Enhancing Observational Data Collection to Inform Precision Cancer Research and Care

Division of Cancer Control and Population Sciences

Division of Cancer Control and Population Sciences

Enhancing Observational Data Collection to Inform Precision Cancer Research and Care

Table of Contents

Message from the Director

A critical mission of both NIH and NCI is ensuring that biomedical data are Findable, Accessible, Interoperable, and Reusable (FAIR) for the cancer research community. In recent years, there has been even greater focus on that mission. In 2016, NCI's Cancer Moonshot[SM] Blue Ribbon Panel emphasized the importance of data sharing and harmonization of results from diverse sources. In 2018, NIH released its first-ever Strategic Plan for Data Science. In 2019, NCI's Center for Biomedical Informatics and Information Technology created its Strategic Plan for NCI Informatics, Data Science, and Information Technology, to ensure research data are optimally managed and the right technologies are available when needed at NCI. And, in June 2019, the NCI Cancer Advisory Board (NCAB) released the NCAB Data Science Working Group Report, which presents interrelated recommendations that address important areas of opportunity for NCI in data science.

The "real world" evidence that observational data provides has the potential to complement traditional sources of clinical trial evidence and allow researchers to answer questions more efficiently and to guide sound decision making in the health care setting. Moreover, observational data provide valuable information concerning questions where the conduct of a clinical trial may be too costly, lengthy, or impossible for ethical reasons.

Since the formation of the Division of Cancer Control and Population Sciences (DCCPS) more than two decades ago, NCI has made significant investments in population-based observational data collection and infrastructure.

While extensive data are collected from research participants, some gaps exist in data collection. For example, detailed treatment information may not be available for all, and recurrence after cancer diagnosis may be difficult to assess. We are focused on defining criteria to identify, prioritize, and close such gaps. Current activities include expansion of capacity and collaborations to better capture the array of relevant real-world evidence required to answer current and future research questions. Our efforts will continue to emphasize ongoing improvements to statistical methods and validation of information.

> **As we explore opportunities to identify and close research data gaps, we welcome engagement with the research community, policy experts, advocates, and other stakeholders, and we hope this report will serve as a resource to provide transparency and facilitate those discussions.**

As we explore opportunities to identify and close research data gaps, we welcome engagement with the research community, policy experts, advocates, and other stakeholders, and we hope this report will serve as a resource to provide transparency and facilitate those discussions. Observational data—when appropriately collected, analyzed, and interpreted—have the potential to significantly improve the quality of cancer care and public health; but progress will depend on our ability to close critical gaps in the data and to make those data FAIR for the cancer research community.

Division of Cancer Control and Population Sciences

Enhancing Observational Data Collection to Inform Precision Cancer Research and Care

Outcomes of Interest Spanning All Programs within DCCPS

Survival Outcomes	Therapeutic Outcomes	Patient Outcomes
- Overall survival - Progression-free survival - Cancer recurrence and/or progression - Secondary malignancies	- Acute and late-onset adverse events - Response to therapy - Resistance to therapy	- Patient-reported outcomes - Quality-of-life outcomes - Cost - Adherence

What Kinds of Data Do We Need to Collect?

Healthcare System Factors

- Free-standing or part of network
- Profit status and facilities ownership
- Participation in formal QI activities
- Availability of clinical trials
- Institutional research context
- Status of health information technology implementation
- Insurance mix and reimbursement model
- Procedures for genomic companion testing

Provider Factors

- Individual physician factors
- Practice size and case mix
- Practice context
- Knowledge and attitudes
- Proximity of comparable expertise

Treatment Factors

- Drug(s) received/concomitant therapies
- Dose and duration
- Course or sequence of therapy
- Off-label use
- Use of companion diagnostic testing

Patient-Level Factors

- Sociodemographic factors
- Health indicators
- Diet and lifestyle factors
- Clinical, pathological, and tumor characteristics
- Somatic tumor characteristics
- Germline alterations
- Adherence
- Health care coverage
- Facilitators or barriers to access to health care

Population-based Observational Data Available through DCCPS and Our Partners

DCCPS has the lead responsibility at NCI for supporting research in surveillance, epidemiology, health care delivery, behavioral science, and cancer survivorship. The division also plays a central role within the federal government as a source of expertise and evidence on issues such as the quality of cancer care, the economic burden of cancer, geographic information systems, statistical methods, communication science, tobacco control, and the translation of research into practice.

To support that mission, the division gathers population-based observational data through initiatives, funding supplements, and requests for applications (RFAs), and makes those data available to the cancer research community. Below are brief summaries and links to examples of data resources available through DCCPS and our partners.

