Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2014 August ; 23(8): 1598–1608. doi:
10.1158/1055-9965.EPI-14-0191.

Pooling prospective studies to investigate the etiology of second cancers

Amanda Black¹, Todd M. Gibson¹, Meredith S. Shiels¹, Yikyung Park¹, Kim Robien², Demetrius Albanes¹, Stephanie J. Weinstein¹, Laura E. Beane Freeman¹, Gabriella Andreotti¹, Mark P. Purdue¹, Joseph F. Fraumeni¹, Patricia Hartge¹, Margaret A. Tucker¹, Robert N. Hoover¹, James R. Cerhan³, Anne Zeleniuch-Jacquotte⁴, Rochelle E. Curtis¹, Joanne Elena⁵, Joshua N. Sampson¹, Amy Berrington de Gonzalez¹, and Lindsay M. Morton¹

¹Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Rockville, MD

²Department of Epidemiology and Biostatistics, George Washington University School of Public Health and Health Services, Washington, DC

³Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, MN

⁴Division of Epidemiology, Department of Population Health, New York University School of Medicine, New York, NY

⁵Division of Cancer Control and Population Sciences, National Cancer Institute, NIH, DHHS, Rockville, MD

Abstract

Background—With over 13 million cancer survivors in the United States today, second cancers are of rapidly growing importance. However, data on non-treatment 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 1985–2009. 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). 7,890 second primary cancers (excluding original primary site) were identified, yielding sufficient statistical power (80%) for detecting modest associations with

Corresponding author: Lindsay M. Morton, PhD; Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS; 9609 Medical Center Drive, Room 7E-454, MSC 9778; Rockville, MD 20892; +1-240-276-7377 (phone), +1-240-276-7840 (fax), mortonli@mail.nih.gov.

Conflicts of interest: The authors have no conflicts of interest to disclose.

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 non-treatment risk factors for second primary cancers appears feasible but there are important methodological issues - some of which are barriers to specific research questions - that require special attention.

Impact—Increased understanding of non-treatment risk factors for second cancers will provide valuable prevention and surveillance information.

INTRODUCTION

The prevalence of cancer survivors in the United States (US) 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 US 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),(3–7) clinical trials (typically smaller with rich and comprehensive treatment data but limited availability of pre-diagnostic risk factors)(8–11), 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, 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),(19) Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC),(20) Iowa Women's Health Study (IWHS),(21) NIH-AARP (formerly American Association of Retired Persons) Diet and Health Study (NIH-AARP),(22) or Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO)(23) (Supplemental Table 1). 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 US-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 post-menopausal women, whereas NIH-AARP and PLCO were general population samples. All studies enrolled adults aged 50–55 years, 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 non-melanoma skin cancer), and having 30 days of follow-up. Within this cohort, we identified sub-cohorts of survivors of the most common first primary malignancies occurring among adults aged 50–79 years: breast, prostate, colon/rectum, lung, or bladder cancer, and non-Hodgkin lymphoma (NHL).(24) The institutional review boards from the National Cancer Institute 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 prior to 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 prior to their second cancer diagnosis.

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 (prior to 1999), in which cancer diagnoses were self-reported and subsequently verified by medical record abstraction (Supplemental Table 1). 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 (ICD-O), we coded all cancers according to the Surveillance, Epidemiology, and End Results (SEER) Program Incidence Site Recode based on ICD-O-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

In order 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 to that expected in the general population from which the cohort was drawn using a standardized incidence ratio (SIR; observed/expected) (Supplemental Table 2). 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 five-year relative survival, comparing the observed survival of each group of patients to that expected in the US 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 US during 1985–2009.(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 1.2–2.0, 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., Cary, NC)) 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) 25kg/m^2 , with 41% categorized as overweight ($25\text{--}29.9\text{ kg/m}^2$) and 21% as obese ($\geq 30\text{kg/m}^2$). 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. Supplemental Figure 1 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 to that in the general population from which the cohort was drawn, we computed population-specific SIRs for each of the selected first primary malignancies (Supplemental Table 2). 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 (Supplemental Table 3). 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, five-year relative survival was consistently poorest for ATBC and best for PLCO (Figure 1). Participants from cohorts other than ATBC generally had better survival than observed in cancer patients with first primary malignancies reported to the SEER9 registry population. The greatest variability in five-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 72–75% in AHS, NIH-AARP, and PLCO. The five-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 Figure 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 prior to their second diagnosis, of which 574 occurred 12 months prior to 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 a 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 a 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 methodological concerns not inherent to most epidemiological investigations, which may, in large part, explain the paucity of studies conducted to date on non-treatment-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 methodological 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 comprised of individuals who are healthier than the populations from which they are drawn (the 'healthy cohort effect'), with lower prevalence of key cancer risk factors.(29) Thus, understanding the context within which the data from individual studies were collected is critical to facilitate identification of potential sources of bias and avoid inaccurate interpretation of findings.

