial treatment to stage at diagnosis for colon, rectal, non–small cell lung and prostate cancers, based on data abstracted from medical records in PLCO (Table 3). For colon and rectal cancers, initial treatment was closely related to stage: surgery alone was the most common treatment for individuals diagnosed with stage I or II colon cancer (98% and 70%, respectively) and stage I rectal cancer (68%), whereas those diagnosed with stage III or IV colon cancer were primarily treated with chemotherapy (82% and 72%, respectively), and those with stage II or III rectal cancers with combined chemotherapy and radiation (66% and 72%, respectively). In contrast, initial treatment for non–small cell lung and prostate cancers was highly variable by stage, with the exception of stage I lung cancer (72% of patients, surgery alone) and stage III prostate cancer (70% of patients, radiation therapy with or without surgery).

Discussion
Second cancers are a leading cause of morbidity and mortality among cancer survivors, but few studies have examined the role of risk factors other than treatment in the etiology of second cancers. The study of second cancers involves particular methodologic concerns not inherent to most epidemiologic investigations, which may, in large part, explain the paucity of studies conducted to date on nontreatment-related risk factors. Pooling prospective cohorts offers an opportunity to leverage the substantial existing resources to yield new clinical, public health, and
scientific benefits, but we highlight five key methodologic issues that warrant consideration in future studies of second cancer etiology in this setting.

**Study design and population differences**

Differences in population selection criteria and the very nature of voluntary participation often result in cohorts...
and, importantly, the risk factors that affect both cancer.

Table 2. Number of second cancer events and median follow-up time [median person years at risk (PYR) and interquartile range (IQR)] between first and second primary cancer diagnosis

<table>
<thead>
<tr>
<th>Primary cancer site</th>
<th>All second cancers</th>
<th>All second cancer cases with a follow-up questionnaire before second diagnosis</th>
<th>All second cancer cases with a follow-up questionnaire at least 12 months before second diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Number of cases</td>
<td>PYR (IQR)</td>
</tr>
<tr>
<td>Breast</td>
<td>17,775</td>
<td>1,186</td>
<td>3.50 (5.09)</td>
</tr>
<tr>
<td>Prostate</td>
<td>36,540</td>
<td>3,300</td>
<td>2.75 (4.08)</td>
</tr>
<tr>
<td>Lung</td>
<td>17,373</td>
<td>646</td>
<td>1.17 (3.50)</td>
</tr>
<tr>
<td>Colorectal</td>
<td>13,058</td>
<td>1,206</td>
<td>2.33 (4.33)</td>
</tr>
<tr>
<td>Bladder</td>
<td>6,806</td>
<td>1,108</td>
<td>1.42 (3.91)</td>
</tr>
<tr>
<td>NHL</td>
<td>4,961</td>
<td>444</td>
<td>2.50 (3.59)</td>
</tr>
</tbody>
</table>

risk and mortality. For example, cigarette smoking may seem protective for developing a second primary cancer in patients with advanced stage lung cancer because these individuals have relatively short survival and do not live long enough to develop a second primary malignancy. Conversely, the lower cumulative incidence of second cancers in AHS compared with the other cohorts is likely explained by younger participants in AHS, who are less likely to develop cancer, than the older participants in the other cohorts. Therefore, when pooling data, it is important to carefully consider the design and characteristics of cohorts in the analysis and interpretation of findings.

**Cancer ascertainment**

In the absence of centralized medical record review, cancer ascertainment via linkage to cancer registries is optimal for studies of second cancer etiology because of the use of consistent methodology to identify all primary cancer diagnoses, allowing for distinction between recurrent or metastatic disease and true second primary diagnoses. However, some differences may exist in the definition of multiple primary malignancies occurring in the same organ, particularly in European versus U.S. registries. Studies such as PLCO that ascertain new cancer diagnoses by self-report can be included in studies of second cancer etiology, but use of additional sources to identify cancers such as death certificate and subsequent medical record confirmation is essential, and study-specific rules—such as PLCO confirmation of only one malignancy in a specific topographical site—must be taken into account. To address these specific issues, we uniformly applied SEER rules for defining multiple same-site primaries to all reports of cancer and excluded same-site second primary cancers (26).

