



Cohort Overview

Mia M. Gaudet, Ph.D.

Connect Senior Scientist

Trans Divisional Research Program
Division of Cancer Epidemiology and Genetics
National Cancer Institute



Last Updated:
April 19, 2024

Study Objectives

Comprehensive research resource using new technologies and methods for the scientific community to study:

Cancer Etiology

Precursor to Tumor Progression

Cancer Risk Prediction

Early Detection of Cancer

Survivorship

Connect Designed to Address Key Research Priorities



Emerging exposures



Novel biomarkers, genomics



Cutting-edge methodology

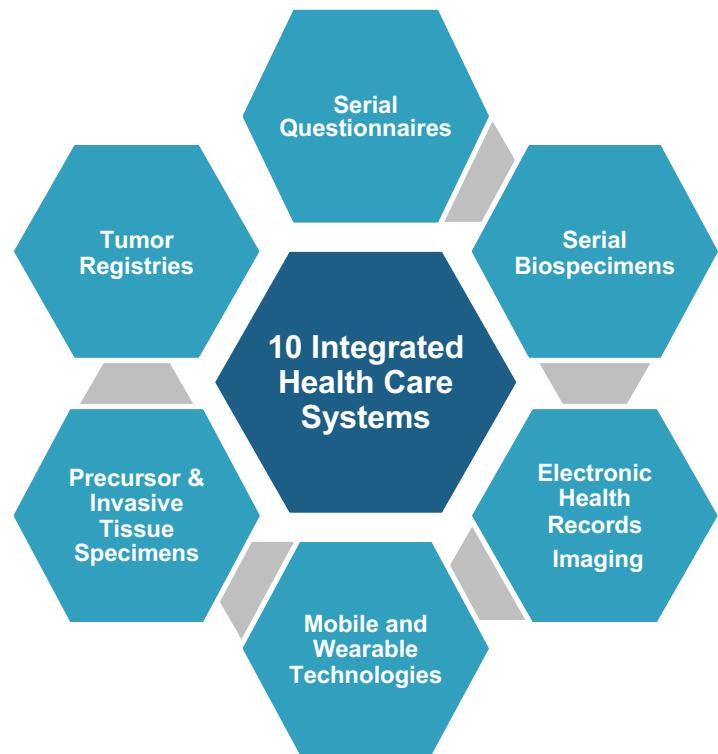


Diverse and special populations

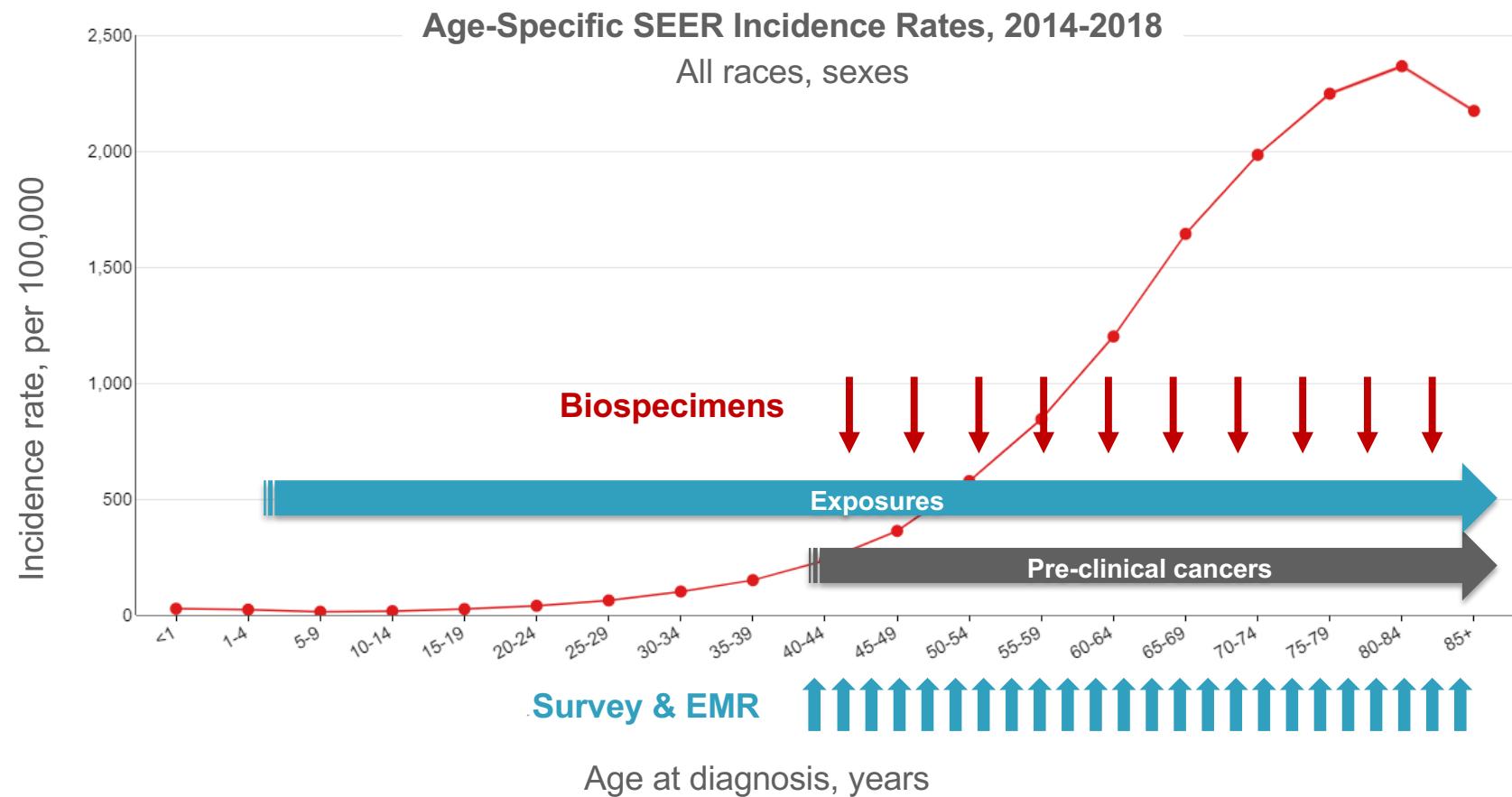
Cohort Study Design Overview & Recruitment Update