Data Types Collected

Cancer Diagnosis

Patient and Treatment Factors	Provider and System Factors	Outcomes
■ Sociodemographic factors	■ Individual physician factors	■ Overall survival
■ Health indicators	■ Practice size and case mix	■ Progression-free survival
■ Clinical, pathological, and tumor characteristics	■ Practice context	■ Cancer recurrence and/or progression
■ Germline alterations	■ Knowledge and attitudes	■ Acute and late-onset adverse events
■ Adherence	■ Proximity of comparable expertise	■ Response to therapy
■ Health care coverage	■ Free-standing or network health system	■ Resistance to therapy
■ Facilitators or barriers to access to health care	■ Profit status and facilities ownership	■ Patient reported outcomes
■ Concomitant therapies	■ Participation in formal QI activities	■ Quality-of-life outcomes
■ Dose and duration	■ Institutional research context	■ Cost
■ Course or sequence of therapy	■ Status of health information technology implementation	
■ Off-label use	■ Insurance and mix model reimbursement	
■ Use of companion Dx testing	■ Procedures for genomic testing	
■ Lifestyle and behavioral factors		

- SEER Program: The Surveillance, Epidemiology, and End Results (SEER) Program provides information on cancer statistics to reduce the burden of cancer among the US population. NCI has funded the SEER Program since 1973 to support research on the diagnosis, treatment, and outcomes of cancer.

- SEER-Medicare: The SEER-Medicare database results from the linkage of two large population-based data sources: the SEER cancer registries data and the Medicare enrollment and claims files for beneficiaries.

- SEER-MHOS: The SEER-Medicare Health Outcomes Survey (SEER-MHOS) linked database is designed to improve understanding of the health-related quality of life of cancer patients and survivors enrolled in Medicare Advantage health plans. The database contains clinical, quality-of-life, socioeconomic, demographic, and other information. SEER-MHOS is sponsored by NCI and the Centers for Medicare & Medicaid Services (CMS).

- SEER-CAHPS: The SEER-CAHPS data set is a resource for quality-of-care research based on a linkage of SEER cancer registry data with CMS' Medicare Consumer Assessment of Healthcare Providers and Systems (CAHPS®) patient surveys. These data provide a rich opportunity for analyses of Medicare beneficiaries' experiences with their care at various stages of the cancer care continuum.

- SEER-Pharmacy: NCI's SEER Program is linking with large pharmacy chain central repositories for oral therapies. These linkages will expand the clinically relevant data available to cancer researchers by giving them a more complete picture of patient treatment and outcomes. While these data are not accessible currently to the research community, a process is in development to provide accessibility in the near future.

- SEER-Genomics: Through collaborative partnerships, SEER is linking to data from genetic and genomic laboratories to characterize each cancer. These types of linkages using real-world data are increasingly important for understanding effectiveness in patients outside the clinical trial setting, who tend to come from more diverse backgrounds than those enrolled in trials.

- HINTS: The Health Information National Trends Survey (HINTS) is a cross-sectional, nationally representative survey of American adults that utilizes self-administered mail questionnaires to monitor the health information environment and assess how people access and use health-related information. NCI and extramural communication researchers analyze HINTS data to gain insight into people's knowledge about cancer, the communication channels through which they obtain health information, and their cancer-related behaviors.

- NHIS: The National Health Interview Survey (NHIS) is an annual nationwide survey of approximately 35,000 households. It is conducted by the National Center for Health Statistics and administered by the US Census Bureau. A Cancer Control Supplement (CCS) has been periodically fielded on the NHIS since 1987, and since 2000 the CCS has been co-sponsored by NCI and CDC.

- Patterns of Care: Patterns of Care data—collected under a Congressional Mandate to NCI—are used to evaluate the dissemination of state-of the-art cancer therapies and diagnostics into community practices and to inform NCI about trends in cancer therapy and survival over time, as well as health disparities. The project is coordinated jointly by DCCPS and NCI's Division of Cancer Treatment and Diagnosis.

- Physician Surveys: NCI conducts periodic assessment of physician practice with respect to new, as well as established, cancer control technologies. The NCI-led surveys collect current, national data on primary care physicians' knowledge, attitudes, recommendations, and practices. The National Survey of Precision Medicine in Cancer Treatment, sponsored by NCI, with support from the American Cancer Society and National Human Genome Research Institute, is the first nationally representative survey of oncologists about the current practice of precision medicine in cancer treatment.

- Etiology and Survivor Cohort Data: The Cancer Epidemiology Descriptive Cohort Database (CEDCD), maintained by EGRP, contains descriptive information about cohort studies that follow groups of persons over time for cancer incidence, mortality, and other health outcomes. The CEDCD is a searchable database that contains general study information (e.g., eligibility criteria and size), the type of data collected at baseline, cancer sites, number of participants diagnosed with cancer, and biospecimen information. All data included in this database are aggregated for each cohort; there are no individual-level data.

