
PLCO USER GUIDE

TABLE OF CONTENTS

| | |
|--|----|
| 1. INTRODUCTION | 2 |
| 2. GENERAL DATA USE | 2 |
| MISSING CODES..... | 2 |
| QUESTIONNAIRE FOLLOW-UP ELIGIBILITY..... | 2 |
| MERGING DATA ACROSS FILES | 3 |
| 3. TIME AND EVENTS..... | 3 |
| DATES AND DAYS (ENTRY, EXIT, AND END OF FOLLOW-UP)..... | 3 |
| CANCER ASCERTAINMENT | 4 |
| STAGE, ICD-O-2 CODES AND OTHER CANCER CHARACTERISTICS..... | 4 |
| DIAGNOSTIC PROCEDURES AND TREATMENT | 4 |
| MORTALITY | 5 |
| DEATHS FROM CANCER WITHOUT A CONFIRMED DIAGNOSIS..... | 5 |
| EXTENDED FOLLOW-UP..... | 5 |
| 4. PROTOCOL CHANGES | 5 |
| PRIOR SCREENING EXAMINATIONS | 5 |
| DUAL CONSENT | 6 |
| AGE RANGE..... | 6 |
| WOMEN WITHOUT OVARIES..... | 6 |
| SCREENING CHANGES | 6 |
| CA-125 ASSAYS, VERSIONS 1 AND 2..... | 7 |
| BROOKLYN AND ALABAMA..... | 7 |
| 5. CANCER SITE SPECIFIC CONSIDERATIONS | 7 |
| LUNG | 7 |
| COLORECTAL AND COLORECTAL ADENOMA..... | 8 |
| OVARIAN..... | 8 |
| FEMALE BREAST..... | 9 |
| GLIOMA..... | 9 |
| HEAD AND NECK..... | 9 |
| HEMATOPOIETIC..... | 9 |
| PANCREAS..... | 10 |
| UPPER GASTROINTESTINAL | 10 |
| 6. OTHER CONSIDERATIONS | 10 |
| BASELINE QUESTIONNAIRE (BQ)..... | 10 |
| QUALITY ASSURANCE (QA) EXAMS..... | 10 |
| SECOND SAME SITE PRIMARY | 10 |
| MENOPAUSE STATUS..... | 11 |
| 7. ACRONYM GLOSSARY | 11 |

1. INTRODUCTION

This document is intended to serve as a guide to using the PLCO data. It covers five main themes related to your dataset: General Data Use, Time and Events, Trial Protocol Changes, Cancer Organ Site-specific Information and Other Considerations. We recommend that you review this document prior to starting your analysis to ensure correct use of the data and to familiarize yourself with how data relevant to your analyses were collected in PLCO. The data delivered to you represent the most up to date information currently available. Every effort was made to ensure the high quality of PLCO data. Thorough quality control protocols and procedures were implemented to see that data were captured accurately and consistently. However, due to routine cleaning and updates, it may not be possible to replicate the published PLCO trial findings using the CDAS datasets. The datasets will be updated periodically and this user guide will be updated accordingly.

2. GENERAL DATA USE

MISSING CODES

Missing codes have been standardized to represent the same thing for all variables across all datasets. They are as follows:

| CODE | FORMAT | FURTHER EXPLANATION |
|------|-----------------------------------|---|
| .A | Ambiguous | Contradictory information resulted in an unclear conclusion. |
| .C | Control Arm | The variable does not apply to participants in the control arm. |
| .F | No Form | The form was not filled out. |
| .G | Wrong Gender | The variable only pertains to one gender. |
| .M | Missing/Not Answered | The question was not answered or data are otherwise missing. |
| .N | Not Applicable | The variable did not apply to this participant. |
| .R | Out of Range | The value was considered either too extreme or impossible. |
| .V | Not Captured on this Form Version | The information was only available on a different form version. |

QUESTIONNAIRE FOLLOW-UP ELIGIBILITY

Multiple questionnaires were completed by participants over the course of the trial. The Baseline Questionnaire (BQ) provides background information and was completed by 97% of participants close to trial entry. Potentially two dietary questionnaires (DQX in the screening arm, DHQ in both arms) were completed within the first few years of follow-up, capturing self-reported dietary habits. A Supplemental Questionnaire (SQX) was completed later in the trial, capturing information similar to that on the BQ.