Key Study Design Features of Connect

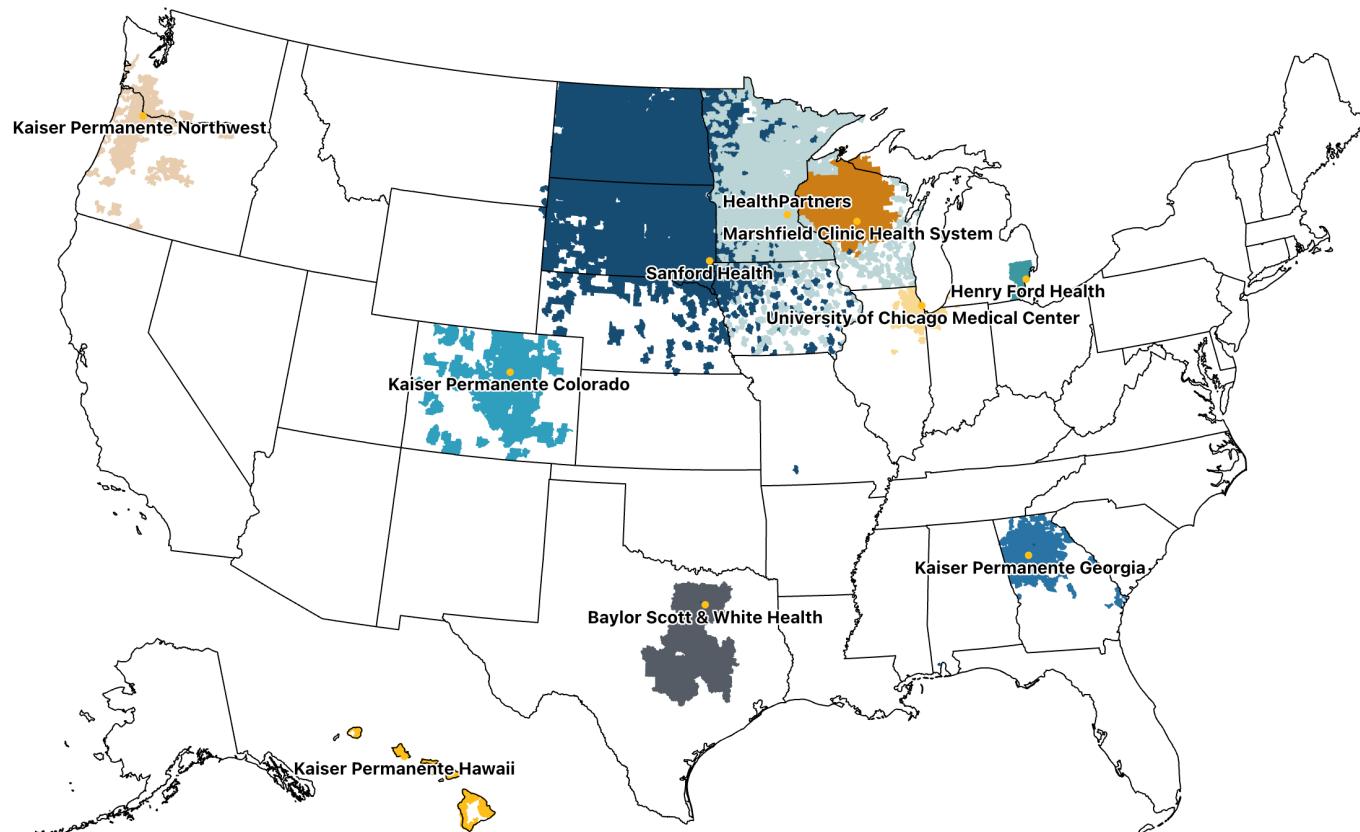
- ▶ 200,000 adults across the US
 - ✓ Aged 30-70 years
 - ✓ No history of cancer
 - ✓ Patients or members of partner health care systems
- ▶ Defined catchment population
- ▶ Long-term follow-up
- ▶ Comprehensive cancer and precancer outcomes
- ▶ English and Spanish **(by Summer 2024)** content
- ▶ Flexible infrastructure for enhancement studies



Serial Assessments to Observe Longitudinal Changes



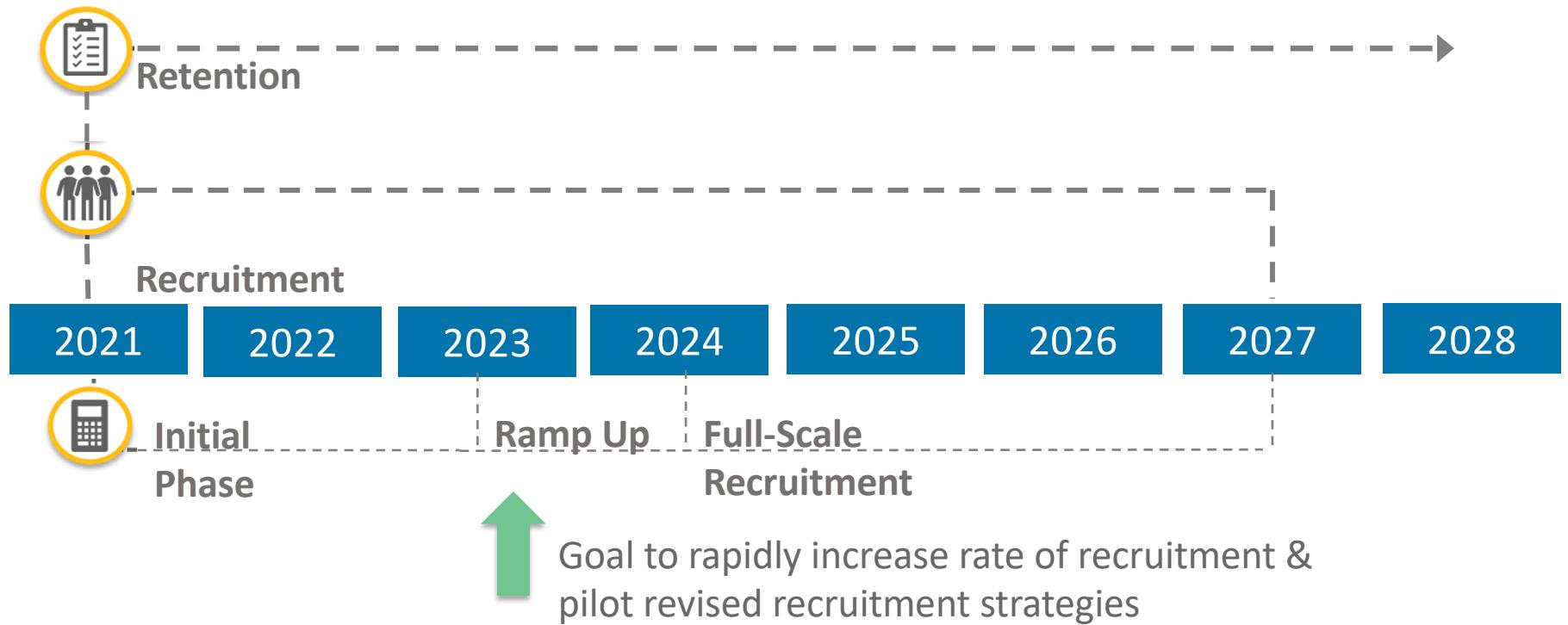
Recruitment through 10 U.S. Integrated Health Care Systems