- dbGaP: The database of Genotypes and Phenotypes (dbGaP), managed by NIH, is a repository for genetic and/or phenotypic data from research studies. Most datasets are managed under controlled access. In addition to NIH-funded studies that fall under NIH's Genomic Data Sharing (GDS) policy, EGRP encourages epidemiologic researchers to make data available through dbGaP or a comparable NIH-supported repository.

Research Challenges Answered with These Data
Selected Examples

Initiative: Surveillance, Epidemiology, and End Results (SEER)

Challenge: Tests for inherited genetic mutations can provide women diagnosed with ovarian or breast cancer with important information that can have implications for family members and potentially guide treatment decisions and longer-term screening for second cancers. Therefore, it's important to determine whether women with ovarian and breast cancers are routinely receiving these genetic tests.

What we have learned from these data: In the first population study of hereditary cancer genetic testing in the United States with laboratory-confirmed testing results, researchers analyzed data on women older than age 20 who were diagnosed with breast or ovarian cancer in California and Georgia from 2013–2014 and whose data were reported to SEER registries. Only a quarter of women with breast cancer and one-third of women with ovarian cancer underwent testing for known harmful variants in breast and ovarian cancer susceptibility genes. The findings point to a need to consider likely barriers to testing outside of cost, to ensure patients benefit from the genetic information that can guide treatment or, possibly, increased surveillance.

Kurian et al. Genetic testing and results in a population-based cohort of breast cancer patients and ovarian cancer patients. J Clin Oncol 2019:37(15):1305-1315.

Initiative: SEER

Challenge: In the United States, approximately 11.2% of men will be diagnosed with prostate cancer at some point during their lifetime. For men with low-risk prostate cancer, "active surveillance" or "watchful waiting" offers an alternative to radical prostatectomy or radiotherapy, and national guidelines began advocating its use in 2010. However, current rates for this conservative management across the US are not well established.

What we have learned from these data: A recent study using SEER data found that among men with low-risk prostate cancer in the US, 42.1% used the watch-and-wait, or active surveillance, approach in 2015—up from 14.5% in 2010. This study provides a benchmark of surveillance rates in the US. It will be important to continue to follow these trends to understand the management rates in the US, as well as to observe population-level outcomes for patients on active surveillance.

Mahal et al. Use of active surveillance or watchful waiting for low-risk prostate cancer and management trends across risk groups in the United States, 2010-2015. JAMA 2019;321(7):704-706.

Initiative: SEER

Challenge: Eliminating racial/ethnic disparities in breast cancer outcomes requires an understanding of the biological and clinical care issues that may lead to those disparities.

What we have learned from these data: Researchers analyzed nearly 450,000 invasive breast cancers diagnosed between 2004 and 2011 and found that compared to black women, non-Hispanic white women were more likely to be diagnosed when their breast cancer was at an early stage, and black women had a higher risk of dying from their breast cancer, compared to white women. Differences remained after the researchers excluded women with triple-negative breast cancer, which is known to be more common among black women. After accounting for various factors that are known to affect breast cancer outcomes, the differences in outcomes persisted, suggesting that biological reasons play a role behind these disparities.

Iqbal et al. Differences in breast cancer stage at diagnosis and cancer-specific survival by race and ethnicity in the United States. JAMA 2015;313(2):165-173.

Initiative: SEER

Challenge: Most types of cancer are decreasing nationwide, but uterine cancer is proving to be an exception. Uterine cancer is the fourth-most common cancer diagnosis among US women and is among the small number of cancers increasing in both incidence and mortality in the country.

How these data are helping to address this challenge: During 1999–2015 and 1999–2016, uterine cancer incidence and mortality rates increased 0.7% and 1.1% per year, respectively, with black women disproportionately affected. These findings emphasize the need to promote awareness among women, African-American particularly, and health care providers of the need for timely evaluation of abnormal vaginal bleeding to increase the chance that uterine cancer is detected early and treated appropriately.

Henley et al. Uterine cancer incidence and mortality—United States, 1999-2016. MMWR 2018;67(48):1333-1338.

Initiative: SEER-Medicare

Challenge: The Surveillance, Epidemiology, and End Results (SEER)-Medicare data combine clinical information from population-based cancer registries with Medicare claims. These data have been used in many studies to understand cancer screening, treatment, outcomes, and costs. However, until recently, these data included limited information related to the characteristics and outcomes of cancer patients residing in or admitted to nursing homes.

What we have learned from these data: SEER-Medicare was linked to the Minimum Data Set (MDS), a standardized, comprehensive assessment instrument that is used for all individuals who receive care in a Medicare and/or Medicaid-certified nursing home, regardless of source of payment. Researchers examined this linkage and found that the inclusion of these data provides insights into cancer patients' health (cognitive status, physical function) and allows for better understandings of patient transitions between care settings (e.g., from hospitals to nursing homes) and their post-acute care, including hospice and palliative care.