You may consider data from these questionnaires useful predictors for your analysis. To identify which questionnaires are eligible for prospective analyses, indicators are provided in the data for each cancer site. These indicators take the form [site]_ELIGIBLE_[qx]. For example, for colorectal cancer analyses using SQX data, limit to records where COLO_ELIGIBLE_SQX=1.

In general, data from a questionnaire is not eligible for site-specific analyses if: the questionnaire itself was not considered valid; the participant had a prior diagnosis of that site cancer; the participant did not have known time

at risk following the questionnaire. Further, as diet may change in response to cancer diagnoses, a dietary questionnaire is indicated ineligible for prospective analyses given history of any cancer prior to the questionnaire.

Also provided in the data are indicators for personal history of site-specific or any cancer prior to trial entry. These take the form PH_[site]_TRIAL and PH_ANY_TRIAL.

MERGING DATA ACROSS FILES

The primary dataset contains all information necessary for most standard analyses. However, more detailed information is available for certain trial events. PLCO cancers have secondary datasets for screens, abnormalities, diagnostic procedures, treatments and medical complications. The breast, hematopoietic and head and neck sites have an associated dataset containing second cancers. All files on CDAS contain the personal identifier variable PLCO_ID. This variable should be used when merging datasets. The screen, abnormality and procedure files also contain STUDY_YR which links any given screen with its findings and follow-up. Lastly, the abnormality file contains the variables SOURCE and VISIT. The first will link the abnormality to either the initial screen or the QA reading, and the second will tell you from which screening visit it came.

3. TIME AND EVENTS

DATES AND DAYS (ENTRY, EXIT, AND END OF FOLLOW-UP)

In accordance with HIPAA, and to protect PLCO participants' privacy, no dates are included in the datasets. Instead, "days" variables are provided to represent the number of days from randomization to each event. Year of randomization (RNDYEAR) is provided to identify the calendar year of trial entry. Specific dates are available only by special request. Two important events at trial entry are randomization and the Baseline Questionnaire (BQ). For 82% of participants, these two events occur within a month of each other. Variables are included to indicate the time by which both events have taken place (ENTRYDAYS_BQ and ENTRYAGE_BQ). Entry days and age variables are available for the main questionnaires.

For any given site, three exit times are supplied: site-specific cancer incidence ([site]_EXITDAYS); first cancer of any type (FSTCAN_EXITDAYS); mortality (MORTALITY_EXITDAYS). Each has a corresponding exit age ([site/FSTCAN/MORTALITY]_EXITAGE).

For cancer incidence exit times, the exit is the day of cancer diagnosis or the day last known cancer free for that site, good through 2009 at the latest.

For mortality exit times, the exit is defined as the day of death or the day last known alive, good through 2018 at the latest.

For all exit times, the status at that exit can be determined using the corresponding variables: [site/FSTCAN/MORTALITY]_EXITSTAT.

Screening study years (T[X]) PLCO nomenclature refers to time on study with the indicator T[X]. The "X" indicates the number of years since randomization. Thus, T0 begins on the date of randomization, and ends exactly one year later. This is where T1 begins, and so forth. The T[X] screening exams are expected at the beginning of the T[X] study year. For the colorectal site, the flexible sigmoidoscopy screening exam was done at T3 for some people and T5 for others (see Protocol Changes) so the terms T35 or T3/5 are used to denote the time point of this second exam.

CANCER ASCERTAINMENT

In the PLCO trial, reports of cancers were collected through various means including, but not limited to, self-reports, family reports and death certificates. Reports of cancer were followed up and any available medical records were abstracted. Follow-up activity continued when records could not be readily obtained. Some suspected cancers remain unresolved. Cancer ascertainment can be determined by cancer indicator variables ([site]_CANCER). Confirmed invasive cancers, *in situ* cancers and borderline malignancies are considered end points for cancer incidence. Once it is recorded that a participant has been diagnosed with one of these conditions, no other subsequent diagnoses are recorded. Other considerations on a site by site basis should be taken into account when determining case status. For these, see Site Specific Considerations.

STAGE, ICD-O-2 CODES AND OTHER CANCER CHARACTERISTICS

Cancer stage is provided for prostate, lung, colorectal, ovarian, breast and bladder cancers only. Staging follows TNM classification; individual T, N and M components are provided in the data. Prostate, lung, colorectal, breast and bladder stage can come from clinical assessment ([site]_CLINSTAGE) and/or pathology ([site]_PATHSTAGE). The default variable [site]_STAGE records pathologic stage if available, clinical stage otherwise. Nearly all cancers with staging data have clinical stage ascertained at diagnosis; fewer cancers have pathologic stage. Ovarian cancer is primarily staged through oophorectomy, so the distinction between clinical and pathologic staging variables is less relevant.

