Published in final edited form as:

Rev Recent Clin Trials. 2015 ; 10(3): 238–245.

PLCO: Evolution of an Epidemiologic Resource and Opportunities for Future Studies

Amanda Black^{a,*}, Wen-Yi Huang^a, Patrick Wright^b, Tom Riley^b, Jerome Mabie^b, Sunitha Mathew^c, Lawrence Ragard^c, Sigurd Hermansen^c, Kelly Yu^d, Paul Pinsky^d, Philip C. Prorok^d, Neal D. Freedman^a, and Robert N. Hoover^a

^aDivision of Cancer Epidemiology and Genetics, National Cancer Institute, NIH;

^bInformation Management Services, Inc.;

^cWestat, Inc.;

^dDivision of Cancer Prevention, National Cancer Institute, NIH

Abstract

The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO), a large-scale, multi-institutional, randomized controlled trial, was launched in 1992 to evaluate the effectiveness of screening modalities for prostate, lung, colorectal, and ovarian cancer. However, PLCO was additionally designed to serve as an epidemiologic resource and the National Cancer Institute has invested substantial resources over the years to accomplish this goal. In this report, we provide a summary of changes to PLCO's follow-up after conclusion of the screening phase of the trial and highlight recent data and biospecimen collections, including ancillary studies, geocoding, administration of a new medication use questionnaire, consent for linkage to Medicare, and additional tissue collection that enhance the richness of the PLCO resource and provide further opportunities for scientific investigation into the prevention, early detection, etiology and treatment of cancer.

Keywords

Cancer research; cohort; epidemiologic resource; PLCO; screening trial

1. INTRODUCTION

The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) was a large-scale randomized controlled trial designed to investigate whether selected screening tests lead to a reduction in mortality from the four targeted cancers. PLCO has evolved into a unique epidemiological cohort that serves as an important resource for the scientific community [1]. Participants were enrolled from 1993 to 2001 in 10 United States screening centers (Birmingham AL, Denver CO, Detroit MI, Honolulu HI, Marshfield WI,

* Address correspondence to this author at the Division of Cancer Epidemiology and Genetics, National Cancer Institute,.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

Minneapolis MN, Pittsburgh PA, Salt Lake City UT, St Louis MO, and Washington DC) and subsequently followed locally by that screening center through 2011. In 2011, participants were re-consented for continued follow-up using a centralized approach, providing the advantages of further follow-up, *e.g.* higher numbers of incident cancer cases and opportunities for supplemental data collection, at substantial cost savings. In addition to changes in follow-up, new data and biospecimen collections have further enhanced the cohort. The aim of this paper is to describe these enhancements in order to provide background for scientists who may wish to use this rich resource.

2. PLCO DESIGN AND METHODS OVERVIEW

The original design of PLCO, as well as previous data and biospecimen collections, has been described in earlier reports [2, 3]. However, we provide a brief background here. A total of 154,897 women and men aged 55 to 74 were enrolled between 1993 and 2001. Study participants were randomized into two arms, with screening-arm participants receiving a series of up to six annual screening examinations for prostate, lung, colorectal, and ovarian cancer. In contrast, participants in the other, usual care, arm received routine health care from their health providers. At baseline, all participants completed a baseline self-report questionnaire (BQX) and screening-arm participants also completed a dietary questionnaire (DQX). Beginning in 1998, a dietary history questionnaire (DHQ) was administered to all participants. The active screening component of the trial concluded in 2006. Blood was collected from screening-arm participants at baseline and at each annual screening visit thereafter for up to 5 additional years. In addition, buccal cells were collected from 58,000 usual care-arm participants in a onetime collection in 2000.

The trial was initially designed to follow participants for up to thirteen years after randomization, allowing sufficient time for the accrual of cancer cases and deaths and proper evaluation of the benefit-risk balance of each cancer screening test. During this period, through 2011, cancers were ascertained through routine follow-up of positive screening tests and through an annual study update (ASU) form that inquired about all cancer diagnoses. Such activities were carried out at each individual screening center, with a coordinating center (CC) managing all PLCO Trial procedures and data acquisition [4]. Cancers reported on the ASU forms then underwent a confirmation process in which the relevant medical records were obtained with cancers identified *via* ICD-O codes and abstracted using standardized forms. Deaths were ascertained primarily through the Annual Study Update (ASU) forms with the collection of copies of death certificates that were subsequently coded for the cause of death. For the PLCO cancers, additional medical records were collected for an independent committee to review to determine the cause of death separate from the death certificate documentation. Information about deaths was also supplemented by periodic linkage to the National Death Index (NDI). For the primary trial endpoints of mortality from the four trial cancers, events were censored at thirteen years of follow-up or on Dec, 31 2009, whichever came first. Currently available PLCO data for analysis are also generally censored at this time point; although in special cases the censoring date has been extended modestly past thirteen years or past 2009.