Thomas et al. New opportunities for cancer health services research: linking the SEER-Medicare Data to the Nursing Home Minimum Data Set. Med Care 2018;56(12):e90-e96.

Initiative: SEER-Medicare

Challenge: Many researchers are interested in determining the practice and treatment patterns of physicians over time. Interest in studying these temporal patterns will likely increase, as more research is directed toward understanding the impact of new health care payment models. Medicare data are an important resource to evaluate these types of questions.

What we have learned from these data: Researchers developed a SEER-Medicare physician NPI (National Provider Identifier)-UPIN (Unique Physician Identification Number) crosswalk file using the National Claims History files from SEER-Medicare data and the CMS NPI and UPIN Directories. They then evaluated the ability to identify an NPI-UPIN match by physician sex and specialty. This crosswalk file will help health services researchers better track individual care providers and their practice and treatment patterns over time.

Parsons et al. Creating a National Provider Identifier (NPI) to Unique Physician Identification Number (UPIN) crosswalk for Medicare data. Med Care 2017;55(12):e113-e119.

Initiative: SEER-Medicare

Challenge: Chronic myeloid leukemia is a treatable cancer with a variety of drugs, including Gleevec, that allows patients to live near-normal lifespans. However, many new cancer patients are not using lifesaving medicine.

What we have learned from these data: An analysis of SEER-Medicare data found that among nearly 400 Medicare patients diagnosed with the disease between 2007 and 2011, nearly one-third never started the drug regimen. Those patients were more likely to be older than 80 or have other medical conditions. The researchers also found that among patients who started taking the drug within 6 months of diagnosis, those who received subsidies to defray the cost were quicker to adopt the drug than those who did not have financial help. The study raises important questions about how cost may be affecting patients' use of drugs.

Winn et al. Factors associated with tyrosine kinase inhibitor initiation and adherence among Medicare beneficiaries with chronic myeloid leukemia. J Clin Oncol 2016;34(36):4323-4328.

Initiative: SEER-Medicare

Challenge: Women in the US with early breast cancer have been increasingly opting to have their entire breast removed and then reconstructed, but the possible harms of the treatment must be quantified as compared to simpler options, so that women can make informed decisions.

What we have learned from these data: Researchers examined SEER-Medicare data on more than 100,000 patients and found that those who had a mastectomy and reconstruction were nearly twice as likely to suffer complications as those who had a lumpectomy followed by radiation. The more extensive treatment cost roughly $88,000 per patient, $23,000 more than the less extensive lumpectomy treatment, with more than $9,000 of that cost attributed to managing complications.

Smith et al. Cost and complications of local therapies for early-stage breast cancer. J Natl Cancer Inst 2017;109(1).

Initiative: SEER-Medicare Health Outcomes Survey (MHOS)

Challenge: Health-related quality of life (HRQOL) for both individuals diagnosed with cancer and those who are cancer free is affected by a wide range of factors, including clinical and initial treatment, and socioeconomic, demographic, co-morbidity, race/ethnicity, and other factors. Clinicians need to understand the impact of these factors, individually and in combination, to help improve the HRQOL for the patients in their care.

What we have learned from these data: Researchers specifically examined the prevalence of falls and balance/walking problems among older cancer survivors and found that prevalence ranged by cancer site and time since diagnosis, with falls ranging from 12%–28% prevalence and significant increases in falls among lung and prostate cancer patients post-diagnosis. The findings point to a need to consider functional deficits and cancer diagnosis during follow-up screening, surveillance, and interventions.

Huang et al. Prevalence of self-reported falls, balance or walking problems in older cancer survivors from Surveillance, Epidemiology and End Results-Medicare Health Outcomes Survey. J Geriatr Oncol 2017;8(4):255-261.

Initiative: SEER-MHOS

Challenge: Health-related quality of life after a cancer diagnosis is predictive of overall survival. Few studies have examined the impact of health-related quality of life pre-diagnosis compared to post-diagnosis in terms of overall survival.

What we have learned from these data: Worse pre-diagnosis HRQOL was significantly associated with greater risk of death across HRQOL measures in older lung cancer patients. The researchers noted that these findings reinforce the importance of self-reported health status as a predictor for overall survival, and that routine HRQOL screening may identify patients who could benefit from early interventions to improve HRQOL. Future studies should explore associations between changes in HRQOL before and after cancer diagnosis and overall survival.

Pinheiro et al. The prognostic value of pre-diagnosis health-related quality of life on survival: a prospective cohort study of older Americans with lung cancer. Qual Life Res 2017;26(7):1703-1712.

Initiative: SEER-Consumer Assessment of Healthcare Providers and Systems (CAHPS)

Challenge: Interventions to improve cancer survivors' health status, such as increased access to supportive care services, may improve experience of care, but we must examine data to understand experiences of care among older cancer survivors.