**Exposure assessment**

A key strength of prospective cohort studies is the availability of detailed and diverse patient exposure data, directly contrasting most previous settings for second
cancers research. Using baseline questionnaire data, we were able to harmonize exposure to several key cancer risk factors. The current effort and others (e.g., refs. 31–33) robustly demonstrate the ability to pool data across studies with different questionnaire formats for many risk factors. However, the use of exposure data may be limited by the format and level of detail, so pooled studies must use either the simplest exposure measures available across all cohorts, or restrict analyses to the cohorts that collected detailed information. Nonetheless, as more cohorts contribute to this pooling effort, the possibility exists to address numerous other risk factors in relation to...
second cancer risk including, but not limited to, reproductive factors, menopausal hormone use, alcohol consumption, physical activity, and dietary factors.

For studies of second cancer etiology, the time period of exposure ascertained by each questionnaire and the availability of follow-up questionnaires also are key considerations. All five studies included in this pooled analysis had at least one follow-up questionnaire, but the frequency of exposure ascertainment and the data collected at each time point varied widely. In addition, questionnaires were inevitably administered over a range of time before the diagnosis of a first and second primary malignancy, likely resulting in varying exposure misclassification for different individuals within a cohort. Therefore, consideration of this exposure misclassification and possible inclusion of time-dependent covariates are warranted in studies of second cancer etiology in this setting. Attention should also be given to the possibility of selection bias, i.e., individuals who survive long enough to complete the follow-up questionnaire, and choose to complete it, may differ from those who do not (e.g., younger, nonsmokers).

**Statistical power**

Collaboration of multiple studies is critical to have sufficient statistical power to investigate second cancer risk factors, particularly if the first cancer, second cancer, or exposure of interest is rare. Even with the five studies included in this pooled analysis, including over 800,000 cancer-free individuals at baseline, the cohorts of survivors of the most common first primary malignancies were relatively modest in size, ranging from 4,961 patients diagnosed with NHL to 36,540 patients with prostate cancer. Even with almost 18,000 incident breast cancers, there is insufficient statistical power for site-specific

### Table 3. Initial treatment, by stage, for selected first primary cancers in the PLCO Cancer Screening Trial

<table>
<thead>
<tr>
<th>Colon cancer</th>
<th>Surgery only</th>
<th>Chemotherapy (without radiation)</th>
<th>Radiation (without chemotherapy)</th>
<th>Chemotherapy and radiation</th>
<th>No treatment/unknown/other</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>474</td>
<td>464 (98%)</td>
<td>6 (1%)</td>
<td>0 (0%)</td>
<td>4 (1%)</td>
</tr>
<tr>
<td>II</td>
<td>409</td>
<td>286 (70%)</td>
<td>112 (27%)</td>
<td>2 (0%)</td>
<td>2 (0%)</td>
</tr>
<tr>
<td>III</td>
<td>407</td>
<td>47 (12%)</td>
<td>334 (82%)</td>
<td>0 (0%)</td>
<td>24 (6%)</td>
</tr>
<tr>
<td>IV</td>
<td>263</td>
<td>42 (16%)</td>
<td>189 (72%)</td>
<td>4 (2%)</td>
<td>11 (4%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rectal cancer</th>
<th>Surgery only</th>
<th>Chemotherapy (without radiation)</th>
<th>Radiation (without chemotherapy)</th>
<th>Chemotherapy and radiation</th>
<th>No treatment/unknown/other</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>164</td>
<td>112 (68%)</td>
<td>4 (2%)</td>
<td>10 (6%)</td>
<td>38 (23%)</td>
</tr>
<tr>
<td>II</td>
<td>58</td>
<td>16 (28%)</td>
<td>2 (3%)</td>
<td>2 (3%)</td>
<td>38 (66%)</td>
</tr>
<tr>
<td>III</td>
<td>79</td>
<td>9 (11%)</td>
<td>11 (14%)</td>
<td>3 (4%)</td>
<td>57 (72%)</td>
</tr>
<tr>
<td>IV</td>
<td>30</td>
<td>1 (3%)</td>
<td>13 (43%)</td>
<td>1 (3%)</td>
<td>10 (33%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non–small cell lung cancer</th>
<th>Surgery only</th>
<th>Chemotherapy (without radiation)</th>
<th>Radiation (without chemotherapy)</th>
<th>Chemotherapy and radiation</th>
<th>No treatment/unknown/other</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>676</td>
<td>487 (72%)</td>
<td>63 (9%)</td>
<td>60 (9%)</td>
<td>23 (3%)</td>
</tr>
<tr>
<td>II</td>
<td>169</td>
<td>52 (31%)</td>
<td>31 (18%)</td>
<td>31 (18%)</td>
<td>42 (25%)</td>
</tr>
<tr>
<td>III</td>
<td>646</td>
<td>59 (9%)</td>
<td>116 (18%)</td>
<td>89 (14%)</td>
<td>305 (47%)</td>
</tr>
<tr>
<td>IV</td>
<td>880</td>
<td>29 (3%)</td>
<td>286 (33%)</td>
<td>152 (17%)</td>
<td>231 (26%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prostate cancer</th>
<th>Surgery only</th>
<th>Chemotherapy (without radiation)</th>
<th>Radiation (without chemotherapy)</th>
<th>Chemotherapy and radiation</th>
<th>No treatment/unknown/other</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>31</td>
<td>18 (58%)</td>
<td>0 (0%)</td>
<td>5 (16%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>II</td>
<td>7,380</td>
<td>3,053 (41%)</td>
<td>1 (0%)</td>
<td>3,180 (43%)</td>
<td>2 (0%)</td>
</tr>
<tr>
<td>III</td>
<td>118</td>
<td>21 (18%)</td>
<td>0 (0%)</td>
<td>83 (70%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>IV</td>
<td>197</td>
<td>19 (10%)</td>
<td>1 (0%)</td>
<td>41 (21%)</td>
<td>1 (0%)</td>
</tr>
</tbody>
</table>
association studies of the most commonly diagnosed second cancers (Fig. 3). For analyses of rare malignancies, and studies of low prevalence risk factors, additional cohorts are required to yield case numbers large enough for adequately powered studies. For example, with 1,186 total second cancers after a primary breast cancer, we have adequate power to detect an OR of 1.3 with a risk factor prevalence of 10%; however, if we limited our analysis to second colorectal cancers, we could only detect an OR of 2.0 or greater.