During the PLCO trial, most staging data conformed to the 5th edition of the AJCC Cancer Staging Manual, so this is the default staging scheme for the provided stage variables. Additional 7th edition variables are provided for PLCO and breast cancers. The 7th edition variables were created from the same information used for 5th edition variables, but with the 7th edition staging scheme applied.

ICD-O-2 behavior, morphology, topography and grade codes are provided for non-PLCO sites. ICD-O-2 codes are not available for the PLCO sites.

Gleason score is provided for prostate cancers. Gleason score can come from biopsy (PROS_GLEASON_BIOP) and/or prostatectomy (PROS_GLEASON_PROST). The default variable PROS_GLEASON records the prostatectomy Gleason score if available, the biopsy Gleason score otherwise.

Female breast cancers also have additional cancer characteristic information gathered through the Breast Cancer Supplemental (BCS) form, described later in this document.

DIAGNOSTIC PROCEDURES AND TREATMENT

Information pertaining to diagnostic procedures and treatment is available only for PLCO cancers, and to a smaller extent for female breast cancer. Diagnostic procedures are recorded for all PLCO cancers. Additionally, procedures done as follow-up to a positive trial administered screen are recorded. Summary data of these procedures are available in the primary site datasets, and additional data are available in the procedures and treatment datasets. Only the initial course of treatment following the diagnosis is available.

MORTALITY

Death certificates are the primary source of dates and underlying causes of death in the trial, and were obtained via annual follow-up and annual searches of the National Death Index (NDI). An independent Death Review Committee further reviewed deaths among those diagnosed with a confirmed PLCO cancer, those whose death certificate stated they died of a PLCO cancer, and those whose death certificate was ambiguous as to whether the cause of death was a PLCO cancer.

Variables prefixed D_ identify the cause of death according to the death certificate. Variables prefixed F_ also use the death certificate, but prioritize the cause of death as determined by the Death Review Committee.

DEATHS FROM CANCER WITHOUT A CONFIRMED DIAGNOSIS

The trial obtained data for cancers and deaths from different sources. As such, occasionally a death attributed to a particular cancer will not always be preceded by a diagnosis of that cancer. It is left to the user to determine how to handle these discrepancies.

EXTENDED FOLLOW-UP

After the trial ended in 2009, follow-up was centralized. Continued follow-up required participant re-consent. In the absence of participant contact, centers determined the status of re-consent by default. Participants who *refused consent* had no extended follow-up. Participants with *passive consent* were no longer contacted, but passive searches such as with the National Death Index (NDI) were conducted. Participants with *active consent* continued to be contacted and could participate in data collection efforts such as the Medication Use Questionnaire (MUQ) and the Brief Survey (BCQ) in addition to being passively searched. For studies concentrated on extended follow-up data you may choose to consider potential biases introduced by the re-consent process. To identify if and why a participant transferred for continued follow-up, re-consent outcome variables have been provided in the data.

For most analytic purposes, the provided incidence and mortality exit days variables suffice; these censor participants at the conclusion of incidence and mortality follow-up, respectively.

4. PROTOCOL CHANGES

Several parts of the protocol were significantly changed during the course of the trial. Users should pay special attention to how these affect the data.

PRIOR SCREENING EXAMINATIONS

Beginning in April 1995, changes were made to enrollment eligibility based on screening contamination. Men who had received two or more PSA exams in the three years prior to enrolling were no longer considered eligible. Similarly, any prospective participant of either gender who had received one or more colorectal cancer screens in the three years prior to enrolling was no longer considered eligible.

DUAL CONSENT

All Study Centers had to use a pre-randomization informed consent procedure in which eligible and interested potential participants would sign a main study consent form prior to randomization. Each Study Center had the option of using one of two informed consent procedures: single consent and dual consent.

Single consent required informed consent from participants before randomization into the trial. It was believed that this consent, covering all aspects of participation including screening and data collection, could lead to greater contamination, but could also lead to greater compliance.