What we have learned from these data: SEER-CAHPS data allow assessment of factors influencing experience of cancer among US cancer survivors. Higher self-reported general health status was associated with better care experiences for breast, colorectal, and prostate cancer survivors, while better mental health status was associated with better care experiences for lung cancer survivors. Interventions to improve cancer survivors' health status, such as increased access to supportive care services, may improve experience of care.

Halpern et al. Healthcare experience among older cancer survivors: Analysis of the SEER-CAHPS dataset. J Geriatr Oncol 2018;9(3):194-203.

Initiative: SEER-CAHPS

Challenge: Patient-provider relationships serve a potentially important role in adherence to office visits for colorectal cancer (CRC) surveillance. As adherence may increase survival among CRC survivors, it's important to investigate components of this relationship that may influence office visit adherence, as well as other potentially modifiable drivers of surveillance guidelines.

What we have learned from these data: By examining SEER-CAHPS data, associations between experiences of care and adherence to office visits and surveillance guidelines can be better understood. Adjusted models showed that ratings of personal doctor and specialist doctor were positively associated with adherence to office visits, and ratings of personal doctor were associated with adherence overall among Medicare Fee-For-Service beneficiaries with colorectal cancer.

Mollica et al. Examining colorectal cancer survivors' surveillance patterns and experiences of care: a SEER-CAHPS study. Cancer Causes Control 2017;28(10):1133-1141.

Initiative: Health Information National Trends Survey (HINTS)

Challenge: In a rapidly evolving health communication environment, program planners need accurate data to overcome barriers to health information usage across populations, while social scientists need data to refine theories of health communication and to offer new and better recommendations for reducing the burden of cancer.

What we have learned with these data: Survey researchers have used the HINTS data collection program to understand how adults 18 years and older use different communication channels, including the Internet, to obtain vital health information for themselves and their loved ones. Cancer information seeking was reported most frequently by cancer survivors (69.8 %), and the percentage of cancer survivors who reported information seeking increased from 66.8% in 2003 to 80.8% in 2013. Cancer information seeking was less likely among older adults, those with less education, and those with lower incomes. Survivors surveyed in 2008 were more than twice as likely (44%) as respondents in 2003 (20%) to report that health care providers were their first source of cancer information; the percentage reporting use of Internet as the first source remained relatively stable over this period, decreasing from 44% to 36%.

Finney Rutten et al. Cancer-related information seeking among cancer survivors: trends over a decade (2003–2013). J Canc Educ 2016;31:348-357.

Initiative: HINTS

Challenge: The Institute of Medicine, now called the Health and Medicine Division of the National Academies of Sciences, Engineering, and Medicine, has recommended that cancer survivors completing treatment be provided with a treatment summary to facilitate delivery of patient-centered survivorship care. However, the relationship between treatment summary receipt and patient-centered communication and overall quality of care are not well understood.

What we have learned with these data: HINTS data showed that cancer survivors who received a treatment summary were three times more likely to report excellent or very good quality of care compared to those survivors who did not receive a treatment summary. They were also significantly more likely to report better overall patient-provider communication. Survivors who had received two types of cancer treatment were significantly more likely than those who had received one type of cancer treatment to report receiving a treatment summary.

Blanch-Hartigan et al. Cancer survivors' receipt of treatment summaries and implications for patient-centered communication and quality of care. Patient Educ Couns 2015;98(10):1274-1279.

Initiative: National Health Interview Survey (NHIS)

Challenge: Screening for colorectal cancer (CRC), the second leading cause of cancer-related death in the US, has been shown to reduce deaths from this disease, yet many adults in the US are not up to date with CRC screening. By examining CRC screening rates by several features of health insurance coverage, we can better understand progress and how we can improve lagging screening rates.

What we have learned from these data: According to NHIS data, recommendation-consistent CRC screening increased from 51.6% in 2008 to 61.3% in 2015. While CRC screening rates have increased over time, certain segments of the population, especially the uninsured, continue to screen below recommended levels. Data from the NHIS can be used to continue to monitor the effects of health insurance coverage and changes in health insurance reform on CRC screening use.

de Moor et al. Colorectal cancer screening in the United States: trends from 2008–2015 and variation by health insurance coverage. Prev Med 2018;112:199-206.

Initiative: NHIS

Challenge: Cancer survivors can face ongoing health issues and need access to affordable health care, yet studies examining health care access and affordability in this population have been lacking. To better understand their health care needs, it is important to evaluate trends in their health care access and its affordability.