Specific differences in study populations between the cohorts in this analysis – which were selected by age, sex, and other exposures – suggest that particular care must be given to evaluate inter-study heterogeneity in studies of second cancer risk factors. For example, as expected, ATBC had higher incidence of first primary lung and bladder cancers than in the Finnish population and in the other cohorts, and poorer survival, because the cohort included only male smokers in ATBC. Differences in risk factor prevalence such as these should not preclude pooling data since they should not affect internal validity, but they do have the potential to influence the risk estimates of second cancers.

Differences in study designs also impacted cancer incidence across cohorts. For example, the overdiagnosis of prostate cancer due to screening with the prostate-specific antigen (PSA) test(30) is likely responsible for the substantially higher than expected incidence rates of prostate cancer in PLCO (a prostate cancer screening trial) than in the general population and compared to the other cohorts. In contrast, ATBC had the lowest prostate cancer incidence, likely reflecting the absence of PSA screening for prostate cancer in Finland.

We observed substantial differences in survival between the cohorts that warrant attention as individuals must live sufficiently long enough to be diagnosed with a second primary cancer, and some risk factors may be associated with survival. The differences in cumulative incidence rates of second cancers, both between cohorts and between site of primary diagnosis, illustrate the complex relationships between risk factors and cancer and, importantly, the risk factors that affect both cancer risk and mortality. For example, cigarette smoking may appear 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 US 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’s 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.(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; 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 prior to 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, non-smokers).

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 with prostate cancer. Even with almost 18,000 incident breast cancers, there is insufficient statistical power for site-specific association studies of the most commonly diagnosed second cancers (Figure 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 prior to 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 non-treatment risk factors for second cancers would ideally control for initial cancer treatment in order to account for any potential relationships between treatment

and exposure (e.g., chest radiotherapy, history of cigarette smoking, and potential risk of 2nd 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, 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 appears 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support: This research was supported by the Intramural Program of the National Cancer Institute, National Institutes of Health. Additionally, the Alpha-Tocopherol Beta Carotene (ATBC) Study was supported by U.S. Public Health Service contracts N01-CN-45165, N01-RC-45035, N01-RC-37004 and HHSN261201000006C and from the National Cancer Institute, Department of Health and Human Services and the Iowa Women's Health Study (IWHS) was supported by a grant RO1-CA39742 from the National Cancer Institute (D Lazovich, PI).

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. The views expressed herein are solely those of the authors and do not necessarily reflect those of the FCDC or FDOH. 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. The Pennsylvania Department of Health specifically disclaims responsibility for any analyses, interpretations or conclusions. 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.