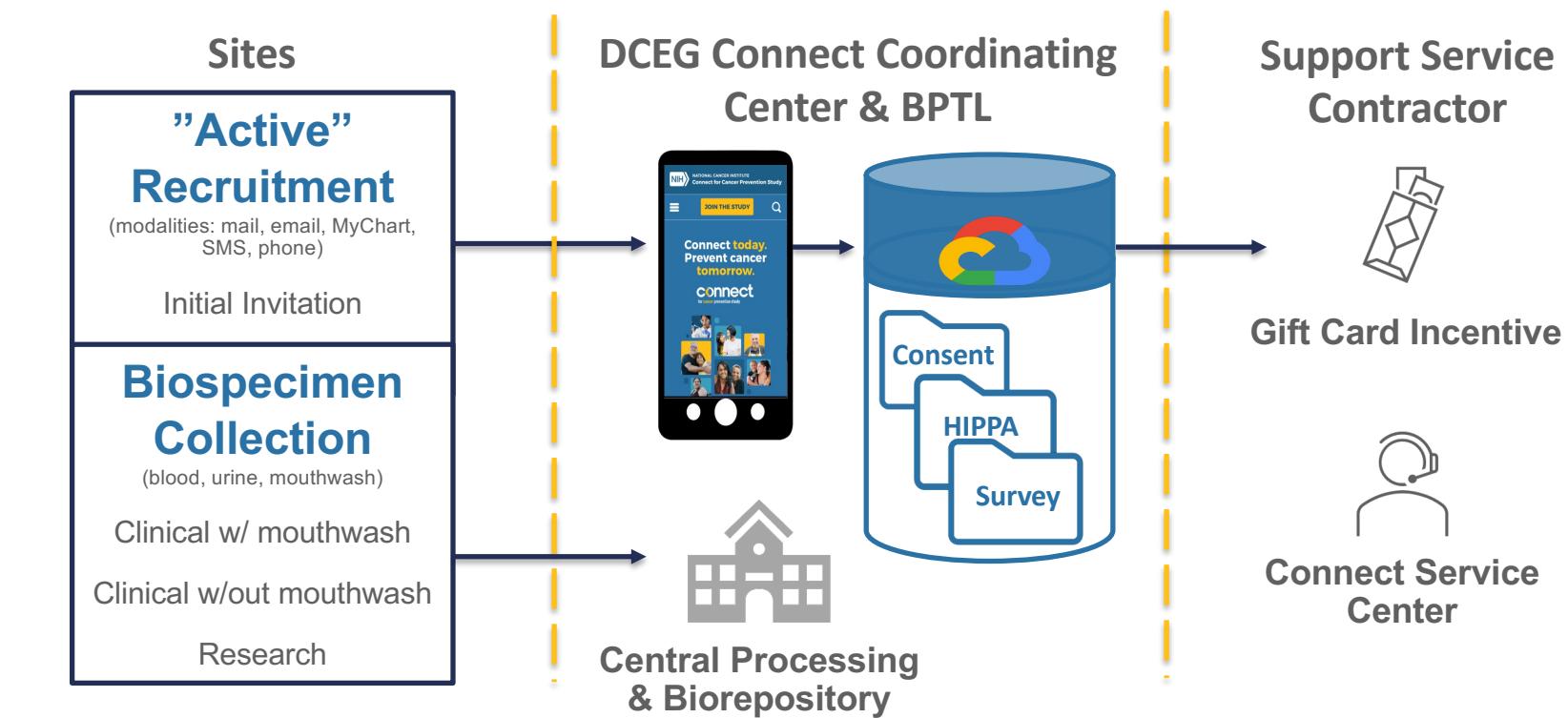
connect
for cancer prevention study

*Baylor Scott & White Health recruitment funded by Cancer Prevention and Research Institute of Texas

Scaling up recruitment



Overview of Recruitment and Biospecimen Collection



connect
for cancer prevention study

Participant Engagement

Recruitment Materials & Strategy

- Value proposition to anchor communications
- Prevention-focused messaging
- Materials designed to engender trust, enthusiasm, and inclusivity
- Monetary and non-monetary incentives
- Active and passive recruitment

**Connect today.
Prevent cancer
tomorrow.**

connect
for cancer prevention study

Retention Strategies

- Candid messaging about participant burden for the long-term
- Responsive communication using multiple channels
- Incentives for retention, such as return of results



Ongoing Engagement Strategies

- Digital engagement
- Regular contact every 6 months beyond those for study activities
- “Study anniversary” mailing
- Monitor real-time metrics to pivot engagement strategy



MyConnect

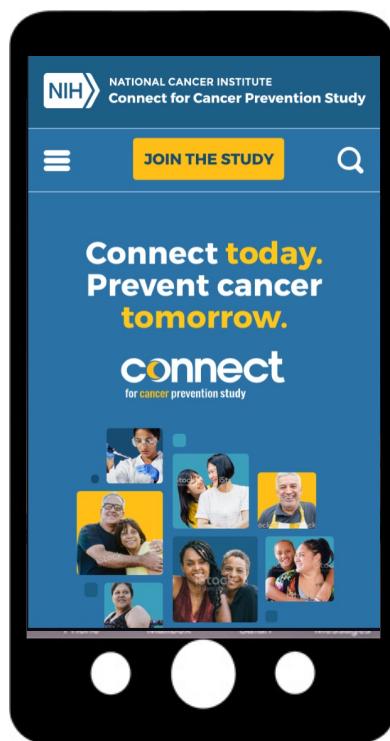
MyConnect: Communication and Study Activity Hub

Existing Functionality

Reminders &
Communication



Study Activities



Planned Functionality

Return of
Results



Mobile/ Wearable
Technologies



connect
for cancer prevention study

Educational Module before Long Form Consent

The screenshot shows the homepage of the Connect for Cancer Prevention Study. At the top, there's a navigation bar with links for Surveys, Samples, Agreements, About the Study, What to Expect, Privacy, My Profile, and Messages. On the right side of the header is a 'Sign Out' button. Below the header, a progress bar shows steps 1 through 8: About (selected), Activities, Privacy, Leaving, Results, Benefits, Consent, and Thank You. Step 1 is highlighted with a blue circle and a thick blue line under the progress bar. The main content area has a dark blue header with the NIH logo and 'NATIONAL CANCER INSTITUTE Connect for Cancer Prevention Study'. The first section, 'What is the Connect for Cancer Prevention Study?', contains text explaining the study's purpose and duration. Below this, another paragraph discusses researchers' goals and the nature of the study. A yellow 'Next' button is located at the bottom right of the content area.

The screenshot shows the 'Consent' module of the Connect for Cancer Prevention Study. At the top, it features the NIH logo, 'NATIONAL CANCER INSTITUTE Connect for Cancer Prevention Study', and a 'MENU' button. Below this is a progress bar indicating step 7 of 8, labeled 'Consent'. The main content area has a dark blue header with the title 'Informed Consent'. The text explains the process of joining Connect, which includes giving informed consent by reviewing the full consent form and electronic health records release form, and signing your name. Below this, another section titled 'Informed consent form' provides a detailed explanation of what it means to take part in Connect. At the bottom, there's a link with a download icon labeled 'Download an unsigned copy of the consent form'.

Messages and Data Input

Messages

New 2 Read 1

Complete Your Profile 6 days ago

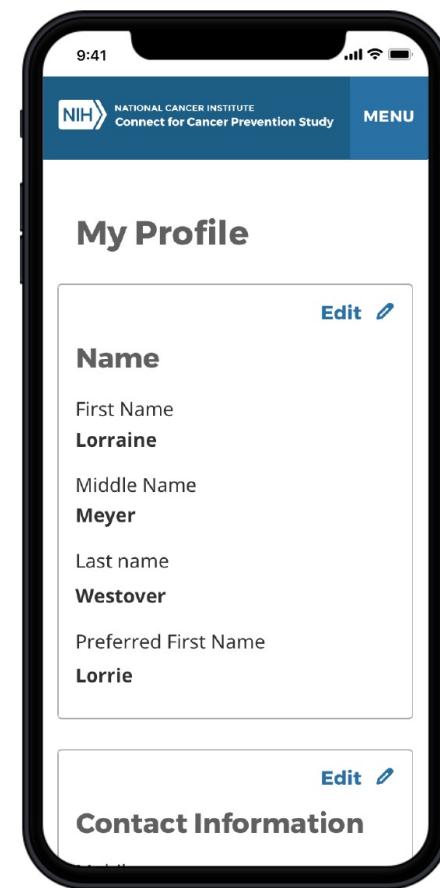
Please complete your profile section so you can move on to surveys!

Start

Sign Your Consent Forms 8 days ago

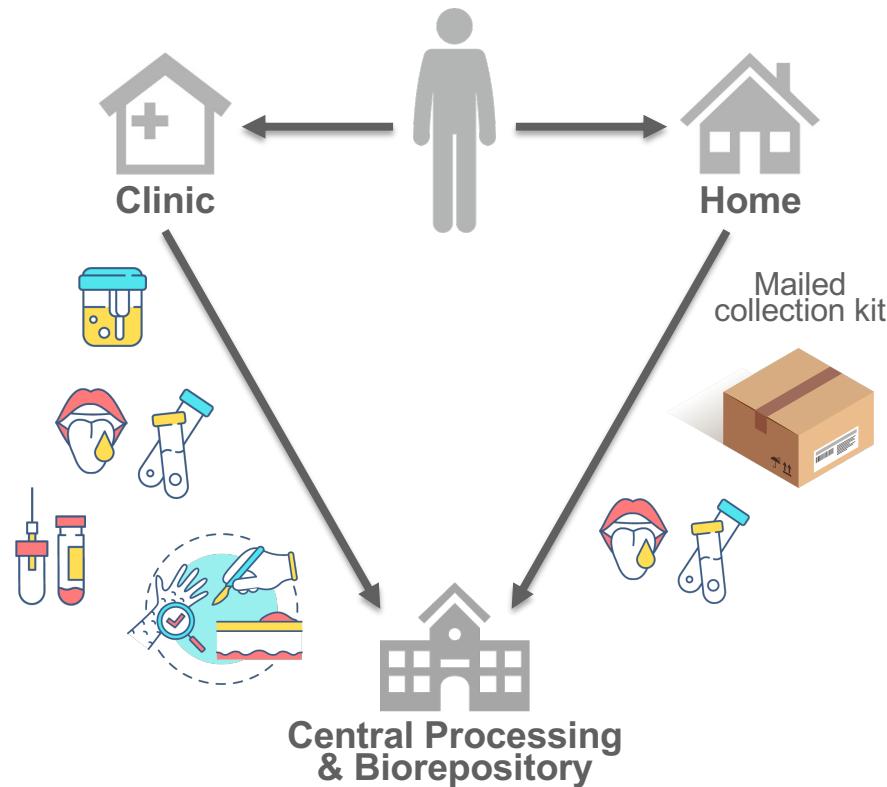
If you would like to participate in this study, please sign the consent forms.