Studies to investigate changes in risk factor behaviors following a cancer diagnosis, such as smoking cessation or weight change, are further restricted to those with a follow-up questionnaire before their second diagnosis, resulting in a marked loss of cases. These studies, and those requiring stratification of patient populations (e.g., by disease subtype, stage, or race), are feasible, but further emphasize the need for a pooled approach.

Clinical data

Because most prospective cohort studies were designed to investigate cancer etiology, clinical data are generally quite limited or not available at all. Consequently, studies of treatment-related second cancer risks are unlikely to be successful in this setting. Investigations of nontreatment risk factors for second cancers would ideally control for initial cancer treatment to account for any potential relationships between treatment and exposure (e.g., chest radiotherapy, history of cigarette smoking, and potential risk of second lung cancer). However, in the absence of such data, when treatment is highly standardized and correlated with stage, stage may be a reasonable proxy for broad classifications of treatment (e.g., received radiation therapy) for select investigations. For example, we observed that stage was a good indicator for treatment of both colon and rectal cancers during the treatment era covered by this study. For other malignancies, stage at diagnosis may help identify patient populations with relatively homogeneous treatment, such as stage I non-small cell lung cancer and stage III prostate cancer. For other sites, such as breast, for which treatment regimens are highly variable, this approach is not valid. The lack of detailed treatment data (e.g., dose, frequency, and duration) likely precludes investigation of potential second cancer risk factors that may be correlated with treatment exposures.

Conclusions

As the population of cancer survivors continues to increase, improved understanding of survivors’ long-term health is of both clinical and public health importance. The descriptive data presented here are intended to demonstrate that pooling existing data from prospective epidemiologic cohorts seems to be a feasible approach to address important questions related to second cancer etiology. Although inherent challenges may restrict the range of research questions that may be pursued in this setting, careful utilization of existing resources may provide opportunities to identify patients for targeted cancer prevention efforts or increased surveillance, as well as new insights into carcinogenesis. Such collaborative efforts will be facilitated by the dramatic increase in epidemiologic consortia, particularly the National Cancer Institute Cohort Consortium (34), which can provide infrastructure for identifying collaborators for new research projects as well as experience in data harmonization.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Disclaimer

The views expressed herein are solely those of the authors and do not necessarily reflect those of the Florida Cancer Data System (FCDS) or Florida Department of Health (FDOH). The Pennsylvania Department of Health specifically disclaims responsibility for any analyses, interpretations, or conclusions.