2.1. Rationale for Continued Follow-Up

In 2011, the trial faced a turning point. Individual contracts with each screening center were scheduled to terminate between 2011 and 2014. Yet, there was strong scientific rationale to leverage the investment in the trial by continuing follow-up. For example, continued follow-up would result in the accrual of substantially more cancer cases and mortality events, strengthening the trial's ability to clarify the long-term effects of screening on cancer mortality and enhance opportunities for basic, clinical, and translational research. Extended follow-up would also provide the opportunity to collect additional exposure data and biological specimens, such as administration of another risk factor questionnaire and developing an expanded tumor pathology resource for molecular and genetic studies. However, PLCO had already incurred an estimated cost of \$454 million by this point [1]. Extending follow-up, as previously performed, was cost prohibitive. For this reason, PLCO implemented an alternative approach for follow-up, in which each study center transferred follow-up activities to a centralized data collection center (CDCC). Between 2011 and 2014, all living PLCO participants were reconsented for continued follow-up and given the option to participate in (a) continued active follow-up performed by the CDCC, (b) passive follow-up performed by their local screening center, or (c) to withdraw from continued follow-up. This approach allows continued follow-up and additional data collections, but at substantial cost savings.

3. SUPPLEMENTARY DATA AND BIOSPECIMEN COLLECTIONS

A secondary objective of the PLCO trial is to identify environmental, molecular, and genetic risk factors for cancer. For this reason, NCI has continually expanded the cohort by funding additional data collections that permit application of rapidly advancing technologies and to address emerging hypotheses. Fig. (1) summarizes the PLCO data and biospecimen collections over the past three decades; from the initial recruitment in 1993 through present day (2015). The trial methods have been published previously and Carrick and colleagues describe the PLCO biospecimen collection in detail [5]. Here, we describe recent ancillary activities which enhance PLCO as a national resource.

3.1. Breast Cancer Supplemental (BCS) Data Collection

By 2005, it was recognized that the association with genetic and lifestyle risk factors may vary according to the breast tumor subtype (based on both established and recently identified histologic and molecular markers) [6]. As PLCO had not routinely collected data on stage, grade, or histology for breast cancer, as it was a non-PLCO cancer site, NCI initiated the BCS data collection project. This project allowed the ascertainment of a list of critical items related to breast cancer diagnosis, including histopathologic type, surgical resection procedure, laterality, multiple simultaneous ipsilateral primary, tumor size, lymph node status, and prognostic test results, *e.g.* estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor 2 receptor (HER2) status. At this time, PLCO collected these variables retrospectively for all previously diagnosed breast cancers. However, these policies were also adapted as standard operating procedures for newly diagnosed cancers. These data have been used widely in both individual and consortial studies of breast cancer [7–10].

3.2. Risk Factor Questionnaires

There have been five self-report risk factor questionnaires administered to PLCO participants between enrollment and 2015, including the BQX, DHQ, DQX, Supplemental Questionnaire (SQX), and the Medication Use Questionnaire (MUQ). The data collected on each of these questionnaires are summarized in Table 1.

It is envisioned that the information in the SQX and MUQ will be used primarily in a prospective fashion (*i.e.* the SQX or MUQ is the starting point for follow-up), with associations examined for endpoints occurring subsequent to the administration of the questionnaire. Due to the relatively short time between SQX and MUQ administration and the current follow-up censor date (December 31, 2009 or 13 years of follow-up) there has been little use of the questionnaires in this fashion to date. However, these data will play an important role in disease association studies in the future as follow-up time and events accrue. Cross-sectional analyses may also be of interest, for example, in combination with the subset of participants who happened to have a temporally relevant blood collection. Of course, it will also be possible to use the data retrospectively. However, caution is merited. For example, there is potential for survival bias, as only those who survived long enough had the opportunity to complete the questionnaire. On the other hand, studies of cancer survival may benefit from careful exposure assessment at two time-points, for example, before and after cancer diagnosis (*e.g.* smoking status before and after a cancer diagnosis).