Start



Biospecimen Collections

Biospecimen Collections at Recruitment



	No. Tubes	Fraction	Tube Type
Buccal	1		Saliva-mouthwash kit
Blood	2	Serum	10ml Red top
	1	Plasma/buffy coat	10ml Lavender top (EDTA)
	1	Plasma/buffy coat	10ml Green top (Heparin)
	1	Whole blood	10ml ACD Streck
Urine	1		10ml Urine kit
Tissue			Precursor tissue blocks, if available

Biospecimen Collections at Follow-up

- Timed on average every three years
 - More frequent for participants at higher risk of cancer
- Series of same biospecimens and new types of biospecimens (e.g., liquid biopsy, stool)
- Located at healthcare system, external clinic, home
- Paired with questionnaires and other relevant data collection

Surveys

Baseline Surveys



Content available on [Connect GitHub](#)

First Survey

This survey is split into four sections that ask about a wide range of topics, including information about your medical history, family, work, and health behaviors. You can answer all of the questions at one time, or pause and return to complete the survey later. If you pause, your answers will be saved so you can pick up where you left off. You can skip any questions that you do not want to answer.

	Background and Overall Health Questions about you, your medical history, and your family history.	Start
	Where You Live and Work Questions about places where you have lived and worked, and your commute to school or work.	Start
	Medications, Reproductive Health, Exercise, and Sleep Questions about your current and past use of medications, your exercise and sleep habits, and your reproductive health.	Start
	Smoking, Alcohol, and Sun Exposure Questions about your use of tobacco, nicotine, marijuana, and alcohol, as well as your sun exposure.	Start

Surveys Accompanying Biospecimen Collection

Blood & Urine

- Recent exposures to food, drink, pharmaceuticals
- Reproductive Health
 - Last menstrual period date

Mouthwash

- Oral health and care
- Recent exposures to toothpaste, mouthwash, gum, tobacco
- Permanent teeth lost

COVID-19

- Exposure
- Complications
- Long-COVID
- Vaccination

Menstrual Cycle

- Next menstrual period date



Surveys Available in MyConnect

Baseline	Background & Overall Health	Demographics, medical history, family history of cancer, education & occupation
	Medications, Reproductive Health, Exercise, Sleep	OTC/rx, menstrual cycle, pregnancy, fertility, hormonal meds & contraception, physical activity, shiftwork
	Smoking, Alcohol, Sun Exposure	Tobacco/nicotine, marijuana, alcohol, tanning
	Where You Live & Work	Residential, seasonal, & childhood addresses, water source, tap water intake, work & school commute
	SSN	Full or partial SSN
Biospecimens	Blood, Urine, Mouthwash (Research Appt)	Symptoms/behaviors on day of donation, meds/antibiotics, recent reproductive health, oral hygiene & health
	Blood/Urine (Clinical Collection)	Symptoms/behaviors on day of donation, meds, recent reproductive health
	COVID-19	# Infections, symptoms, hospitalization, vaccination
	Menstrual Cycle Survey	Last period, triggered by research appt/clinical collection response
Follow up	Quality of Life Survey	Physical function, anxiety, depression, fatigue, sleep disturbances, pain interferences, social roles, social isolation, cognitive function

Access to human- and machine-readable copies on [Connect Wiki](#)

Surveys In Development

Participant Experience	MyConnect, study communications, activities completed, overall experience, motivation for joining
Cancer Screening History	Organ Inventory (born with/current), history of cancer screening tests
Cancer Diagnosis Surveys (17)	Dx, symptoms, Patient-Provider interaction, medical history repeat assessment
Menstrual & Intimate Care Products	Vaginoplasty, powder, douching, vaginal cleansing products, menstrual products
Fecal Collection	Donation, bowel movements, meds, supplements, probiotics
Social & Structural Determinants of Health (SSDOH)	Discrimination, police interaction, medical mistrust, social support, and financial, food, and housing insecurity
Hair Products	Dyes, relaxers, straighteners, perms, oils
Menstrual Experience Survey	Menstrual problems, endometriosis dx and treatment
Mothballs & Scented Products	Household exposure to p-DCB & Napthalene

Detailed Status Available [Here](#)

Planned Survey Cadence

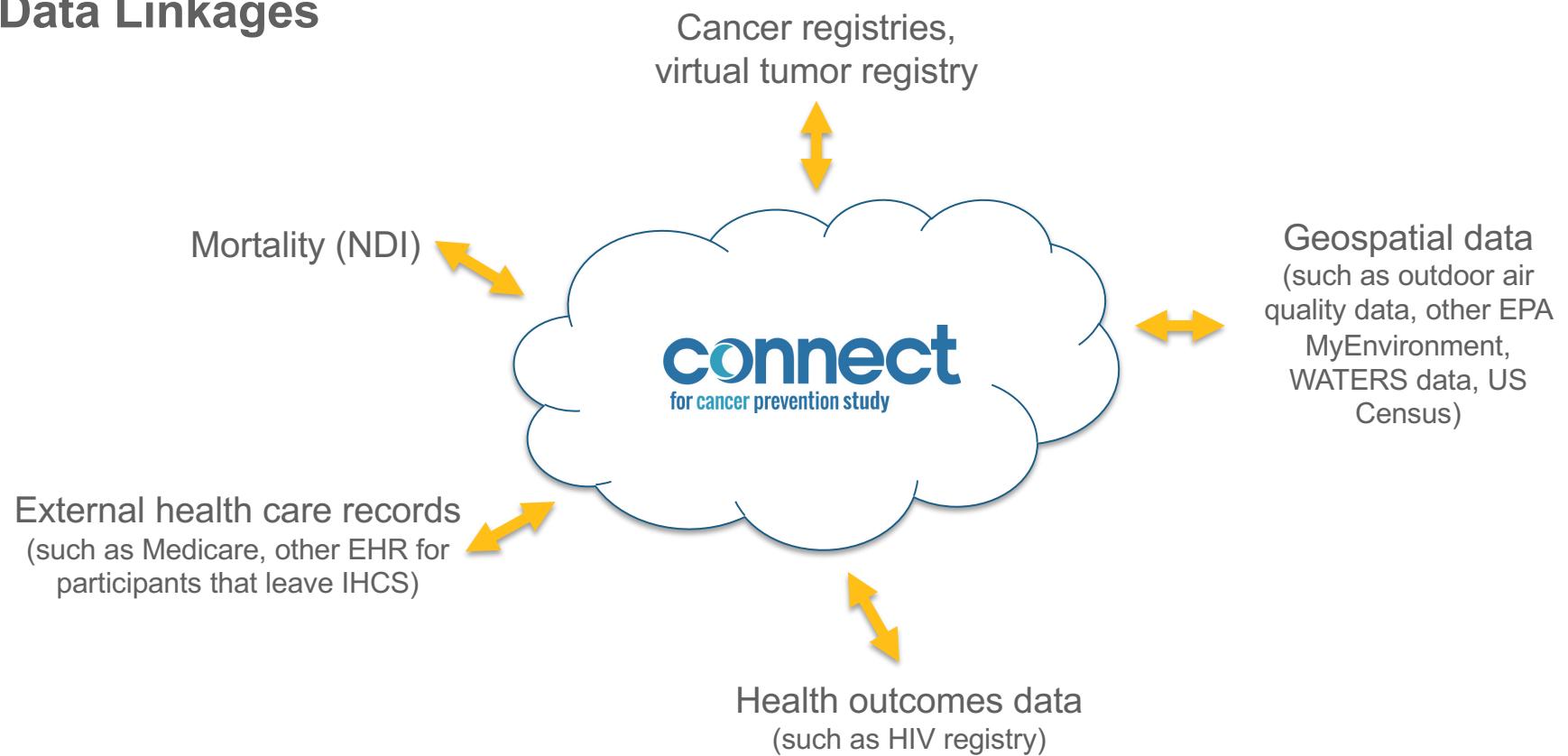
6 Months	Within 24 Months	12 Months	2 Years
<ul style="list-style-type: none">Diet food frequency questionnaire (DHQ)	<ul style="list-style-type: none">Repeated 24-hour recalls for diet (ASA24) and physical activity (ACT24)	<ul style="list-style-type: none">Screening HistoryQuality of life (PROMIS)PsychosocialSocial determinants of health	<ul style="list-style-type: none">Baseline exposures repeat assessmentConsumer products (Menstrual & intimate care, Hair, mothballs & scented)Menstrual experience (Endometriosis)