References

1. Howlader, N.; Noone, A.; Krapcho, M.; Neyman, N.; Aminou, R.; Waldron, W., et al. SEER Cancer Statistics Review, 1975–2008. National Cancer Institute; Bethesda, MD: 2011. Available from: http://seer.cancer.gov/csr/1975_2008/
2. Berrington de Gonzalez A, Curtis RE, Kry SF, Gilbert E, Lamart S, Berg CD, et al. Proportion of second cancers attributable to radiotherapy treatment in adults: a cohort study in the US SEER cancer registries. *The Lancet Oncology*. 2011; 12:353–60. [PubMed: 21454129]
3. Dores GM, Anderson WF, Beane Freeman LE, Fraumeni JF Jr, Curtis RE. Risk of breast cancer according to clinicopathologic features among long-term survivors of Hodgkin's lymphoma treated with radiotherapy. *British Journal of Cancer*. 2010; 103:1081–4. [PubMed: 20842115]
4. Lonn S, Gilbert ES, Ron E, Smith SA, Stovall M, Curtis RE. Comparison of second cancer risks from brachytherapy and external beam therapy after uterine corpus cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2010; 19:464–74.
5. Morton LM, Curtis RE, Linet MS, Bluhm EC, Tucker MA, Caporaso N, et al. Second malignancy risks after non-Hodgkin's lymphoma and chronic lymphocytic leukemia: differences by lymphoma subtype. *Journal of Clinical Oncology*. 2010; 28:4935–44. [PubMed: 20940199]
6. Boukheris H, Ron E, Dores GM, Stovall M, Smith SA, Curtis RE. Risk of radiation-related salivary gland carcinomas among survivors of Hodgkin lymphoma: a population-based analysis. *Cancer*. 2008; 113:3153–9. [PubMed: 18823043]
7. Curtis, REFD.; Ron, E.; Ries, LAG.; Hacker, DG.; Edwards, BK.; Tucker, MA., et al. NIH Publ. No. 05-5302. National Cancer Institute; Bethesda, MD: 2006. New Malignancies Among Cancer Survivors: SEER Cancer Registries, 1973–2000.
8. Fornander T, Rutqvist LE, Cedermark B, Glas U, Mattsson A, Silfversward C, et al. Adjuvant tamoxifen in early breast cancer: occurrence of new primary cancers. *Lancet*. 1989; 1:117–20. [PubMed: 2563046]
9. Veronesi U, Mariani L, Decensi A, Formelli F, Camerini T, Miceli R, et al. Fifteen-year results of a randomized phase III trial of fenretinide to prevent second breast cancer. *Annals of oncology*. 2006; 17:1065–71. [PubMed: 16675486]
10. Lippman SM, Lee JJ, Karp DD, Vokes EE, Benner SE, Goodman GE, et al. Randomized phase III intergroup trial of isotretinoin to prevent second primary tumors in stage I non-small-cell lung cancer. *Journal of the National Cancer Institute*. 2001; 93:605–18. [PubMed: 11309437]
11. Khuri FR, Lee JJ, Lippman SM, Kim ES, Cooper JS, Benner SE, et al. Randomized phase III trial of low-dose isotretinoin for prevention of second primary tumors in stage I and II head and neck cancer patients. *Journal of the National Cancer Institute*. 2006; 98:441–50. [PubMed: 16595780]
12. Travis LB, Gospodarowicz M, Curtis RE, Clarke EA, Andersson M, Glimelius B, et al. Lung cancer following chemotherapy and radiotherapy for Hodgkin's disease. *Journal of the National Cancer Institute*. 2002; 94:182–92. [PubMed: 11830608]
13. Travis LB, Hill DA, Dores GM, Gospodarowicz M, van Leeuwen FE, Holowaty E, et al. Breast cancer following radiotherapy and chemotherapy among young women with Hodgkin disease. *JAMA*. 2003; 290:465–75. [PubMed: 12876089]
14. Travis LB, Curtis RE, Glimelius B, Holowaty EJ, Van Leeuwen FE, Lynch CF, et al. Bladder and kidney cancer following cyclophosphamide therapy for non-Hodgkin's lymphoma. *Journal of the National Cancer Institute*. 1995; 87:524–30. [PubMed: 7707439]