3.2.1. Supplemental Questionnaire (SQX)—In 2006, soon after the screening component of the trial had ended, participants were invited to complete a self-report SQX designed to update risk factor data ascertained at baseline (*e.g.*, family history of cancer, current smoking status, current body mass index, and comorbidities) as well as to collect relevant new information (*e.g.*, detailed information on hormone therapy, cigarette smoking, physical activity, waist-hip fraction, and use of anti-inflammatory, pain relievers, and bone strengthening medications). All living and actively followed participants were mailed the SQX (134,848 participants; 87% of the 154,897 originally enrolled). Overall, 104,478 individuals responded to the one-time mailing (a response rate of 77.5%).

To facilitate the use of the information on the SQX, we compare baseline characteristics for those completing the questionnaire to those who did not in Table 2. In Table 3, we provide the number of prevalent cancer cases among those completing the SQX (*i.e.* diagnosed prior to completing the questionnaire) and detail the number of incident cases that have occurred between completion of the SQX and the current follow-up censor date. With additional follow-up, these numbers will continue to increase.

3.2.2. Medication Use Questionnaire (MUQ) & Medicare Linkage—In light of the paucity of data on medication use (prescription and over-the-counter) and cancer in the elderly, PLCO invited participants to complete a comprehensive medication use questionnaire (MUQ) in 2013, at which point PLCO participants ranged in age from 66–93 years. In addition to asking about Nonsteroidal Anti-Inflammatory Drugs (NSAIDs), which had been assessed on prior questionnaires, the MUQ asked participants to report all prescription medications used in the past year, as well as their frequency and duration. It also

asked participants to update their smoking status and provide their current weight. Finally, it asked whether participants would permit electronic linkage to electronic health systems such as Medicare and Medicaid. Although PLCO collected a wealth of information at baseline and during follow-up, it lacks information on cancer treatment of non-PLCO cancers and on incident comorbidities. These factors are important predictors of cancer outcomes and potential confounders. Preexisting medical conditions also influence cancer treatment decisions, cancer screening, cancer survival, and the development of second cancers. In addition, chronic diseases such as stroke and diabetes are of interest in their own right, in addition to serving as important comorbidities for cancer. Linkage with Medicare provides an important potential source of additional data, including demographic and lifestyle exposures, disease diagnosis, treatment for cancer and other diseases, and non-cancer comorbidities.

Of 75,344 participants who were mailed the MUQ, 59,898 completed and returned the questionnaire (a response rate of 79%). Of the responders, 49,507 (83%) also consented to allow their data to be linked with electronic health data such as Medicare. More detailed methods for the MUQ and its initial findings will be reported elsewhere.

3.3. Contract Closeout Studies

With the anticipated expiration of the individual screening center contracts in 2011, a special call for proposals was announced in 2010. This call allowed researchers from the scientific community the opportunity to request new or additional data prior to final archival by each screening center. These studies are described below.

3.3.1. Pathology Report Retention—Over the course of PLCO, pathology reports were collected for all cancers but only abstracted for the four PLCO cancers (prostate, lung, colorectum, ovary) and breast cancer. To gain pathology data (*e.g.* stage, grade) for other cancer sites, screening centers were asked to retain copies of the pathology reports during centralization. As these reports contain personally identifiable data, this effort was restricted to deceased participants and those who opted for centralized active follow-up during the re-consent process. Each center mailed copies of available pathology reports for cancers diagnosed through December 2009 to the CDCC. These reports, as well as pathology reports from confirmed cases diagnosed after centralization to 2014, have been fully redacted of identifiable information and electronically scanned. To date, approximately 24,603 pathology reports have been redacted and scanned. In addition, most of these reports have been abstracted for the following variables: stage, grade, histology, tumor size, primary site, and diagnostic procedure.

3.3.2. Pancreatic and Liver Cancers—An additional effort was made for pancreatic and liver cancers diagnosed in PLCO participants. Histology, stage, and treatment data were abstracted for approximately 700 confirmed pancreatic cancers that had occurred prior to December 2009. These data have been utilized in several studies to date and hopefully will help understand this typically lethal cancer.