Authors’ Contributions

Conception and design: A. Black, T.M. Gibson, Y. Park, K. Robien, J.F. Fraumeni, P. Hartige, J.R. Cerhan, A. Zeleniuch-Jacquotte, A. Berrington de Gonzalez, L.M. Morton

Development of methodology: A. Black, T.M. Gibson, K. Robien, J.N. Sampson, L.M. Morton

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A. Black, Y. Park, K. Robien, D. Albanese, S.J. Weinstein, L.E. Beane Freeman, G. Andreotti, M.P. Purdue, R.N. Hoover, J.R. Cerhan, L.M. Morton

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A. Black, T.M. Gibson, M.S. Shiels, Y. Park, K. Robien, S.J. Weinstein, J.F. Fraumeni, R.E. Curtis, J.N. Sampson, A. Berrington de Gonzalez, L.M. Morton

Writing, review, and/or revision of the manuscript: A. Black, T.M. Gibson, M.S. Shiels, Y. Park, K. Robien, D. Albanese, S.J. Weinstein, L.E. Beane Freeman, G. Andreotti, M.P. Purdue, J.F. Fraumeni, P. Hartige, M.A. Tucker, R.N. Hoover, J.R. Cerhan, A. Zeleniuch-Jacquotte, R.E. Curtis, J. Elena, A. Berrington de Gonzalez, L.M. Morton

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A. Black, Y. Park, K. Robien, L.E. Beane Freeman, L.M. Morton

Study supervision: A. Black, K. Robien, M.A. Tucker, L.M. Morton

Other (final approval of the manuscript): J.F. Fraumeni

Acknowledgments

NIH-AARP ascertained cancer incidence from individual cancer registries and would like to acknowledge their contributions as follows: cancer incidence data from the Atlanta metropolitan area were collected by the Georgia Center for Cancer Statistics, Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia. Cancer incidence data from California were collected by the California Cancer Registry, California Department of Public Health’s Cancer Surveillance and Research Branch, Sacramento, California. Cancer incidence data from the Detroit metropolitan area were collected by the Michigan Cancer Surveillance Program, Community Health Administration, Lansing, Michigan. The Florida cancer incidence data used in this report were collected by the Florida Cancer Data System (Miami, Florida) under contract with the Florida Department of Health, Tallahassee, Florida. Cancer incidence data from Louisiana were collected by the Louisiana Tumor Registry, Louisiana State University Health Sciences Center School of Public Health, New Orleans, Louisiana. Cancer incidence data from New Jersey were collected by the New Jersey State Cancer Registry, Cancer Epidemiology Services, New Jersey State Department of Health, Trenton, New Jersey. Cancer incidence data from North Carolina were collected by the North Carolina Central Cancer Registry, Raleigh, North Carolina. Cancer incidence data from Pennsylvania were supplied by the Division of Health Statistics and Research, Pennsylvania Department of Health, Harrisburg, Pennsylvania. Cancer incidence data from Arizona were collected by the Arizona Cancer Registry, Division of Public Health Services, Arizona Department of Health Services, Phoenix, Arizona. Cancer incidence data from Texas were collected by the Texas Cancer Registry, Cancer Epidemiology and Surveillance Branch, Texas Department of State.
Health Services, Austin, Texas. Cancer incidence data from Nevada were collected by the Nevada Central Cancer Registry, State Health Division, State of Nevada Department of Health and Human Services, Las Vegas, Nevada.

Grant Support
This research was supported by the Intramural Program of the NCI, NIH. In addition, the ATBC study was supported by U.S. Public Health Service contracts N01-CN-45165, N01-RC-45035, N01-RC-37004, and HHSN261201000056C and from the NCI, Department of Health and Human Services, the AHS study also was supported (in part) by the National Institute of Environmental Health Sciences (Z01-ES049030) and National Cancer Institute (Z01-CP011019), and the IWHS was supported by a grant RO1-CA39742 from the NCI (D. Lazovich, principal investigator).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 14, 2014; revised May 8, 2014; accepted May 9, 2014; published OnlineFirst May 15, 2014.

References


Downloaded from cebp.aacrjournals.org on June 20, 2017. © 2014 American Association for Cancer Research.
Cancer Epidemiology, Biomarkers & Prevention

Pooling Prospective Studies to Investigate the Etiology of Second Cancers

Amanda Black, Todd M. Gibson, Meredith S. Shiels, et al.

Cancer Epidemiol Biomarkers Prev  Published OnlineFirst May 15, 2014.

Updated version

Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-14-0191

Supplementary Material

Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2014/05/15/1055-9965.EPI-14-0191.DC1

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, contact the AACR Publications Department at permissions@aacr.org.