15. van den Belt-Dusebout AW, Aleman BM, Besseling G, de Bruin ML, Hauptmann M, van 't Veer MB, et al. Roles of radiation dose and chemotherapy in the etiology of stomach cancer as a second malignancy. *International journal of radiation oncology, biology, physics*. 2009; 75:1420–9.
16. Travis LB, Andersson M, Gospodarowicz M, van Leeuwen FE, Bergfeldt K, Lynch CF, et al. Treatment-associated leukemia following testicular cancer. *Journal of the National Cancer Institute*. 2000; 92:1165–71. [PubMed: 10904090]
17. Elena JW, Travis LB, Simonds NI, Ambrosone CB, Ballard-Barbash R, Bhatia S, et al. Leveraging epidemiology and clinical studies of cancer outcomes: recommendations and opportunities for translational research. *Journal of the National Cancer Institute*. 2013; 105:85–94. [PubMed: 23197494]
18. Berrington de Gonzalez A, Morton LM. Converting epidemiologic studies of cancer etiology to survivorship studies: approaches and challenges. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2012; 21:875–80.
19. Alavanja MC, Sandler DP, McMaster SB, Zahm SH, McDonnell CJ, Lynch CF, et al. The Agricultural Health Study. *Environmental health perspectives*. 1996; 104:362–9. [PubMed: 8732939]
20. The ATBC Cancer Prevention Study Group. The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods participant characteristics, and compliance. *Annals of Epidemiology*. 1994; 4:1–10. [PubMed: 8205268]
21. Folsom AR, Kaye SA, Sellers TA, Hong CP, Cerhan JR, Potter JD, et al. Body fat distribution and 5-year risk of death in older women. *JAMA*. 1993; 269:483–7. [PubMed: 8419667]
22. Schatzkin A, Subar AF, Thompson FE, Harlan LC, Tangrea J, Hollenbeck AR, et al. Design and serendipity in establishing a large cohort with wide dietary intake distributions : the National Institutes of Health-American Association of Retired Persons Diet and Health Study. *American Journal of Epidemiology*. 2001; 154:1119–25. [PubMed: 11744517]
23. Prorok PC, Andriole GL, Bresalier RS, Buys SS, Chia D, Crawford ED, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Controlled Clinical Trials*. 2000; 21:273S–309S. [PubMed: 11189684]
24. Society AC. American Cancer Society: Cancer Facts and Figures 2013. Atlanta, GA: 2013. [September 24, 2013]. Available from: <http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2013/index>
25. Michaud DH, Hermansen S, Hermansen S, Leitzmann M, Harlan LC, Kipnis V, et al. Comparison of cancer registry case ascertainment with SEER-estimates and self-reporting in a subset of the NIH-AARP Diet and Health Study. *Journal of Registry Management*. 2005; 32:70–5.
26. Howlader, N.; Krapcho, M.; Neyman, N.; Aminou, R.; Waldron, W.; Altekruse, SF., et al. SEER Cancer Statistics Review, 1975–2009. National Cancer Institute; Bethesda, MD: 2012. Available from: http://seer.cancer.gov/csr/1975_2009_pops09/, based on November 2011 SEER data submission, posted to the SEER web site, April 2012
27. Fritz, AGPC.; Jack, A.; Shanmugaratnam, K.; Sobin, L.; Parkin, DM.; Whelan, S. International classification of diseases for oncology (ICD-O), 3rd edition. Geneva: World Health Organization; 2000.
28. Surveillance Epidemiology and End Results, SEER Registry Groupings for Analyses. Available from: <http://seer.cancer.gov/registries/terms.html>
29. Pinsky PF, Miller A, Kramer BS, Church T, Reding D, Prorok P, et al. Evidence of a healthy volunteer effect in the prostate, lung, colorectal, and ovarian cancer screening trial. *American Journal of Epidemiology*. 2007; 165:874–81. [PubMed: 17244633]
30. Welch HG, Black WC. Overdiagnosis in cancer. *Journal of the National Cancer Institute*. 2010; 102:605–13. [PubMed: 20413742]
31. Berrington de Gonzalez A, Hartge P, Cerhan JR, Flint AJ, Hannan L, MacInnis RJ, et al. Body-mass index and mortality among 1.46 million white adults. *The New England Journal of Medicine*. 2010; 363:2211–9. [PubMed: 21121834]

32. Elena JW, Stepkowski E, Yu K, Hartge P, Tobias GS, Brotzman MJ, et al. Diabetes and risk of pancreatic cancer: a pooled analysis from the pancreatic cancer cohort consortium. *Cancer Causes & Control* : CCC. 2013; 24:13–25. [PubMed: 23112111]
33. Moore SC, Patel AV, Matthews CE, Berrington de Gonzalez A, Park Y, Katki HA, et al. Leisure time physical activity of moderate to vigorous intensity and mortality: a large pooled cohort analysis. *PLoS Medicine*. 2012; 9:e1001335. [PubMed: 23139642]
34. National Cancer Institute Cohort Consortium. Available from: <http://epi.grants.cancer.gov/Consortia/cohort.html>

Figure 1.
Relative survival following first primary cancer diagnosis

Figure 2.
Cumulative incidence of second cancers (excluding same site) following first primary cancer diagnosis

Figure 3.

