WHI LILAC Participants and Age-Matched Cancer-Free Cohort – Dataset Guide
Investigator Datasets
October 2023

Introduction

Life and Longevity after Cancer (LILAC) is a cancer survivor cohort embedded within the Women's Health Initiative (WHI). The primary focus of LILAC, when initiated in 2013, was to fill critical gaps in our information on cancer survivorship in post-menopausal women. We limited this work to eight cancers, hereafter referred to as LILAC cancers: invasive breast, colorectal, lung, endometrial, ovarian (including fallopian tube and primary peritoneum), melanoma, leukemia, and lymphoma. A description of study methods and characteristics obtained in the initial LILAC survey of consenting participants has been published (1). Women with a LILAC cancer who were no longer alive when the study was initiated were enrolled under a partial waiver of consent. For detailed information about datasets for LILAC cancer survivors, please see the companion document titled **WHI LILAC Cohort – Data Preparation and Use Guide.**

Identifying a cancer-free Cohort

To understand the diagnosis of cancer and its treatment on the trajectories of aging, the accelerated aging phenotype, and age-related comorbidities, we established a cohort of age-matched WHI participants with similar data who have remained cancer free. The rationale for this approach is that matching in cohort studies ensures that the exposed cohort (cancer patients in LILAC) and the unexposed cohort (the cancer-free WHI participants) will have the same distributions over categorical levels of the matching variables.

To be eligible for the cancer-free cohort, WHI participants had to be alive and free of cancer (other than non-melanoma skin cancer) at the time of the LILAC cancer diagnosis (risk-set sampling). The selected controls further remained cancer free until their death/end-of-follow-up/February 28th 2020, whichever occurred first. Participants with hysterectomy or bilateral oophorectomy were excluded from the eligible controls for LILAC participants with endometrial or ovarian cancer. No participant served as a control for more than one LILAC case.

We identified up to 5 cancer-free WHI participants matched to each LILAC cancer participant. Matching was completed on the factors below.

Matching factors:

- Age at WHI screening (+/- 1 year)
- WHI enrollment date (+/- 30 days for the initial identification of matched controls; criterion was relaxed to +/- 90 days in order to match controls to 45 LILAC participants)
- WHI Study Component: Clinical trial (CT) or Observational Study (OS)
- Hormone Trial (HT) enrollment
- Long Life Study (LLS) enrollment

The last three factors, defining the enrollment status in the various WHI components, were used to assure comparable data collection for cancer cases and their matched cancer free-controls. See the WHI website (Working with WHI Data) for additional information.

Matching results:

Among 13,412 LILAC cases and 114,873 potential controls, a total of 66,144 controls were matched to 13,405 cases:

- 12,956 cases (96.60%) with 5 controls each
- 196 cases (1.46%) with 4 controls each
- 123 cases (0.92%) with 3 controls each

- 81 cases (0.60%) with 2 controls each
- 49 cases (0.37%) with 1 control each
- 7 cases (0.05%) had no matched controls

Database for analyses of LILAC participants and cancer-free cohort

To facilitate data analyses, we developed a database populated with extensive demographic, physical, mental, social health, and clinical event data collected under the WHI protocol based on this common control group and standardized variables collected across time and data collection forms.

For confidentiality reasons, no calendar dates are included. The diagnosis date defines a key time point for each LILAC participant and her matched cancer-free controls, referred to here as the index date and coded as time zero. The timing of other data elements is calculated as days from the index date. Negative values indicate pre-index date events and positive values represent post-index date events.

Description of individual datasets in the LILAC database

The first variable in each file (ID) is the unique WHI participant ID, the same variable used in the files used for the parent study. All files are linked by this identifier. In addition, each dataset includes: case/control flag, matched set number, control number within matched set, WHI CT participant flag, LLS participant flag, LILAC-designated cancer.

For datasets that include items collected over time, variables are included that provide the number of days from diagnosis (LILAC cases) or reference date (matched controls) to the collection time point.

Images of WHI questionnaires are available on the WHI website. LILAC study forms are available at https://www.whi.org/md/370/home.

LILAC participants and matched controls are included in datasets numbered 1 – 7 below. Datasets for items 8-10 include only LILAC participants.