A second effort, focused on distinguishing primary liver cancer from metastases, was considered important as the incidence of primary liver cancer and cancers metastatic to the

liver are roughly similar in the US population. As such, we reviewed the medical records of all confirmed liver cancers in PLCO that were diagnosed prior to December 2009 (n=246). Primary liver cancer was verified in 186 of these cases. Of the remaining 60 cases, 2 reported cancer, but the records could not confirm any cancer diagnosis. All other cases were confirmed to be metastatic to the liver. We recorded histologic data from the verified cancers, enabling future studies that address specific tumor subtypes such as the most common histologic type, hepatocellular carcinoma.

3.4. The Prostate Cancer Progression Study (PCP)

Despite its very high incidence, the etiology and natural history of prostate cancer remains poorly understood. For this reason, PLCO has established a natural history study of prostate cancer, the Prostate Cancer Progression Study (PCP). Between 1993 and 2009, 7,538 men were diagnosed with localized prostate cancer and eligible for inclusion. Of the 6,113 men with a prior history of prostate cancer who were alive at the time of selection, 5,255 (86%) participated in a telephone survey. Relevant medical records were obtained from hospitals and clinicians for a total of 2517 men; 87% (n=1,151) of men who completed the survey and reported recurrence or progression of disease and for an additional 1,236 participants who had died prior to the initial selection. From these records, we abstracted detailed information on progression over time (e.g. PSA levels, secondary treatment) and also validated the self-reported data. Tumor tissue was obtained from 1073 men included in the PCP study, allowing the construction of tissue microarrays, with accompanying tumor cores saved for nucleic acid extraction. This unique resource, comprising pre-diagnostic exposure data and blood specimens, tumor tissue, treatment information, and detailed outcomes data allows comprehensive interrogation by both established and emerging approaches and will hopefully contribute to our understanding of prostate cancer progression.

3.5. Tumor Tissue Collection

PLCO is one of only a few cohort studies to have collected tumor tissue, with cores and tissue microarrays (TMAs) available for a number of different cancer types. As of 2011, PLCO had created tissue cores and constructed TMAs from cancers of the prostate (N=1095), lung (N=435), colon/rectum (N=678), adenoma (N=658), ovaries (N=212), and breast (N=800 female; 18 male). In 2013, PLCO undertook collection of additional tumor tissue including from bladder (N=214), breast ductal carcinoma *in situ* (DCIS) (N=95), and additional colon/rectum cases (N=226). A separate collection targeted at HPV related cancers has resulted in the successful collection of 29 cases of head, neck, and anus tumors that will be used for a global cohort consortium of HPV-driven tumors. Our tissue collection and processing protocols will be described in detail elsewhere. These tissue collections provide the opportunity to measure genetic alterations and other biological changes in tumors and link them to abundant existing clinical and risk factor data as well as other specimens. Together, these resources make PLCO a unique resource for studying the pathobiology of cancer, understanding etiological heterogeneity, and identifying biomarkers for early detection.

3.6. Geocoding

Geocoding is the process of assigning a geographical identifier (latitude and longitude) to an address in a geographic information system (GIS). Geocoded data can then be used to link study participants to georeferenced databases such as U.S. census data, environmental monitoring data (e.g. UV exposure), and locations of health care facilities. The use of geocoded data has the potential to support research by providing contextual and environmental data (e.g. neighborhood characteristics, levels of pollutants in the environment) that cannot be provided through individual interviews.

PLCO recognized that geocoding participant address data would provide opportunities to obtain a wide array of additional information about cohort participants and potentially open up new areas of research. Therefore, in 2012, PLCO geocoded addresses of eligible participants and linked to the 2000 and 2010 U.S. Censuses. Eligible participants included those who were deceased prior to centralization and those who consented to centralized follow-up. Overall, 84,412 study participants from 7 screening centers were successfully geocoded and linked to Census data. The geocoding process, resulting data, and data access policies are described below.

3.6.1. PLCO Addresses—Until recently, PLCO only maintained a record of each participant's most recent address *i.e.*, if a participant moved residence during follow-up their previous address was overwritten with the new address. Therefore, only a single address was available to generate geographical coordinates. Participants who died during follow-up had their last known address used for geocoding, whereas participants who were alive in 2011 had their most recent address used. As such, the 2010 Census data was most relevant for the majority of geocoded PLCO participants.