What we have learned from these data: NHIS data show that cancer survivors were more likely than individuals without a cancer history to delay or forego medical care, and/or to be unable to afford medications and health care services. The proportion of survivors reporting problems with health care access and affordability decreased, however, following implementation of the Affordable Care Act. Such incremental improvement in health care access and affordability after recent health care reform provides an important benchmark for additional potential changes in the future.

Nipp et al. Patterns in health care access and affordability among cancer survivors during implementation of the Affordable Care Act. JAMA Oncol 2018;4(6):791-797.

Initiative: Patterns of Care

Challenge: Ovarian cancer is the most common gynecologic cancer and the leading cause of gynecologic cancer deaths in the US. While studies have demonstrated that guideline care is associated with improved survival for patients with ovarian cancer, many women with ovarian cancer do not receive guideline care.

What we have learned from these data: Researchers evaluated population-based trends in ovarian cancer treatment and survival, focusing on receipt of guideline care for women diagnosed with ovarian cancer in 2002 and 2011. They also assessed patient and provider characteristics associated with receipt of guideline care. For women diagnosed with ovarian cancer, there is underuse of guideline care, suboptimal rates of guideline care from gynecologic oncologists, and lack of improvement in 2-year survival, which might be explained by the lack for guideline care.

Warren et al. Trends in the receipt of guideline care and survival for women with ovarian cancer: population-based study. Gynecol Oncol 2017;145(3):486-492.

Initiative: Patterns of Care

Challenge: Lung cancer, most of which is non-small cell lung cancer (NSCLC), is the leading cause of cancer mortality in the US. Evidence has demonstrated disparities by race, age, and insurance status with respect to receipt of appropriate care. A better understanding of the treatment patterns and outcomes among NSCLC patients, particularly the influence of sociodemographic characteristics, could lead to more equitable evidence-based care.

What we have learned from these data: Researchers found treatment disparities among patients with stages I–IV NSCLC, notably among older patients, and increased chemotherapy receipt and improved survival for stages IIIB/IV. Specifically, between 1996 and 2010, use of systemic therapy increased significantly, regardless of stage. Older patients (≥70 vs. <70) were less likely to receive surgery and chemotherapy. Median survival improved significantly among III–IV patients. However, the statistically significant one-month increase in median survival among stage IIIB/IV patients, from 4 to 5 months, highlights the need for additional research.

Kaniski et al. Temporal patterns of care and outcomes of non-small cell lung cancer patients in the United States diagnosed in 1996, 2005, and 2010. Lung Cancer 2017;103:66-74.

Initiative: Physician Surveys

Challenge: For cancer survivors, there are often many unknowns about their care. To ensure that they receive the most appropriate care moving forward, it's important to understand trends in the frequency of and factors associated with oncologists' and primary care physicians' discussion of survivorship care recommendations with survivors.

What we have learned from these data: Survey data showed that a majority of oncologists always or almost always discuss survivorship care recommendations with survivors; however, discussions about whom survivors should see for cancer-related and other follow-up care were far less common, and written survivorship care plans were almost never provided. Survivorship care recommendations and provider responsibility were rarely discussed between primary care physicians and survivors, indicating that new efforts are needed to improve the follow-up care of survivors.

Blanch-Hartigan et al. Provision and discussion of survivorship care plans among cancer survivors: results of a nationally representative survey of oncologists and primary care physicians. J Clin Oncol 2014;32(15):1578-85.

Initiative: Physician Surveys

Challenge: It is important that oncologists communicate information about long-term and late-onset treatment side effects to primary care providers as patients transition back to primary care settings after completion of treatment for cancer. While many of these patients may experience chemotherapy-related late or long-term effects, the extent to which physicians are aware of these effects is unknown.

What we have learned from these data: Using survey data, researchers described and compared oncologists' and primary care physicians' awareness of select chemotherapy-related late or long-term effects and found that although primary care providers are aware of some long-term and late effects, knowledge of other common treatment side effects is much more limited. The findings suggest that education for all providers caring for the growing population of cancer survivors is needed.

Nekhlyudov et al. Oncologists' and primary care physicians' awareness of late and long-term effects of chemotherapy: implications for care of the growing population of survivors. J Oncol Pract 2014;10(2):e29-36.

Division of Cancer Control and Population Sciences

Enhancing Observational Data Collection to Inform Precision Cancer Research and Care

Initiative: Physician Surveys

Challenge: With the increasing application of precision oncology, understanding how oncologists use genomic tests in their practice and the factors that affect their use is essential to ensure that patients who can benefit receive appropriate testing and follow-up.

What we have learned from this data initiative: The purpose of this study was to investigate how oncologists in the United States use next-generation sequencing (NGS) tests to evaluate patients with cancer and inform treatment recommendations. In 2017, most oncologists in the US were using NGS tests to guide treatment decisions for their patients. More research is needed to establish the clinical usefulness of these tests, to develop evidence-based clinical guidelines for their use in practice, and to ensure that patients who can benefit from these new technologies receive appropriate testing and treatment.