1. **Demographics (***lilac_demographics_inv.dat***)**
   Age at WHI screening, flags for WHI and ancillary study participation, days from WHI enrollment to LILAC diagnosis, age at index date (diagnosis/reference), race and ethnicity, education, family income, US region, marital status, and insurance status at WHI baseline. One row per woman.
2. **Blood draws and urine collection (***lilac_blood_urine_inv.dat***)**
   Specimen type, days from diagnosis/reference to blood draw or urine collection, WHI visit year and type, LLS draw flag. One row per draw or collection for each woman.
3. **Quality of life, physical function, and psychosocial assessments (***lilac_quality_of_life_inv.dat***)**
   Extensive data on activities of daily living, general health, and emotional health, physical function, quality of sleep, stressful life events, and social support. One row per assessment across WHI Forms 37, 38, 151, 155, 157, 159, 151b.
4. **Objective physical measurements (***lilac_physical_measures_inv.dat***)**
   Systolic and diastolic blood pressure, height, weight, waist, and hip measurements. Calculated body-mass index, waist/hip ratio. One row per assessment across WHI Form 80 or Form 301 (LLS home visit).
5. **Personal habits (***lilac_personal_habits_inv.dat***)**
   Alcohol intake, smoking, recreational physical activity. One row per assessment across WHI Forms 34 and 35, 143-148 (OS only), 155, and LILAC Form 370 (LILAC participants only).
6. **Clinical history at WHI enrollment (***lilac_clinical_history_whi_enrol_inv.dat***)**

Self-reported conditions at WHI baseline (from WHI Forms 2 and 30) including fracture, myocardial infarction, stroke, diabetes, hypertension, and a WHI-adapted Charlson comorbidity index. One row per woman.

7. **Clinical events during WHI follow-up (***lilac_clinical_events_inv.dat***)**
   Adjudicated or self-reported fractures, myocardial infarction, stroke, diabetes, or hypertension, plus other chronic conditions that might impact aging. Time to each of these events is from diagnosis/reference to the first adjudication/self-report of each condition and can be used to construct time-dependent ever/never variables for each condition. Time from diagnosis/reference to end of continuous follow-up/death for participants is included. One row per woman.
8. **Clinical characteristics of cancer for LILAC participants (***lilac_cancer_characteristics_inv.dat***)**
   Age at cancer diagnosis, site code, histology, SEER summary stage, ER, PR, and HER2/neu receptor status for breast cancer cases, summary variables for cancer treatment (surgery, chemotherapy, radiation), flags indicating medical record abstraction, Medicare enrollment at diagnosis, tissue collection eligibility and availability of tissue. One row per cancer outcome and participant. Information about SEER coding of tumor characteristics can be found in the WHI data preparation guide.

For specific details about the following datasets, please see the companion document in this submission titled **WHI LILAC Cohort – Data Preparation and Use Guide.**

9. **LILAC baseline and follow-up questionnaires for LILAC participants**
   - LILAC baseline questionnaire (Form 340) was completed by living participants who consented to the study. Some participants completed questions 1-4 and 6 only, by phone. Data collected includes symptoms after treatment completion, pain, depression, anxiety, fatigue, distress, social support, weight, marital status, and insurance coverage.
   - LILAC Form 370 was the first annual survey completed after the baseline survey. It includes variables for weight and weight loss; selected medications; financial and cancer worry; peripheral neuropathy; cognitive and physical functioning; selected symptoms.
   - Additional measures collected in the second annual follow-up questionnaire (Form 371), include: depression, anxiety, fatigue and distress; and unmet needs of cancer survivors (e.g., pain, physical functioning, memory/concentration, weight changes, end of life planning).

10. **Medical record abstraction for LILAC participants**
    Treatment information was obtained through Medicare linkage for participants who were continuously enrolled in Medicare fee for service (FFS) A+B at diagnosis and for one year after diagnosis. Per the agreement with CMS, these data cannot be released. Investigators may apply to use the CMS data in the WHI Virtual Data Enclave. For the remaining LILAC women whose cancer diagnoses occurred in the year 2000 or later, and who either consented to medical records release or were deceased and enrolled in LILAC under partial waiver of consent, medical records were retrieved. From these, data on cancer treatment, molecular testing, and cancer recurrences were abstracted. For women with double primaries (two LILAC cancers diagnosed on the same day), two abstraction forms were completed, one for each cancer.

Questions about this resource may be submitted to the WHI Help Desk or sent to helpdesk@WHI.org.

(1) Paskett, E.D., Caan, B.J., Johnson, L., Bernardo, B.M., Young, G.S., Pennell, M.L., Ray, R.M., Kroenke, C.H., Porter, P.L. and Anderson, G.L., 2018. The Women's Health Initiative (WHI) Life and Longevity

After Cancer (LILAC) study: description and baseline characteristics of participants. Cancer Epidemiology and Prevention Biomarkers, 27(2), pp.125-137.