3.6.2. Geocoding Addresses and U.S. Census Linkage—To geocode the addresses, Westat, Inc. used ESRI's *Ar-cGIS Desktop* in conjunction with the most recent *StreetMap Premium NAVTEQ* street reference database (*StreetMap Premium NAVTEQ 2012 Version 1*, ESRI, Redlands, CA), a process that meets the World Geodetic System 1984 (WGS84) standard (National Geospatial-Intelligence Agency, WGS 84 (G1674), 2005.0). The geocoding engine extracts interpolated latitude and longitude coordinates for each input address. The most accurate match occurs when available address information matches directly to the street segment information (*i.e.* street name and address range) stored within the NAVTEQ database (more accurate "point estimates" of addresses now available for many residential addresses were not available in 2012); if this initial exact match doesn't occur, then the geocode is assigned to the center of the address's ZIP code, or to the center of the city if the ZIP code cannot be matched. Using this method, only 0.2% of all addresses submitted could not be matched.

Once records are assigned their coordinates, a point file is generated. This geographic point layer file allows each address to be spatially assigned to its proper Census 2000 and Census 2010 boundaries. NAVTEQ maintains Census boundary layers that are properly nested within the NAVTEQ street database. This results in a more accurate assignment of Census variables, as opposed to using data from multiple sources to do the spatial assignments.

As the year of last known address for participants spanned almost 20 years between 1993 and 2011, PLCO linked geocoded addresses to both the 2000 census and 2010 census demographic and sociodemographic variables. Nonetheless, the Census in 2010 was dramatically abbreviated, consisting of only 10 questions, compared to the long-form questionnaire used in previous years. As such, only a few of the socioeconomic and demographic variables of interest from the 2000 census have equivalents in the 2010 census data. Others, however, have comparable variables in the American Community Surveys (ACS) - a mandatory, annual survey completed by a small percentage of the population. Therefore, in addition to linking to the censuses, we also linked to the 2006–2010 ACS data to supplement the 2010 census variables. The 2006–2010 ACS data were selected because the 2010 census tract boundaries were used for these surveys whereas earlier surveys used census 2000 tract boundaries. Overall, we identified a set of 29 variables of primary interest, 21 of which are comparable between both the 2000 census and the 2010/ACS combined source. The key variables included from each data source are presented in Table 4. These variables may be used independently or used to construct indicators of neighborhood deprivation such as the composite census-tract-level socioeconomic deprivation index to explore the effect of neighborhood contextual factors on person-level exposure-disease relationships, for example.

3.6.3. Strengths and Limitations of PLCO Geocode Data—PLCO comprises a large and geographically diverse US population making it ideal for spatial analyses. Although the entire PLCO cohort was not eligible for geocoding, we nevertheless have geocoded addresses of more than 84,000 individuals from across the U.S. within the setting of a large prospective cohort. Such numbers are sufficient to facilitate a diverse spectrum of research questions and can also be pooled with similarly geocoded cohorts to investigate population subsets, including underserved and/or minority populations.

Having only a single residential address for each participant - the last known address during follow-up - is a noteworthy limitation. The timing of last known address should be considered not only when deciding which census data to use (2000 vs 2010) for each individual but also when planning studies to link geocoded coordinates to other exposure databases. Using a single address or geographic location is unlikely to reflect cumulative lifetime exposure and may not capture exposures at a time when they are most etiologically relevant. Nevertheless, the geocoded data can be linked to demographic, social, and environmental data sources allowing researchers to use spatial approaches to investigate and identify potential relationships between exposures and disease. Moreover, a major strength of PLCO lies in the repository of biological specimens which can be used for direct measurement of exposure levels within individuals and the results used in combination with derived geospatial exposure data, self-reported data, and clinical outcome data. For these analyses, it is particularly important to consider the timing of data and biospecimen collections vis-à-vis our available data on residential history. Participant addresses spanned the period 1993 to 2011 (with the majority being last known after 2005), bloods were collected between 1993 and 2006, and census data were obtained from both the 2000 and 2010 census databases.

3.6.4. Data Access—Each geocoded address was assigned a latitude and longitude. However, these geocoded coordinates are personally identifiable data that must be kept confidential to preserve participant privacy and, as such, these coordinates cannot be released to researchers under any circumstances. The data available to investigators include “pseudo codes” which can be used in multilevel analyses to account for clustering of participants who live in the same census tract. These pseudo census tract identifiers, along with the aggregate data at the census tract level, will enable investigators to account for clustering in statistical analyses.

4. SUMMARY

More than twenty years after its founding, PLCO remains a unique resource for the scientific community. Continuing follow-up and additional data and biospecimen collections have strengthened the value of the resource and provide numerous opportunities for scientific investigation into the prevention, early detection, etiology and treatment of cancer.