Freedman AN et al. Use of next-generation sequencing tests to guide cancer treatment: results from a nationally representative survey of oncologists in the United States. JCO Precis Oncol 2018;2:1-13.

Initiative: Etiology and Survivor Cohorts

Challenge: A long-standing question has been whether diabetes is a consequence of pancreatic cancer or whether it is a risk factor that contributes to the development of pancreatic cancer.

How these data are helping to address this challenge: The Multiethnic Cohort Study, which was established to investigate possible etiologic factors associated with cancer risk, examined the risk of pancreatic cancer following recent and long-term diabetes among African Americans and Latinos. Diabetes was associated with a more than twofold higher risk of pancreatic cancer in African Americans and Latinos, but recent-onset diabetes was associated with a 2.3-fold greater increase in risk than long-standing diabetes. The findings support the hypothesis that recent-onset diabetes is a manifestation of pancreatic cancer and that long-standing diabetes is a risk factor for this malignancy.

Setiawan et al. Pancreatic cancer following incident diabetes in African Americans and Latinos: the Multiethnic Cohort. J Natl Cancer Inst 2019;111(1):27-33.

Initiative: Etiology and Survivor Cohorts

Challenge: It is well known that survivors of childhood cancer are at risk of developing early and often severe chronic health conditions. However, a quantitative landscape of morbidity of survivors—critical to informing future clinical guidelines, research investigations, and health services planning—had not been described.

How these data are helping to address this challenge: The St. Jude Lifetime (SJLIFE) Cohort is a retrospective cohort study with prospective follow-up and ongoing data accrual designed to facilitate longitudinal, clinically based assessment of health outcomes among survivors of pediatric malignancies. Using SJLIFE data, researchers were able to quantify the chronic health conditions in survivors of childhood cancer, presenting a detailed condition-by-condition assessment of morbidity in this growing high-risk population.

Bhakta et al. The cumulative burden of surviving childhood cancer: an initial report from the St Jude Lifetime Cohort Study (SJLIFE). Lancet 2017;390(10112):2569-2582.

NCI's Surveillance, Epidemiology, and End Results (SEER) Program

Funded by NCI to support research on the diagnosis, treatment, and outcomes of cancer since 1973, the SEER Program now includes 16 population-based registries, covering 35% of the US population. Approximately 550,000 incident cases are received annually through the registries, 85% of which include real-time electronic pathology (e-path) reporting. And, in the near future, all registries will be on a common data platform (SEER DMS) that permits central linkages with external partners and facilitates scaling of new initiatives across all registries simultaneously.

NCI is currently undertaking a series of innovative pilot studies and developing tools to improve collection methods and expand the size and diversity of the population that SEER covers. Ultimately, the main goal in enhancing SEER is to create a system representing population-level real-world data to supplement clinical trials and understand effectiveness of oncology care for the 95% of patients treated outside the clinical trial setting.