The dual consent process was to consent all willing participants into an observational study of cancer risk, without mention of the interventions. The purpose of this process was to keep subjects blinded as to the study protocol with hopes of minimizing contamination in the control arm. After randomization, the intervention subjects were consented a second time, receiving information regarding the screening protocol and being advised they will be invited to come into the screening center for the study examinations. At this time, subjects were also informed of the screening center locations. This approach could reduce contamination but could also lead to lowered compliance.

Three of the initial ten screening centers (Henry Ford Health System (HFHS), Washington University in St Louis (WUSTL) and Pacific Health Research Institute in Honolulu (PHRI)) chose the dual consent approach. However, after several years of recruitment, all centers switched to a single consent. PHRI stopped dual consent on Feb 28, 1995 and WUSTL stopped on March 30, 1995. HFHS stopped gradually around 1997.

It should be noted that HFHS and PHRI enrolled a large percentage of the racial minorities in PLCO; HFHS with African Americans, and PHRI with Asians and Pacific Islanders.

AGE RANGE

At the start of the trial, the eligible age range of participants was 60-74. In January 1996, the eligibility criteria were changed such that prospective participants aged 55-59 could be randomized.

WOMEN WITHOUT OVARIES

At trial inception, women without ovaries were not considered eligible to be randomized. The protocol was changed in October 1996 to allow these women into the trial.

SCREENING CHANGES

In December 1998 and September 1999, changes were made to the screening protocols affecting all four sites.

PSA AND CA-125

In September 1999, PSA and CA-125 screening exams were added in T4 and T5. There are participants randomized towards the beginning of the trial who were not expected for their T4 or T5 screen and some in the transition period who were not expected for T4, but were expected for T5. PSA_PROT and CA125_PROT provide information on who was expected for which screens. This compliance is also taken into account in the creation of the results variables (PSA_RESULT4-5 and CA125_RESULT4-5).

X-RAY

In December 1998, the T3 X-ray exam was discontinued for non-smokers. XRY_PROT provides information on who was expected for this T3 screen. This compliance is also taken into account in the creation of the results variables (XRY_RESULT3).

FLEXIBLE SIGMOIDOSCOPY

Participants were expected to receive one baseline colorectal screen (T0 FSG) and one follow-up screen. Initially, enrolled participants were scheduled for T0 and T3 FSG screens. In December 1998, the trial discontinued the T3 colorectal screen and in September 1999 began administering this exam at T5. The result of the follow-up screen is provided in the variable FSG_RESULT35, though there are variables that keep these two study years apart (FSG_RESULT3, FSG_RESULT5). The variable PROTOCOLC indicates which study year (T3 or T5) the participant was expected to return for the follow-up screen.

OVARIAN PALPATION

At the start of the trial, a third ovarian cancer screening test, ovarian palpation, was used. In December 1998, this exam was discontinued. No data from this screen are included in these datasets.

CA-125 ASSAYS, VERSIONS 1 AND 2

When the trial began, PLCO used the first version of the assay for all CA-125 exams. On October 1, 1995, CA-125II became available and the protocol was switched to use this for all subsequent screens. A few years after this transition, all of the original samples were re-assayed using version two. So, use CA125_LEVEL0-5 to get the result of a screen from a clinical perspective of what the participant was told following their screening visit. Use CA125II_LEVEL0-5 to get the result of a screen from an epidemiologic perspective, with all values coming from the same assay.

BROOKLYN AND ALABAMA

When the trial began, there was a study center located in Brooklyn, New York. In 1998, the Brooklyn center was closed. There are no data from these participants in the datasets. Shortly thereafter, the site was replaced by the center at University of Alabama at Birmingham. At this new site, an added emphasis was placed on recruiting minorities, notably African-Americans.

5. CANCER SITE SPECIFIC CONSIDERATIONS

LUNG

CARCINOIDS

The PLCO trial does not consider carcinoid lung cancer to be a target of lung cancer screening and therefore participants with confirmed lung carcinoids officially have “no confirmed lung cancer.” These individuals can still be identified using the variable LUNG_CARCINOID.

COLORECTAL AND COLORECTAL ADENOMA

CARCINOID, APPENDIX AND *IN SITU* CANCERS

Carcinoid colorectal cancer is considered to be a target of colorectal cancer screening, and is thus included in the definition of confirmed colorectal cancers. They can be identified with the variable COLO_CANCER_TYPE. Appendix cancers are not considered confirmed colorectal cancers; they can be identified with COLO_APPENDIX_CANCER. *In situ* cases, only collected during the early years of the trial operation, are not considered colorectal cancers; time at risk for colorectal cancer is censored at *in situ* for a few participants.