PLCO data and biospecimens are a resource for use by the entire scientific community. Further information on PLCO and how to access PLCO data and biospecimens can be found at <https://biometry.nci.nih.gov/cdas/studies/plco/> and www.plcostars.com, respectively.

ACKNOWLEDGEMENTS

This research was supported by the Intramural Research Program of the National Cancer Institute, National Institutes of Health and the Division of Cancer Prevention, National Institutes of Health.

Biography

Amanda Black

REFERENCES

- [1]. Zhu CS, Pinsky PF, Kramer BS, et al. The prostate, lung, colorectal, and ovarian cancer screening trial and its associated research resource. *J Natl Cancer Inst.* 2013;105(22):1684–93. [PubMed: 24115361]
- [2]. Hayes RB, Reding D, Kopp W, et al. Etiologic and early marker studies in the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial. *Controlled Clin Trials.* 2000;21(6 Suppl): 349S–55S. [PubMed: 11189687]
- [3]. Prorok PC, Andriole GL, Bresalier RS, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Controlled Clin Trials.* 2000;21(6 Suppl):273S–309S. [PubMed: 11189684]
- [4]. O'Brien B, Nichaman L, Browne JE, Levin DL, Prorok PC, Gohagan JK. Coordination and management of a large multicenter screening trial: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Controlled Clin Trials.* 2000;21(6 Suppl):310S–28S. [PubMed: 11189685]
- [5]. Carrick DM, Black A, Gohagan J, et al. The PLCO Biorepository: Creating, maintaining, and administering a unique biospecimen resource. *Rev Recent Clin Trials*, In Press.

- [6]. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747–52 [PubMed: 10963602]
- [7]. Schairer C, McCarty CA, Isaacs C, et al. Circulating insulin-like growth factor (IGF)-I and IGF binding protein (IGFBP)-3 levels and postmenopausal breast cancer risk in the prostate, lung, colorectal, and ovarian cancer screening trial (PLCO) cohort. *Horm Cancer*. 2010;1(2): 100–11. [PubMed: 21761353]
- [8]. Barndahl M, Canzian F, Lindstrom S, et al. Association of breast cancer risk loci with breast cancer survival. *Int J Cancer*. 2015; [Epub ahead of print].
- [9]. Mondul AM, Shui IM, Yu K, et al. Vitamin D-associated genetic variation and risk of breast cancer in the Breast and Prostate Cancer Cohort Consortium (BPC3). *Cancer Epidemiol Biomarkers Prev*. 2015; 24(3):627–30 [PubMed: 25542828]
- [10]. Falk RT, Maas P, Schairer C, et al. Alcohol and risk of breast cancer in postmenopausal women: an analysis of etiological heterogeneity by multiple tumor characteristics. *Am J Epidemiol*. 2014;180(7):705–17 [PubMed: 25150269]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. (1).
PLCO Timeline highlighting select data and biospecimen collections during follow-up (1993–2015).

Summary of data collected on each of the PLCO questionnaires administered between 1993 and 2015.

Table 1.

	Questionnaire	BQX	DHQ	DQX	SQX	MUQ
	Year(s) Administered	1993–2001	1998–2001	1993–2001	2006–2007	2013
	Intervention/Control Arm	Both	Both	Intervention Only	Both	Both
	N Completed	~150,000	~113,000	~61,000	~104,000	~60,000
Risk Factor Categories	Risk Factor Variables					
	Race/Ethnicity	✓			✓	
Demographics	Income				✓	
	Marital Status	✓			✓	
	Education	✓				
	Smoking	✓			✓	✓
Lifestyle & Anthropometry	Alcohol use		✓	✓		
	Physical Activity		✓	✓	✓	
	Anthropometry	✓			✓	✓
	Family history of cancer	✓			✓	
Medical History	Personal history of cancer	✓				
	Personal medical history	✓			✓	
	N SAIDs/Aspirin/Acetaminophen	✓			✓	✓
Medication Use	Prescription medication use					✓
	Reproductive history	✓			✓	
F female-specific	Cancer screening	✓			✓	
	Hormone therapy	✓			✓	
	Male pattern baldness				✓	
Male-specific	Urinary/prostate symptoms	✓			✓	
	Dietary Intake		✓	✓		
Diet	Supplemental Vitamin Use		✓	✓		

Table 2.