3-6 Months Post-Cancer Diagnosis Rapid Case Ascertainment

- Cancer Diagnosis Surveys (Dx, symptoms, patient-provider interaction, medical history)
 - Quality of life (PROMIS: repeat assessment & patient specific)

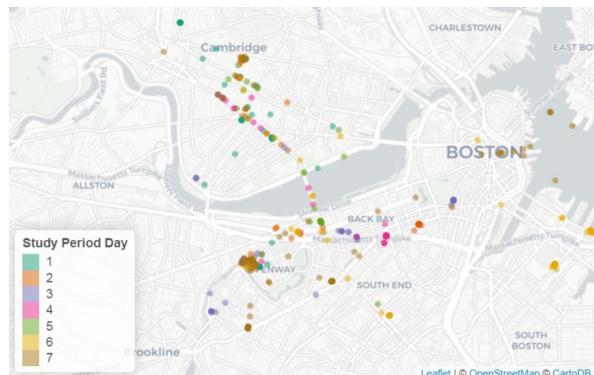
Data Linkages

Data Linkages



Geographic Information

- Geocode 20-year address histories (collected on baseline survey)
- Enable linkages to datasets for numerous external risk factor and exposure information
- Useful for planning ancillary field studies
- Add daily movement and travel to exposure information via MyConnect



Fore et al., 2020

connect
for cancer prevention study



Wearable Devices and Apps

- Connect will have research-grade and consumer-oriented devices available
- Key domains where tech can improve assessment of exposures and risk factors:
 - sleep
 - light at night
 - air pollution
 - physical activity



Estimated Outcomes

Estimated total number of incident invasive cancers

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
Colorectal	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

Estimated Total Number of Precursor Lesions

Lesions	Total Number of Expected Precursor Lesions	
	Retrospective at Baseline	Prospective at 10 Years of Follow Up
Colorectal Adenomas	5,778	18,311
Benign Breast Biopsies (females only)	2,339	4,942
Dysplastic Nevus	2,311	3,779
Cervical Biopsies (females only)	2,137	2,822
High-Grade Prostatic Intraepithelial Neoplasia (males only)	238	384
Lung Nodules	165	349

Data System Infrastructure & Digital Tools

Designed to maximize F.A.I.R. principles

F.A.I.R. Data Ecosystem Makes F.A.I.R. Data

Principles



Findable

- Unique data identifiers and metadata that is searchable using web-based tools



Accessible

- Retrievable data by identifiers based on a shared governance model



Interoperable

- Data architecture that allows for integration of data from a range of sources and applications



Reusable

- Goal of F, A, and I to allow reuse of data, code, and tools

Implementation in Connect

- Online, searchable data dictionary
- Map variables to common thesaurus to aid pooling (planned)

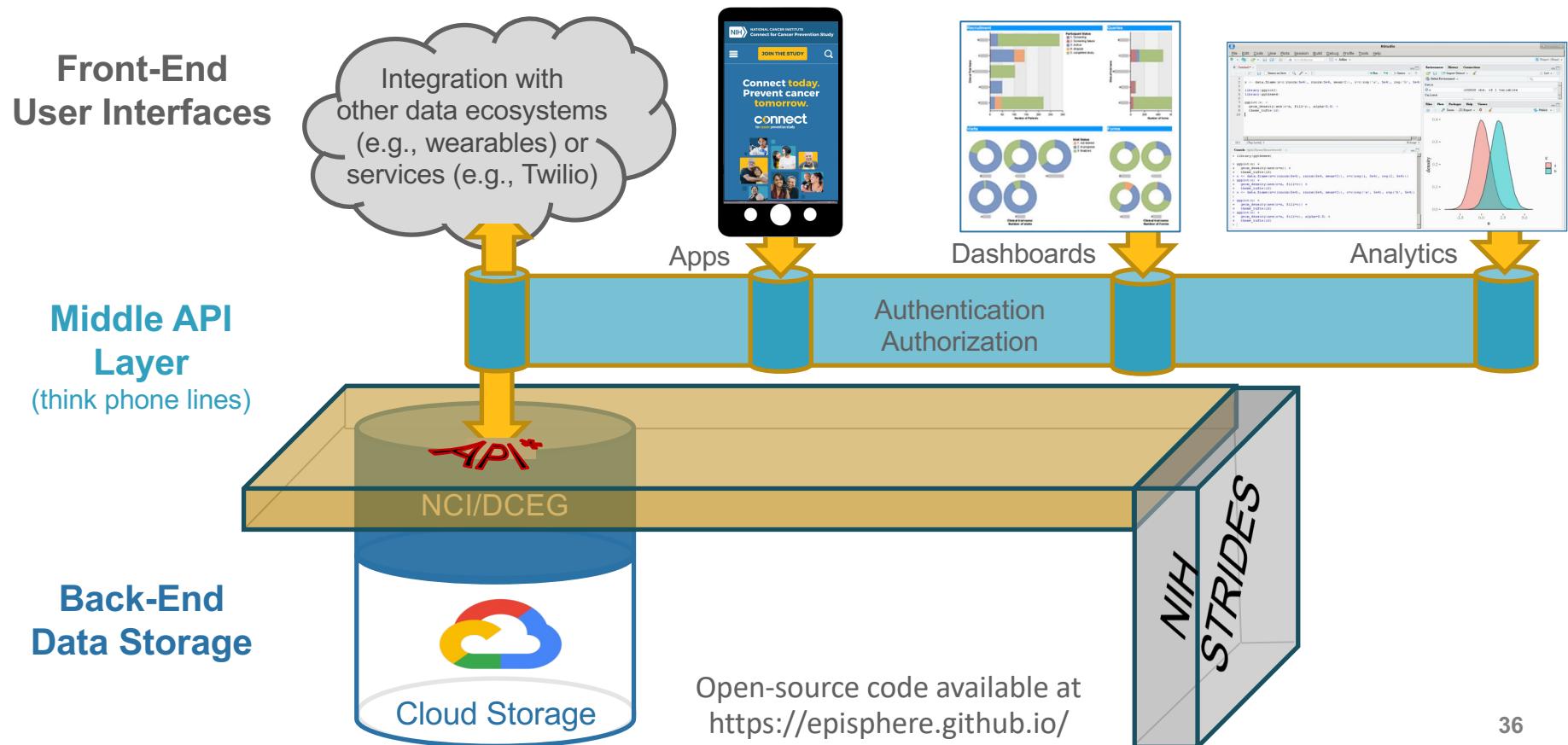
- Authentication and authorization parameters to enforce governance
- Online governance and instructions for data access (when data available)

- Cloud-based data available through governed APIs

- Data, metadata, code, tools and infrastructure available through public, versioned, open-source code on [GitHub](#)

<https://www.go-fair.org/fair-principles/>; <https://episphere.github.io/>

Connect Data Ecosystem provides an interoperable infrastructure



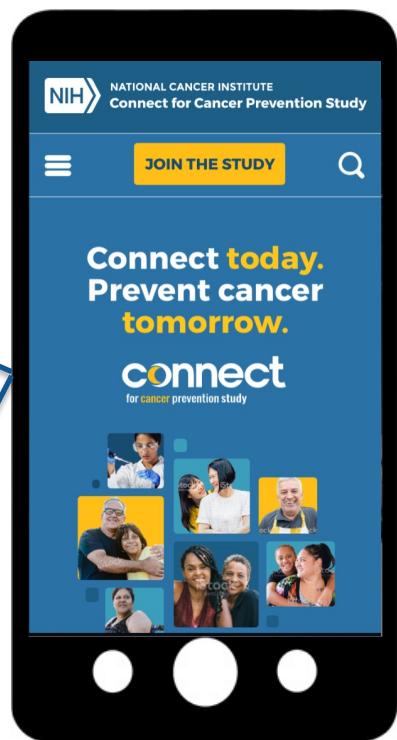
Example of Front-End User Interface: MyConnect

Study Activity Communication

The diagram illustrates the MyConnect user interface across three main sections: Study Activity Communication, Survey Completion, and Informed Consent.

