NATIONAL CANCER INSTITUTE
Division of Cancer Epidemiology & Genetics

connect for cancer prevention study

Connect for Cancer Prevention:

A Prospective Cohort Study within Integrated Healthcare Systems in the US

EXECUTIVE SUMMARY

November 12, 2020

Principal Investigator: Montserrat Garcia-Closas, M.D., Dr.P.H.

Steering Committee: Christian C. Abnet, Ph.D., M.P.H.; Paul S. Albert, Ph.D.; Jonas S. Almeida, Ph.D.; Amanda Black, Ph.D., M.P.H.; Michelle Brotzman, M.P.H.; Mia M. Gaudet, Ph.D.; Nicole Gerlanc, Ph.D.; Amy Berrington de Gonzalez, D.Phil.; Stephanie Weinstein, Ph.D.; Nicolas Wentzensen, M.D., Ph.D., M.S.; Hannah Yang, Ph.D., Sc.M.

Executive Committee: Habib Ashan, M.D., University of Chicago; Chun-Hung Chan, Ph.D., Sanford Health; Heather Feigelson, Ph.D., Kaiser Permanente – Colorado; Jennifer Gander, Ph.D., Kaiser Permanente – Georgia; Katrina Goddard, Ph.D., Kaiser Permanente – Northwest; Robert Greenlee, Ph.D., M.P.H., Marshfield Clinic; Stacey Honda, M.D., Kaiser Permanente – Hawaii; Carmit McMullen, Ph.D., Kaiser Permanente – Northwest; Pamala Pawloski, Ph.D., Health Partners; Benjamin Rybicki, Ph.D., Henry Ford Health System

Coordinating Center: Division of Cancer Epidemiology and Genetics, National Cancer Institute

Point of Contact: Mia M. Gaudet, PhD, Senior Scientist, E-mail: mia.gaudet@nih.gov

NIH-IRB Protocol number: # 000034

Date of NIH-IRB Approval: October 23, 2020

Overview

The Connect for Cancer Prevention Study ("Connect") is a prospective cohort study of 200,000 adults participating through integrated health care systems in the US (Figure 1). The primary objective is to study the etiology and natural history of cancer, and to inform new approaches in precision prevention and early detection of cancer. To this end, participants aged 40-65 years and cancer-free at recruitment will be enrolled and followed up for at least 25 years to study cancer incidence and mortality. Recruitment efforts will promote participation of individuals from diverse backgrounds and geographic areas. During follow up, participants will provide information on a wide range of possible risk factors, including behavioral, lifestyle and environmental, through serial surveys, as well as serial biological specimens to study biomarkers of risk and early detection for cancer. In addition, participants will provide access to their electronic health records (EHR), as well as cancer precursor and tumor tissue specimens for those diagnoses with these diseases. Connect will become a rich trans-NCI research resource to advance cancer research by intramural and extramural scientists in the US and internationally.

Key features

Connect is designed to securely and efficiently collect serial health information through surveys and EHRs together with serial biospecimens, to advance the study of suspected and emerging factors that influence cancer development. Biospecimen collections will include blood, urine, saliva and other materials to study changing exposures/biomarkers and evaluate critical exposure windows before and after disease diagnosis. Connect will leverage recent developments in digital technologies, biomarkers and exposure assessments including:

Study Design Features for Connect for Cancer Prevention

- Cloud-hosted data system with modern interoperability standards for secure and efficient data sharing and powerful analytics following F.A.I.R (Findable, Accessible, Interoperable, Reusable principles, enabled through the NIH STRIDES program.
- Widespread use of electronic devices and applications in the digital age that facilitate engagement of study participants to provide and receive health information.
- High-quality exposure assessment using innovative technologies such as wearable devices to measure behavior and environment.
- Increased availability and quality of EHR and medical imaging for assessment of health conditions and medications.
- Increased availability and quality of databases for data linkages, including pollution monitoring, pharmaceutical records and cancer registries.
- Technological advances to interrogate the genome, epigenome, transcriptome, proteome, metabolome, microbiome and other biological processes, using of serial biological specimens
- Digital image analyses/AI and molecular/genomic profiling of tumors and precursor lesions to study the natural history of cancer and etiologic heterogeneity across tumor subtypes.
- Digital analyses /AI of medical images characterizing normal and diseased organs.

These technologies coupled with new methods in complex analytics of integrated high-dimensional data will provide powerful tools to address the primary objectives of the Connect for Cancer Prevention Study.

In addition, linkage to health (e.g. mortality and tumor registries) and exposure monitoring databases (e.g. geospatial analyses of air pollution) for exposure and health outcomes assessment will greatly enlarge the data resources available from this study.