POLYPS

The trial obtained polyp data only from endoscopic follow-up to positive screening FSGs or from diagnostic procedures in the year prior to colorectal cancer diagnoses. Polyp data is available in 3 of the colorectal datasets: the primary Colorectal dataset summarizes polyp findings following positive T0, T3 and T5 screens; the Colorectal Endoscopies dataset summarizes polyps from each recorded endoscopy; and the Colorectal Polyps dataset represents individual polyp information when available. From trial inception to 2000 any polyp was recorded, but from 2000 onward, only adenomas were documented.

A polyp is considered a left-sided (distal) polyp if found in the rectum, sigmoid colon or descending colon, and right-sided (proximal) if found in the splenic flexure, transverse colon, ascending colon or cecum. If no anatomic location was specified, the polyp is considered left-sided if medical records indicated a distance less than 50 cm from the anal verge. This distinction was made because this was the threshold distance for which an FSG was considered adequate.

An advanced adenoma is an adenoma that was villous, tubulovillous, had severe dysplasia, or was greater than or equal to one centimeter in size. *In situ* and borderline malignant carcinomas are also considered advanced adenomas. Variables that begin with COLSW_ summarize the most severe/advanced finding from that study year or endoscopy.

INCIDENT ADENOMA ANALYSES

The primary Colorectal dataset contains variables prepared for incident adenoma analyses. To be eligible for incident adenoma analyses, participants must have a negative T0 FSG and either a positive T3/T5 FSG with adenoma found on follow-up (for cases) or a negative T3/T5 FSG (for non-cases). Because eligibility is dependent on screening FSGs, the incident adenoma variables describe findings in the distal colon, the region inspected by the screen.

OVARIAN

EXTRA OVARIAN CANCERS

In PLCO, ovarian cancer is grouped along with primary peritoneal and fallopian tube cancers. All of their data have been combined in the ovarian datasets. To differentiate them, the variable OVAR_CANCER_SITE can be used.

LOW MALIGNANT POTENTIALS (LMPs)

LMPs (Borderline) are not considered invasive. While the trial considers LMPs to be ovarian cancers, they were reported separately in the main PLCO publications. LMPs can be identified with OVAR_CANCER_SITE = 4, 5 or 6.

OVARY REMOVAL

There are four ways that the trial learns about a female participant in PLCO having her ovaries removed. The primary source is the Baseline Questionnaire (BQ). All participants were asked upon entry if they had previous surgery on their ovaries and what was removed (OVARY_TRIAL_FLAG and OVARIESR_F). Second, for women in the intervention arm, before every screen for ovarian cancer, they were able to notify the trial that they no longer needed screens due to ovary removal. These data are captured in the variable OREM_FYRO which denotes the first screening year in which they reported ovaries removed. Third, the follow-up of all positive screens was recorded. Sometimes, this follow-up resulted in oophorectomies (PROC_NUMO=33 in the ovarian Procedures dataset). Lastly, each participant was asked her ovary status on the supplemental questionnaire (SQX). The SQX is only available through a special request.

For analyses using both arms of the trial, it is recommended that only the BQ data be used. When dealing with the intervention arm only, or doing a screening analysis, the screen data can be used. It is not advised when using a cohort to use the data from the follow-up of positive screens.

FEMALE BREAST

BREAST CANCER SUPPLEMENTAL (BCS) FORM

At the start of the trial, breast cancers were followed like all non-PLCO cancers. The only information received was a confirmation, a diagnosis date and an ICD-O-2 code. Beginning in 2007, the BCS form was implemented to capture many more details about the cancer including, but not limited to, stage, estrogen receptor (ER) status, progesterone receptor (PR) status and HER2 status. Prospectively, these data were collected at confirmation of diagnosis, and retrospectively they were collected for all previously identified breast cancer cases, both invasive and *in situ*.

BEHAVIOR CODES

There are two sources of behavior for breast cancers: the initial confirmation of breast cancer diagnosis and the surgery resulting in the definitive breast cancer information. These two sources do not always agree. Participants who do not have a BCS form and those with a BCS form but lacking surgery will only have the initial confirmation in the form of an ICD-O-2 code. BREAST_BEHAVIOR is the variable in use to definitively determine which cancers are considered *in situ* and which invasive. The source is the result from surgery where available, and when not available, the initial confirmation.