Study Activity Communication: This section shows a "Messages" screen with two notifications. The first notification, "Complete Your Profile" (6 days ago), instructs the user to complete their profile section. The second notification, "Sign Your Consent Forms" (8 days ago), informs the user to sign consent forms to participate. Both notifications have "Start" buttons.

Survey Completion: This section shows a survey step titled "Background Information". The question asks how often the user has children in the household from birth to 5 years of age. The user has selected the option "Every day, at least 2 hours each day". Other options include "Every day, less than 2 hours each day", "Most days of the week", "One Day per week", "A few days per year", and "Never".

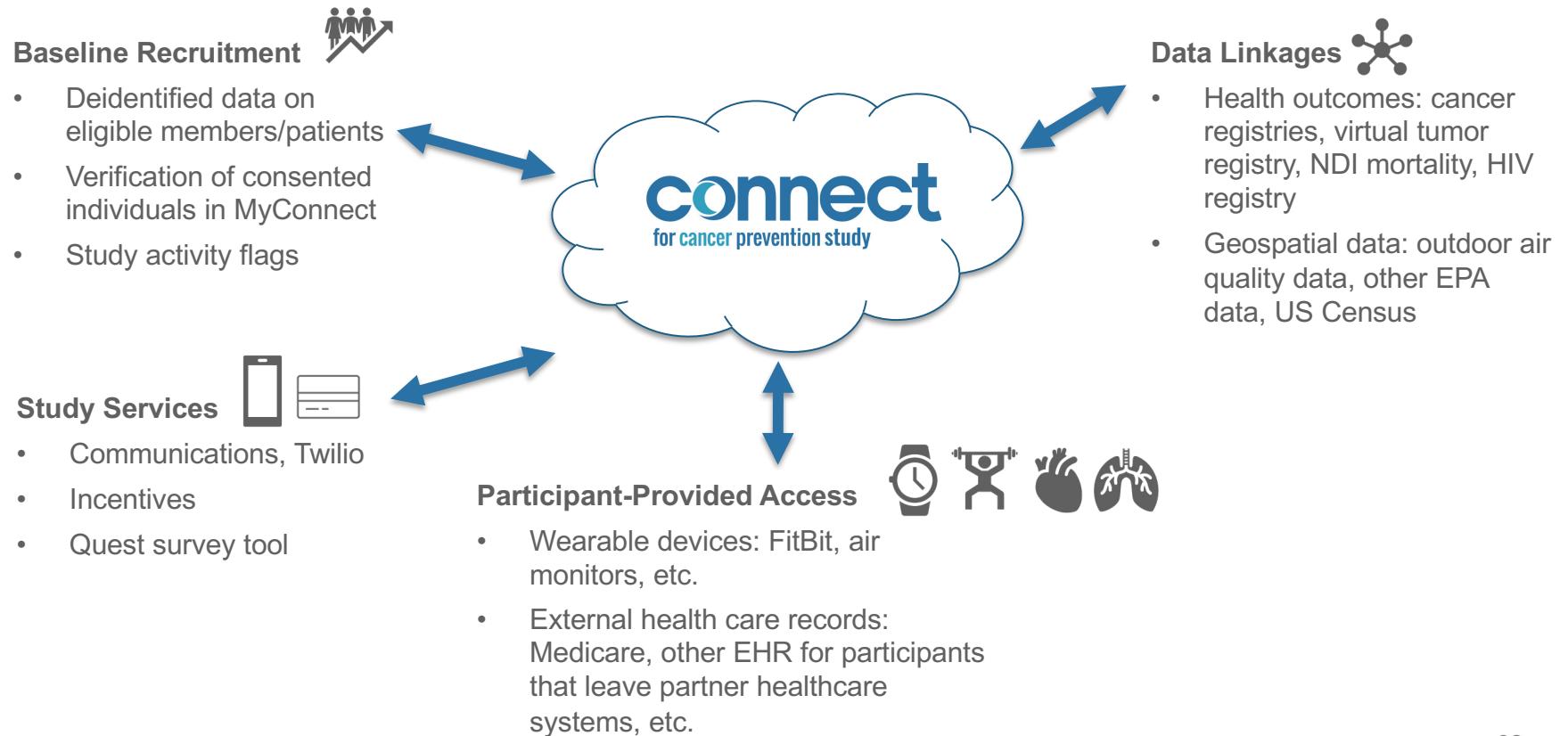


Informed Consent

This section shows the "Informed Consent" screen. It displays a progress bar indicating "7 of 8 Consent". The main content area contains the text: "In order to join Connect, we need to you to give your informed consent by reviewing the full consent form and electronic health records release form, and signing your name." Below this, a section titled "Informed consent form" provides a detailed explanation: "This is a more detailed explanation of what it means to take part in Connect." At the bottom, there is a link: "Download an unsigned copy of the consent form".

<https://github.com/episphere/connectApp>

Benefit of Middle Layer of APIs: Pulling from multiple resources



Open-Source Code Available Now: Quest render surveys into progressive web applications

The diagram illustrates the process of rendering a survey from source code to a final user interface.

Quest.js (Left):

- Questionnaire Options:** Includes fields for "Choose File" (No file chosen), "Enter URL:" (with a "Submit" button), "File Name:" (.txt) (with a "Save" button), and "Upload Questionnaire" (with a "Upload" button).
- Markup (Demo):** Displays JSON-like demo data: {"name": "D_72669695"} [INTROM1] Welcome, {\$u:firstName}! This survey is split into sections. Each section has questions that ask you about a wide range of topics. Our goal is to collect information about your medical history, family, work, and health behaviors. You can answer all of the questions in each survey section at one time, or answer some questions, pause, and return to answer the rest later. If you pause, your answers will be saved and you can pick up where you left off. You can also skip any questions that you do not want to answer.
For some questions, you may see a word or phrase that appears as a button. Clicking the button will show more information that might help you answer the question. Here is an [popup|example|example|This is an example of how additional information will be displayed.| s get started.
- Rendering:** Shows the rendered survey interface for the "Background Information" section. It includes:
 - A header with "NIH NATIONAL CANCER INSTITUTE Connect for Cancer Prevention Study" and "Sign Out".
 - A navigation bar with "Surveys", "Samples", "Agreements", "About the Study", "What to Expect", "Privacy", "My Profile", and "Messages".
 - A progress bar showing steps 1 through 6: Introduction, Background Information (highlighted), Medical History, General Health, Family History, and Education and Occupation.
 - A question: "How often do you have children in the household from birth to 5 years of age in your household?"
 - Options:
 - Every day, at least 2 hours each day
 - Every day, less than 2 hours each day
 - Most days of the week
 - One Day per week
 - A few days per year
 - Never
 - A "Clear" button.

NCI/DCEG Database (Bottom Right): Represented by a cylinder containing a cloud icon, indicating the data is being stored in a database.