Power curves showing the number of second cancer cases (y-axis) required to achieve 80% power for relative risks (RR) ranging from 1.2 to 2.0 (x-axis) for a binary risk factor prevalence of 10, 20 and 50%

Table 1

Baseline characteristics of participants for each prospective cohort study

Cohort	N	Median age at entry (years)	% Male	% White	% Ever smoker		% Overweight/obese	
					Former	Current	Overweight	Obese
AHS	84,920	45.2	62.7	94.4	24.2	14.2	32.1	16.0
ATBC	29,107	57.1	100.0	100.0	-	100.0	46.0	15.3
IWHS	37,976	61.9	0.0	97.8	19.0	14.6	36.9	23.3
NIH-AARP	513,356	62.5	60.6	91.1	49.1	12.1	41.6	21.3
PLCO	142,921	62.0	50.4	88.3	43.2	10.7	41.9	23.6
Pooled	808,280	61.4	57.6	91.6	42.2	15.3	40.6	21.0

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

AHS- Agricultural Health Study; ATBC-Alpa-Tocopherol Beta Carotene Study; IWHS-Iowa Women's Health Study; NIH-AARP-National Institutes of Health-(formerly) American Association of Retired Persons; PLCO-Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial

Table 2

Number of second cancer events and median follow-up time (median person years at risk (PYR) and interquartile range (IQR)) between 1st and 2nd primary cancer diagnosis

		Number of second primary cancer cases and follow-up time				
Primary Cancer Site		All second cancers		All second cancer cases with a follow-up questionnaire prior to second diagnosis		All second cancer cases with a follow-up questionnaire at least 12 months prior to second diagnosis
Site	N	N Cases	PYR (IQR)	N Cases	PYR (IQR)	Cases
Breast	17,775	1,186	3.50 (5.09)	165	6.67 (4.84)	121
Prostate	36,540	3,300	2.75 (4.08)	374	5.66 (4.83)	249
Lung	17,373	646	1.17 (3.50)	38	3.42 (4.83)	23
Colorectal	13,058	1,206	2.33 (4.33)	106	6.13 (5.83)	68
Bladder	6,806	1,108	1.42 (3.91)	102	5.29 (5.08)	79
NHL	4,961	444	2.50 (3.59)	52	5.25 (5.96)	34

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

N	Initial Treatment				No Treatment/Unknown/Other
	Surgery Only	Chemotherapy (w/o radiation)	Radiation (w/o chemotherapy)	Chemotherapy and Radiation	
Colon cancer					
I 474	464 (98%)	6 (1%)	0 (0%)	0 (0%)	4 (1%)
II 409	286 (70%)	112 (27%)	2 (0%)	7 (2%)	2 (0%)
III 407	47 (12%)	334 (82%)	0 (0%)	24 (6%)	2 (0%)
IV 263	42 (16%)	189 (72%)	4 (2%)	11 (4%)	17 (6%)
Rectal cancer					
I 164	112 (68%)	4 (2%)	10 (6%)	38 (23%)	0 (0%)
II 58	16 (28%)	2 (3%)	2 (3%)	38 (66%)	0 (0%)
III 79	9 (11%)	11 (14%)	3 (4%)	57 (72%)	0 (0%)
IV 30	1 (3%)	13 (43%)	1 (3%)	10 (33%)	5 (17%)
Non-small cell lung cancer					
I 676	487 (72%)	63 (9%)	60 (9%)	23 (3%)	43 (6%)
II 169	52 (31%)	31 (18%)	31 (18%)	42 (25%)	13 (8%)
III 646	59 (9%)	116 (18%)	89 (14%)	305 (47%)	87 (13%)
IV 880	29 (3%)	286 (33%)	152 (17%)	231 (26%)	182 (21%)
Prostate cancer					
I 31	18 (58%)	0 (0%)	5 (16%)	0 (0%)	8 (26%)
II 7,380	3,053 (41%)	1 (0%)	3,180 (43%)	2 (0%)	1,144 (16%)
III 118	21 (18%)	0 (0%)	83 (70%)	0 (0%)	14 (12%)
IV 197	19 (10%)	1 (0%)	41 (21%)	1 (0%)	135 (69%)