GLIOMA

In addition to gliomas, the glioma dataset contains other brain and spinal cord cancers. These can be identified with the variable GLIO_TYPE.

HEAD AND NECK

The head and neck datasets contain cancers from the following sites: lip, tongue, salivary gland, floor of mouth, nasopharynx, tonsil, oropharynx, hypopharynx, nose, nasal cavity, middle ear, larynx and other oral cavity and pharynx. The sub-site can be identified using the variable HNC_SEER. This file does not contain esophageal carcinomas. They can be obtained by requesting the upper GI dataset.

HEMATOPOIETIC

The hematopoietic datasets are an amalgam of all cancers relating to the blood and lymphatic systems. Standard sub-classifications such as non-Hodgkin's lymphoma, multiple myeloma and chronic lymphocytic leukemia are identifiable using the variables HEMA_SEER, HEMA_TYPE and HEMA_SUBTYPE. For further differentiation, HEMA_ICDMOR can be used. Lymphoid neoplasms can be identified with the variable IS_LYMPHOIDNEOPLASM.

PANCREAS

Both exocrine and endocrine pancreatic cancers are included in the pancreas dataset. Endocrine pancreatic cancers can be identified and excluded with the variable PANC_TYPE. They include ICD-O-2 morphology codes 8150, 8151, 8152, 8153, 8155, 8156, 8240, 8241, 8242, 8243, 8244, 8245 and 8246.

UPPER GASTROINTESTINAL

The upper gastrointestinal dataset contains all gastric and esophageal cancers collected in PLCO. There are three sets of variables pertaining to these cancers. All upper gastrointestinal cancers are represented in the variables with the prefix UPGI_. They are also represented either in the gastric set (GAST_) or the esophagus set (ESOPH_).

6. OTHER CONSIDERATIONS

BASELINE QUESTIONNAIRE (BQ)

The BQ was given to all participants upon enrollment and provides the background information on participants. 97% of the participants returned a questionnaire. All data on the BQ were self-reported. Every effort was made to clean these data and present them in a consistent manner.

QUALITY ASSURANCE (QA) EXAMS

For each of the qualitative screening exams (non-blood tests), roughly 3-5% of tests were given a QA reading. The types of readings vary by site and include a repeat of the examination, direct observation of the examination, and a review of the materials (films, videotapes or photographs). The overall result of the screen, of which the participant was notified, reflects the most severe of the initial and QA screens. Both examination results and findings are available in the screen and abnormality files.

SECOND SAME SITE PRIMARY

PLCO only collects the first of any given cancer site, and doesn't record any subsequent diagnoses from that site, including recurrence and progression. However, there are three sites of interest that have secondary datasets with second primaries. The datasets for these sites, including breast, hematopoietic and head and neck cancers, exist for different reasons. Chronologically, the first trial confirmed diagnosis is in the primary dataset, while the second is located in the secondary dataset.

The hematopoietic and head and neck datasets combine different cancers, and therefore it is possible for participants to have cancers in multiple sub-sites. The hematopoietic file represents six sites in terms of data collection: Hodgkin's disease; leukemia; lymph nodes; non-Hodgkin's lymphoma; multiple myeloma; and hematopoietic and reticuloendothelial systems. Head and neck cancers were grouped into three site categorizations when the data was collected: (1) Larynx; (2) lip, oral cavity and pharynx; and (3) nasopharynx, nasal cavity, middle ear and sinuses.

During the Breast Cancer Supplemental (BCS) form collection, instructions indicated that data collection should be done for all breast primaries, regardless of whether they were the first. See above in Site Specific Considerations for more information on the BCS form.

MENOPAUSE STATUS

PLCO participants were not directly asked if they were post-menopausal. Through the use of the various questions that were asked, the vast majority of participants can be confidently classified as post-menopausal, but some remain unclear. No one in the trial is classified as pre-menopausal. The source variables are age at last menstrual period (LMENSTR), reasons that periods stopped (MENSTRS), age at Baseline Questionnaire (ENTRYAGE_BQ), ovary removal status (OVARIESR_F), and hysterectomy status (HYSTER_F). The resulting post-menopausal variables are POST_MENOPAUSAL and MENSTRS_STAT_TYPE.