Code available on [Connect GitHub](#)

Governance & Organizational Structure

Connect Governance

Connect Leadership and Coordinating Center

Governing Board

Stephen Chanock (Chair)

Executive Leadership

Steering Committee

Senior Scientist (Chair), DCEG Connect PIs, Coordinating Center Leadership, DCEG Associate Directors

Executive Committee

Senior Scientist (co-Chair), Rotating Site PI (co-Chair), Steering Committee Members, Site PIs, Participant Advisory Board Lead, Support Service Contractor Leads

External Advisory and Review Groups

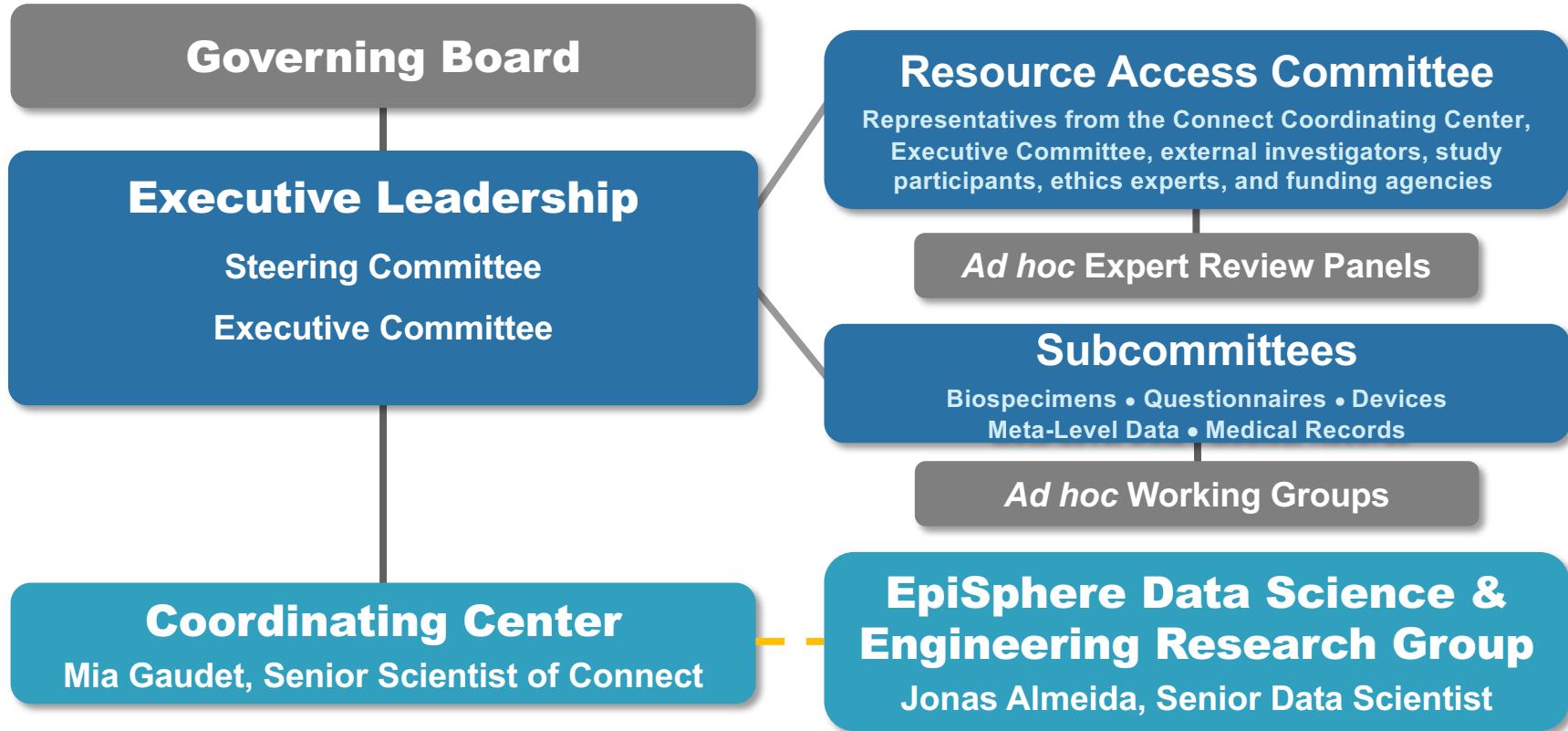
International Scientific Advisory Board

Naomi Allen, Eric Boerwinkle, Julie Buring, Carmen Guerra, Chris Haiman, David Hunter, Roger Milne, Alpa Patel, Karriem Watson

National Cancer Institute (NCI) Board of Scientific Counselors (BSC) – Clinical Sciences and Epidemiology

Participant Advisory Board (PAB)

Connect Leadership and Organization



Connect Resource Development Structure

- Subcommittees (standing)
 - Guide development of Connect infrastructure and enhancements
 - Led by DCEG with NIH investigators
 - Report to Steering Committee
- Working Groups (*ad hoc*)
 - Led by nominator with investigators from NIH, recruitment sites, and scientific community
 - Develop scientific and technical details of proposals and protocols
 - Report to relevant Subcommittees