Primary Aims

Connect will be a large trans-NCI research resource for the scientific community to address scientific questions related to:

- **Etiology of cancer:** identify and characterize biological, behavioral and environmental risk factors, and the interactions among them, associated with the incidence of different cancers. Serial exposure and biomarker assessments prior to cancer diagnosis will allow to study how changes over many years could influence cancer development and to identify critical exposure periods.
- **Natural history from precursor to tumor transformation:** The availability of repeat biospecimens, tissue samples, and imaging data from participants in Connect will provide an opportunity to characterize and identify biomarkers of transformation from precursors to cancer in a large population.
- **Cancer risk assessment:** Connect will be instrumental in the further development and validation of risk prediction models for precision prevention through the integration of risk factors from multiple sources (e.g., questionnaires, EHR, imaging, personal monitors, germline genetics, clonal hematopoiesis and biomarkers in blood, urine, saliva, tissue and other specimens like stool, accounting for changes over time and more precise definitions of endpoints (e.g., molecular subtypes of cancer). Integral to these risk prediction efforts, the study questionnaire has been developed to provide relevant risk factor information for many cancer endpoints. As new risk factors are identified, they can be integrated into follow-up questionnaires to allow for continuous updating and improvement of cancer risk prediction models within integrated health care systems.
- **Early detection of cancer:** Repeated biomarker measurements using specimens collected prior to diagnosis of cancer from Connect participants will provide extensive opportunities to identify biomarkers and develop algorithms for the early detection of cancer based on the trajectories of high-dimensional biomarkers. Specifically, studies of the development of marker discovery, assessment of clinical performance for detection of disease, evaluation of detection windows and lead time for specific markers and cancer sites, as well as comparative studies of cancer detection for several early detection approaches are expected in Connect.

Although Connect will be able to address a wide range of scientific questions, as for any epidemiological study, replication of findings in independent study populations and pooling data with other studies will be essential to confirm findings or address questions that require larger sample sizes (e.g. for the study of rare cancers or exposures). To this end, the cohort has been designed to facilitate data sharing and future data pooling efforts. Connect will be a large and comprehensive data resource and biospecimen repository also available for general research use.

Expected Diagnoses of Precursors and Cancer

Approximately, 14,000 participants (Table 1) are expected to develop cancer within the first fifteen years of follow up. This illustrates the power to address questions related to cancer, particularly common cancers such as prostate, breast, lung and colorectal. Precursor lesions are expected to accrue at a faster rate providing early outcomes for cancer-related publications. For example, 2,339 Connect participants are expected to have a benign breast biopsy in the 10 years after enrollment and an additional 4,942 participants with retrospective biopsies.

Resource Sharing Principles for Research

The data systems, infrastructure and policies for the Connect for Cancer Prevention Study are designed with data sharing in mind, to ensure extensive and appropriate use of the resource according to F.A.I.R. principles and the NIH data sharing policy. Safeguards for privacy and confidentiality will be in place to protect participants' data and biospecimens.

To improve data/code traceability and reproducibility of analyses and results, Connect data access will be granted through a cloud-hosted Connect for Cancer Prevention Data Platform. The Data Platform will facilitate access to raw and derived data, metadata and annotated code to all platform users. There will be different levels of access (Figure 2) to data determined by the level of sensitivity (public, registered and controlled access), as well as controlled access to depletable biospecimen resources.

DCEG is working in coordination with other NCI divisions and centers, in particular the Division of Cancer Prevention (DCP) and Division of Cancer Control and Population Sciences (DCCPS) as well as other divisions, to ensure a broad use of the Connect research resource through extramural funding opportunities. Although the primary endpoints of interest for this cohort are cancer incidence and mortality, the infrastructure has been designed to enable collaborative studies of general research use and ancillary enhancements, which will open opportunities for collaboration with other NIH institutes, who could provide partial funding.

Timeline

Recruitment through the partnering health care institutions will occur between 2021 and 2026. Serial questionnaires, biospecimen collection, EHR data collection, data linkages and passive follow-up activities will occur during the duration of the study expected to be for decades. Data and biospecimens for research use will be available for request after participant recruitment.

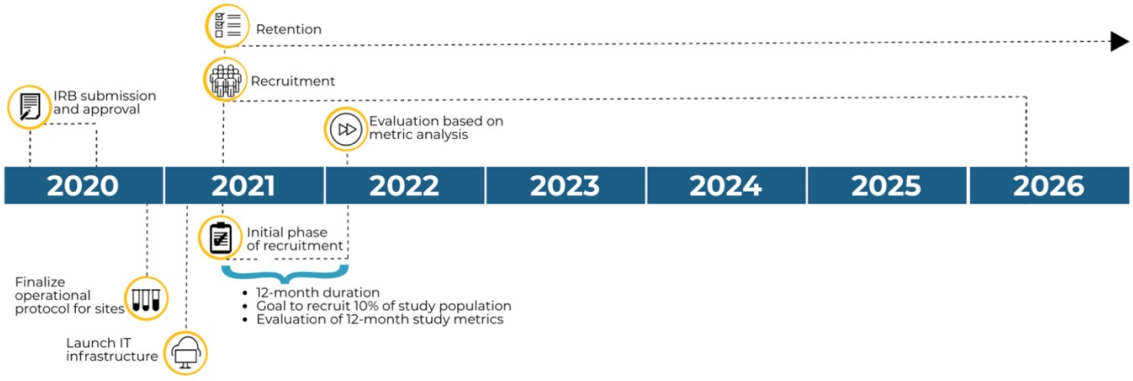

Figure 1: Location of DCEG Coordinating Center, Participating Integrative Health Systems and Support Services Contractor