7. ACRONYM GLOSSARY

Below is a list of common acronyms found in the PLCO data and on the CDAS site.

| ACRONYM | DEFINITION | CATEGORY | USE |
|----------------------|--|---------------|---|
| AJCC | American Joint Committee on Cancer | Resource | Provides standards for staging. PLCO used the 5 th edition of the AJCC Cancer Staging Manual, but 7 th edition variables are available. |
| ASU | Annual Study Update | Questionnaire | Given to participants every year to ascertain whether they were diagnosed with cancer. |
| BCS | Breast Cancer Supplemental | Form | Captures female breast cancer characteristics |
| BMI | Body Mass Index | Medical | Anthropomorphic measure |
| BPH | Benign Prostatic Hypertrophy | Medical | Prostate condition |
| BQ (BQF, BQM) | Baseline Questionnaire (Female, Male) | Questionnaire | Provides background information |
| Bx | Biopsy | Medical | |
| CA-125 | Cancer Antigen 125 | Screen | Ovarian cancer blood screen |
| CDAS | Cancer Data Access System | Resource | Website from which PLCO data are downloaded |
| CDQ | Cause of Death Questionnaire | Form | Records the results of the death review |
| CSFII | Continuing Survey of Food Intakes by Individuals | Resource | Standard for calculating nutrient intake |
| DCF | Death Certificate Form | Form | Captures the death certificate |

| | | | |
|------------------|--|---------------|--|
| DE[X] | Diagnostic Evaluation for site [X] | Form | Provides cancer information for PLCO cancers. Tracks diagnostic procedures for cancers as well as positive screens |
| DHQ | Diet History Questionnaire | Questionnaire | Both arms of the trial. T0 for control arm, T3 for intervention arm. |
| DQX | Dietary Questionnaire | Questionnaire | Intervention arm only at T0 |
| DRE (DRQ) | Digital Rectal Exam (QA Exam) | Screen/Form | Prostate cancer screen |
| Dx | Diagnosis | Medical | |
| FSG (FSQ) | Flexible Sigmoidoscopy (QA Exam) | Screen/Form | Colon cancer screen |
| ICD-O-2 | International Classification of Disease for Oncology 2 nd Edition | Resource | All cancer data use these standards |
| ICD-9 | International Classification of Disease 9 th Revision | Resource | Catalogs non-cancer specific cause of death |
| LMP | Low Malignant Potential | Medical | Ovarian borderline tumor |
| MDF | Missing Data Form | Form | Fills in the gaps, letting the trial know why certain data are not available. |
| MPED | My Pyramid Equivalents Database | Resource | Standard for food intake measurement |
| MUQ | Medical Use Questionnaire | Questionnaire | Administered 2013 (post transfer to CDCC) |
| NDI | National Death Index | Resource | Queried yearly to identify deceased participants. Used to ensure complete ascertainment of death. |
| NDS-R | Nutrition Data System for Research | Resource | Standard for calculating nutrient intake |
| NOS | Not Otherwise Specified | Abbreviation | |
| NRF | Non-Response Form | Form | Keeps a record of participants who leave the trial |
| NSAID | Non-Steroidal Anti-Inflammatory Drug | Medical | Aspirin and ibuprofen |
| OCF | Other Cancer Form | Form | Captures non-PLCO cancer information |
| PLCO | Prostate, Lung, Colorectal and Ovarian | Study | Can refer either to the trial as a whole or specifically to these four cancers |
| PSA | Prostate Specific Antigen | Screen | Prostate cancer blood screen |
| QA | Quality Assurance | Abbreviation | Denotes replicated screens that assure consistently administered exams. |
| Rx | Treatment | Medical | |
| SEER | Surveillance Epidemiology & Endpoint Results | Resource | Used to classify cancers by site |
| SQX | Supplemental Questionnaire | Questionnaire | Background questionnaire administered in 2006 and 2007 |
| T[X] | Time Point [X] | Reference | Denotes years since randomization |
| TI[X] | Treatment Information for site [X] | Form | Records initial treatment information |
| TNM | Tumor, Nodal, Metastases | Medical | Determines the cancer stage |
| TRUS | Transrectal Ultrasonography | Medical | Prostate procedure |
| TURP | Transurethral Resection of the Prostate | Medical | Prostate procedure |
| TVU (TVQ) | Transvaginal Ultrasound (QA Exam) | Screen/Form | Ovarian cancer screen |
| XRY (XRQ) | Chest X-Ray (QA Exam) | Screen/Form | Lung cancer screen |