Baseline characteristics of PLCO participants who responded to the Supplemental Questionnaire (SQX), non-responders, and those who were deceased or otherwise ineligible to complete the SQX

	SQX Responders N=104,478	SQX Non-Responders N=30,370	Deceased/Ineligible* N=20,049
	N (%)		
Gender			
Male	49,602 (47.5)	14,878 (49.0)	12,202 (60.9)
Female	54,876 (52.5)	15,492 (51.0)	7,847 (39.1)
Age at Randomization (years)			
55–59	37,385 (35.8)	10,403 (34.3)	3,899 (19.4)
60–64	33,226 (31.8)	8,850 (29.1)	5,475 (27.3)
65–69	22,549 (21.6)	6,695 (22.0)	5,686 (28.4)
70+	11,318 (10.8)	4,422 (14.6)	4,989 (24.9)
Race			
White	93,536 (89.5)	23,248 (76.6)	15,795 (78.8)
Black	3,225 (3.1)	3,165 (10.4)	1,318 (6.6)
Other/Missing	7,717 (7.4)	3,957 (13.0)	2,936 (14.6)
BMI (kg/m²)			
18.5–25	34,588 (33.1)	8,888 (29.3)	6,087 (30.4)
25–30	43,193 (41.3)	11,862 (39.1)	7,429 (37.1)
>30	23,283 (22.3)	7,744 (25.5)	4,592 (22.9)
Unknown/Missing	3,414 (3.3)	1,876 (6.2)	1,941 (9.7)
Smoking Status			
Never smoker	49,584 (47.5)	13,214 (43.5)	6,474 (32.3)
Former Smoker	44,035 (42.2)	11,916 (39.2)	8,666 (43.2)
Current Smoker	8,777 (8.4)	3,893 (12.8)	3,385 (16.9)
Unknown/Missing	2,082 (2.0)	1,347 (4.4)	1,524 (7.6)

* Ineligibles include those who withdrew from follow-up or were lost to follow-up prior to administration of the SQX

Table 3.

Prevalent* and incident** cancers among 104,478 Supplemental Questionnaire (SQX) respondents.

	Prevalent Cases	Incident Cases
Cancer Site	N	
All sites [^]	12,012	5,094
Prostate	4,591	1,306
Breast	2,409	816
Lung	523	604
Colorectal	972	311
Bladder	606	306
NHL	489	263
Pancreas	51	149

* Prevalent cases are defined as those diagnosed in participants after enrollment and prior to completion of the SQX (ca. 2006)

** Incident cases include all cancer cases diagnosed between SQX completion and 13 years of follow-up from study entry or December 31st, 2009, whichever occurred earliest.

[^] All confirmed cancers diagnosed in PLCO including those sites not listed individually here

Table 4.

Key variables obtained through linkage to the 2000 and 2010 U.S. Censuses and to the American Community Survey.

Category	Variable	2000 Census	2010 Census	ACS
Demographics (2 variables)	Percent of non-hispanic blacks	✓	✓	✓
	Percent of blacks who are hispanic	✓	✓	✓
Income (7 variables)	Percent with no car	✓		✓
	Percent of poverty	✓		✓
	Percent on public assistance	✓		✓
	Percent below poverty	✓		✓
	Percent of households with income less than 30,000	✓		
	Median Household Income	✓		✓
	Less than a high school degree	✓		✓
Housing (11 variables)	Housing units with a mortgage 50.0 percent or more	✓		
	Housing units without a mortgage 50.0 percent or more	✓		
	Median value for all owner occupied housing units	✓		✓
	Percent of renter occupied housing units	✓	✓	✓
	Percent of housing units vacant	✓	✓	✓
	Percent of crowding	✓		✓
	Census Population	✓	✓	✓
	Percent of female head of households with children	✓		
	Count of residents 65 and older	✓	✓	✓
	Percent living in the same residence for 5 years or more	✓		✓
	Percent of urban population	✓		
Employment (9 variables)	Percent of Unemployed Men	✓		✓
	Percent in lower social class	✓		
	Percent of females in management occupations	✓		✓
	Percent of males in management occupations	✓		✓
	Percent of females not in labor force	✓		✓
	Percent of males not in labor force	✓		✓
	Percent of females in professional occupations	✓		

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Category	Variable	2000 Census	2010 Census	ACS
	Percent of males in professional occupations	✓		
	Unemployed	✓		✓