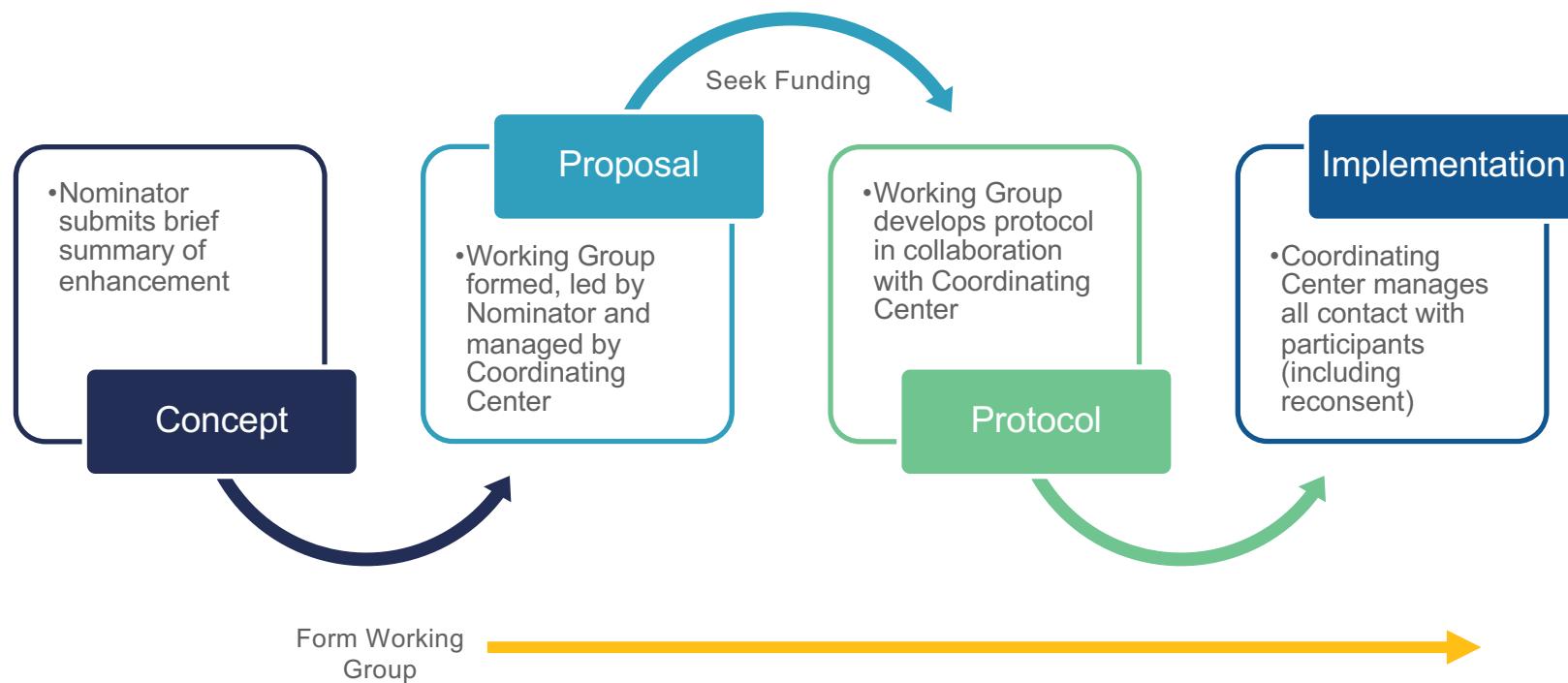
Resource Enhancements

New questionnaire modules, biospecimens, other sources of data

Connect Resource Building Principles



Connect Resource Enhancement Approval Process

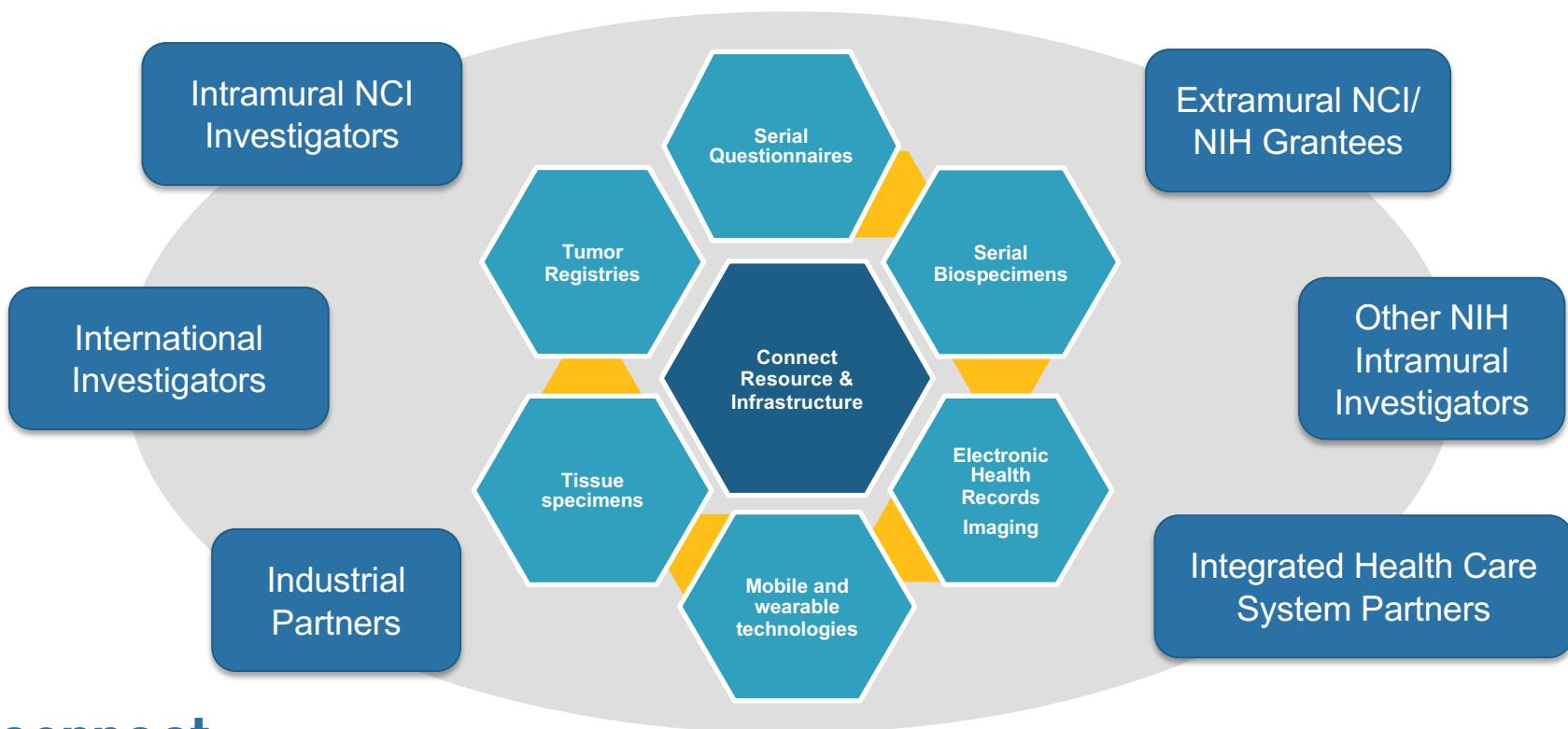


Resource Access Principles

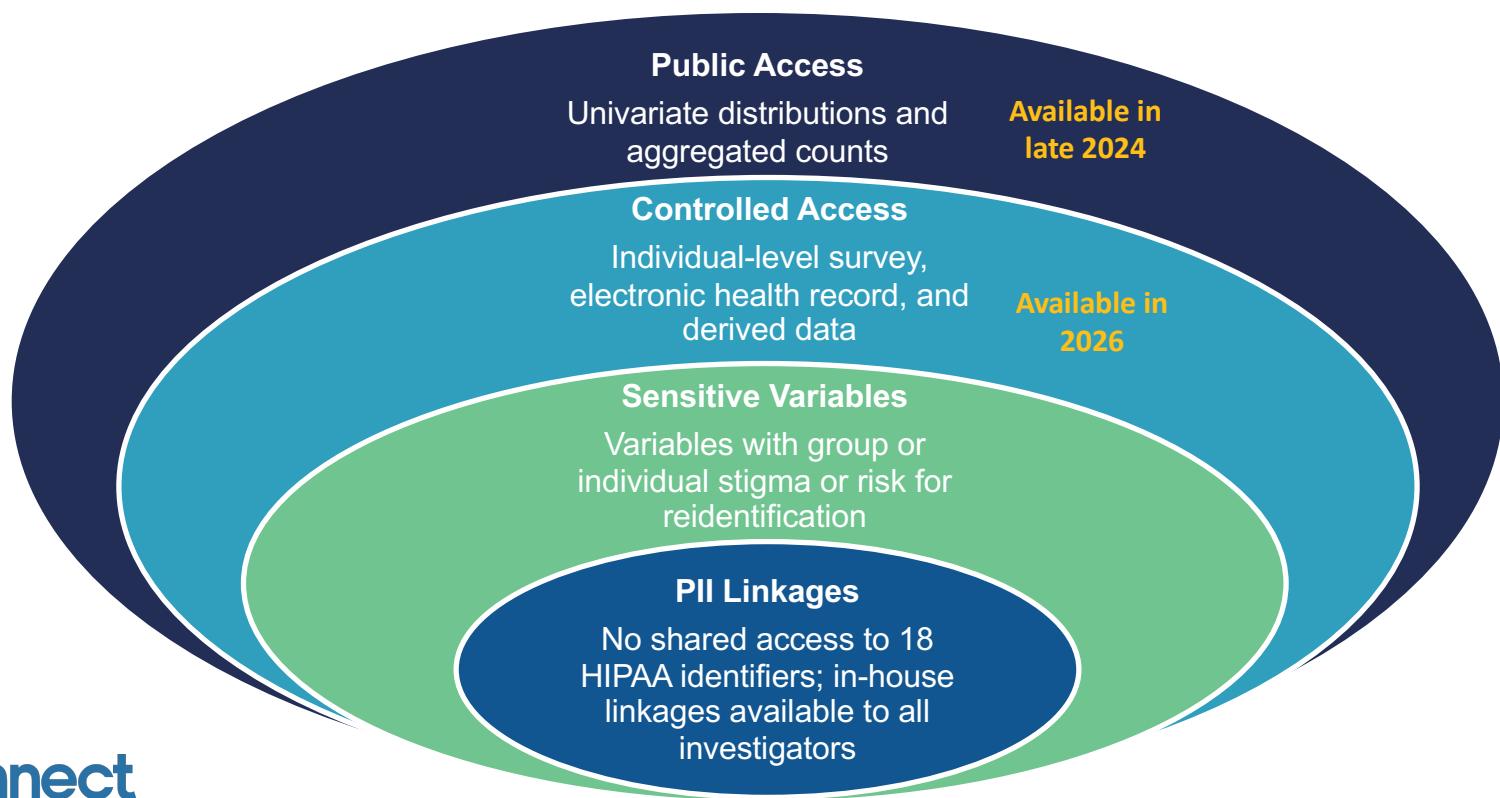
Connect Resource Access Principles



Resource Access for All*



Levels of Data Access



Acknowledgements

Acknowledging all contributors to the development of Connect





The logo consists of the word "connect" in a white, lowercase, sans-serif font. The letter "o" is stylized with a yellow crescent shape on its left side. Below the main word, the words "for cancer prevention study" are written in a smaller, white, lowercase, sans-serif font.

connect
for cancer prevention study