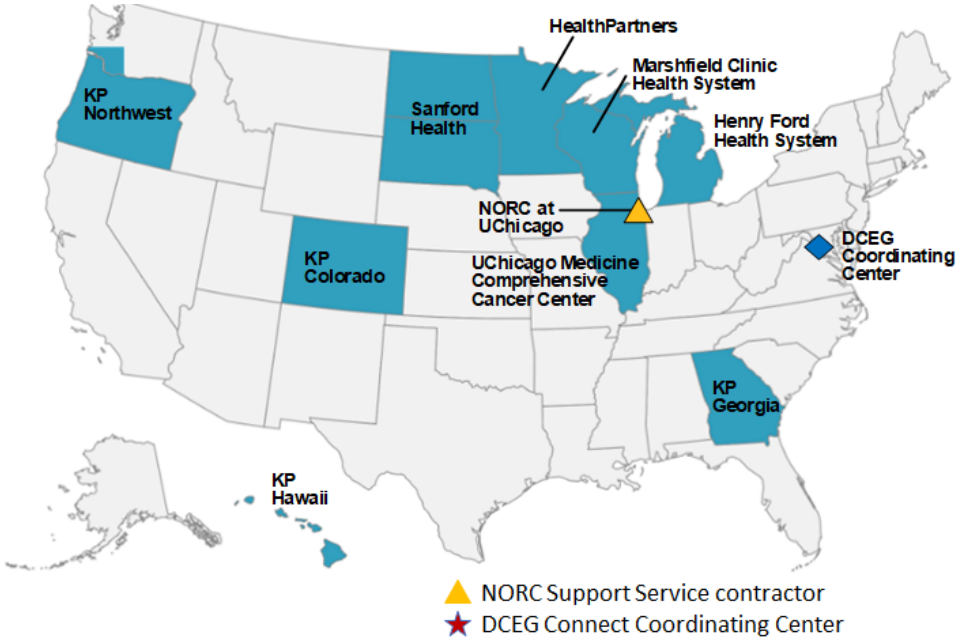

Figure 2: Levels of access to resources from the Connect for Cancer Prevention Study

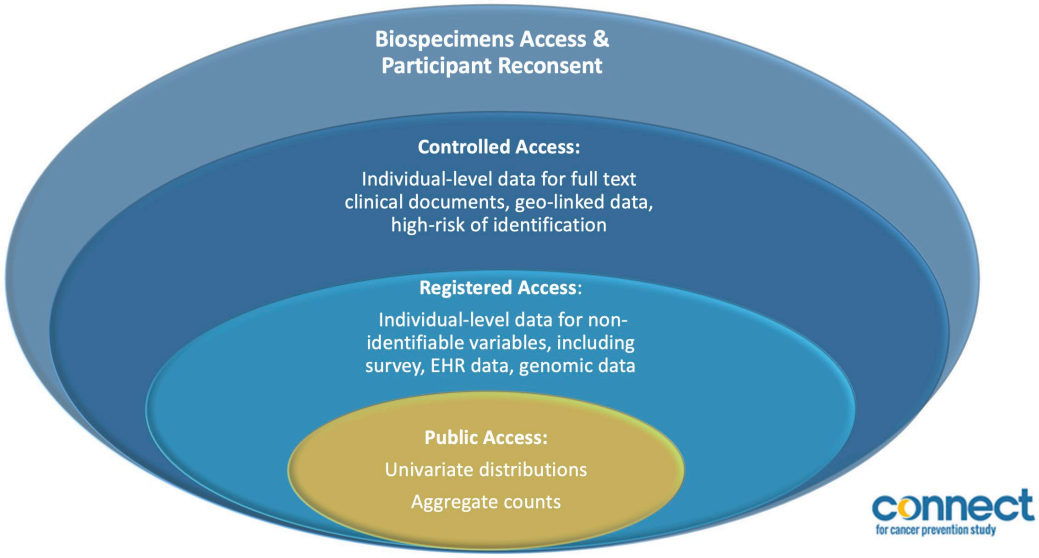

Table 1: Expected number of incident cancer diagnoses in the Connect for Cancer Prevention Study after 10 and 15 years of follow up

Cancer site	Total Number of Expected Incident Cancers	
	10 years of follow up	15 years of follow up
Prostate (males only)	1,951	3,595
Breast (females only)	1,566	2,553
Lung & Bronchus	1,151	2,165
Colorectum	877	1,566
Melanoma of the skin	410	681
Non-Hodgkin lymphoma	369	649
Uterine corpus	345	579
Urinary bladder	327	622
Thyroid	232	356
Pancreas	221	410
Leukemia	203	365
Ovary	146	246