SEER datasets are listed above under "Population-based Observational Data Available through DCCPS and Our Partners."
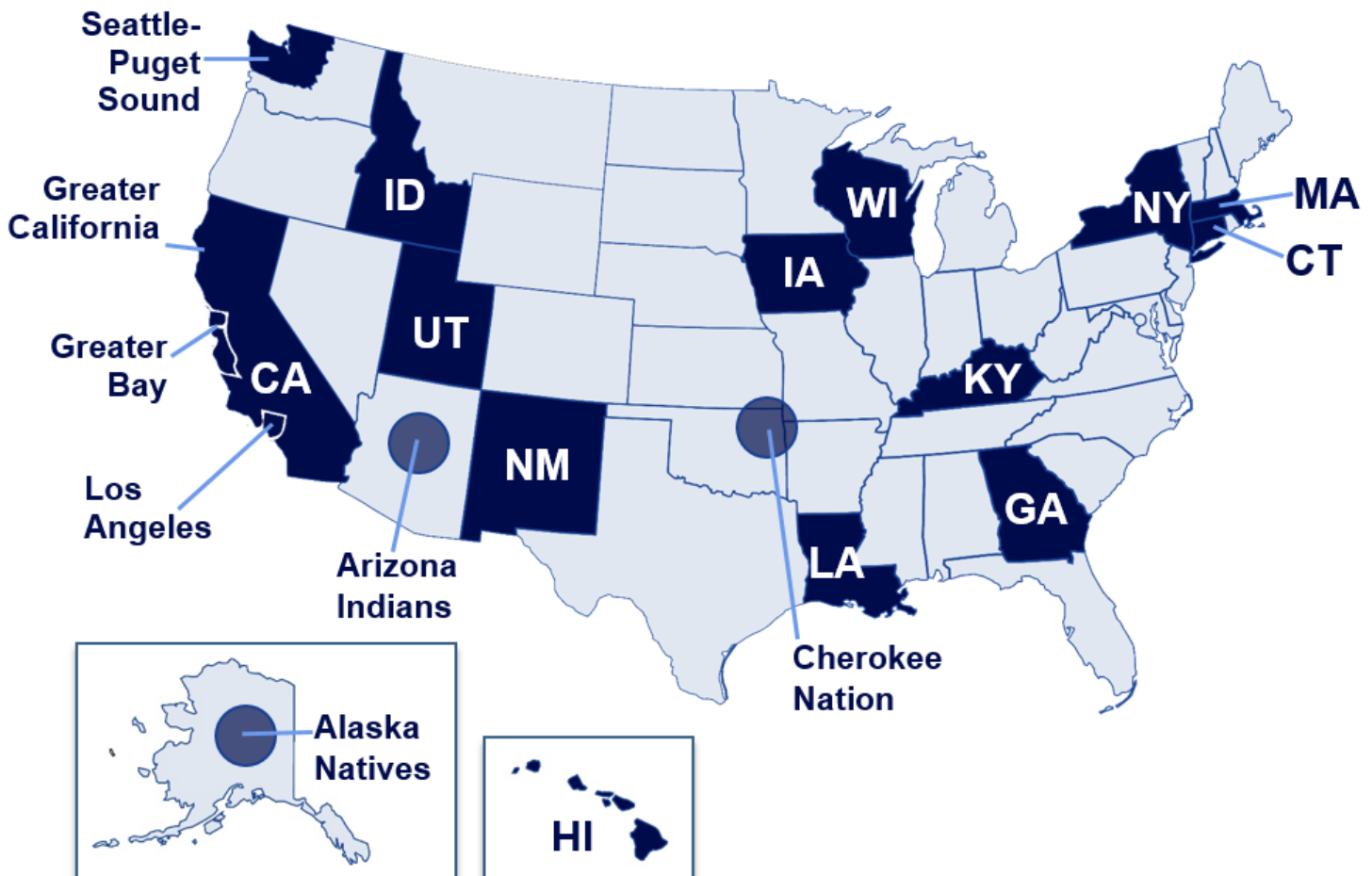
SEER Data Sources: Current and in Testing

Data Sources Currently Used

Data Sources Being Piloted

Hospital Registries
Abstracts, EHR, other data

Pathology Labs
E-path, images, other data

Oncology/Physician Offices
E-path, images, claims, other data

Other (providers, surgical centers)
Administrative claims, other data

Other Registries
Infectious disease, HIV/AIDS, other

DMV, NDI, Voter Registration, SSN, Other
Images (PDFs), other data files

Census
Demographic data, SES, geography

SEER Central Cancer Registry

SEER DMS Database

Biomarkers (genetic, genomic)

Pharmacy Claims

Electronic Health Records

Patient-Reported Outcomes

Administrative Medical Claims

Surveys

Advancing Cancer Control through Etiologic Research

Cohort and case-control studies are the fundamental observational designs for epidemiological research that have discovered and validated the risk of cancer from smoking, other chemicals (e.g., benzene), obesity, radiation exposure, sunlight, genetic variants, and many other factors. These studies have also assured us that coffee is not a risk factor for cancer, have shown the benefit of healthy eating, physical activity, and maintaining a healthy weight, and have helped to determine who falls into high-risk groups, suggesting the need for more screening and prevention interventions. These studies also contribute to the evaluation of screening interventions and how to improve outcomes after cancer has been diagnosed. This long-standing investment also means there is a wealth of data and biospecimens available for the next discovery of factors associated with increased or decreased risk of cancer.

Cancer epidemiology cohorts are large observational population studies in which groups of people with a set of characteristics or exposures are prospectively followed for the incidence of new cancers and cancer-related outcomes. NCI supports numerous cancer epidemiology cohort studies, including both etiology cohorts and cancer survivor cohorts, through investigator-initiated research project mechanisms and a funding opportunity announcement that supports core infrastructure and methodological research, PAR-17-233. The majority of these cohorts are required by the National Institutes of Health (NIH) to share final datasets (or data used for their main publications), through controlled-access models for de-identified data from human participants. NIH's database of Genotypes and Phenotypes (dbGap) (https://www.ncbi.nlm.nih.gov/gap/) is one example of an NIH data repository through which investigators may choose to share their data.

The Cancer Epidemiology Descriptive Cohort Database (CEDCD) (https://cedcd.nci.nih.gov/) is another resource available to the research community to facilitate collaboration and highlight research opportunities within existing cohort studies. The CEDCD contains descriptive information about existing etiologic and cancer survivor cohort studies that follow groups of persons over time for cancer incidence, mortality, and other health outcomes. CEDCD is a searchable database that contains general study information (e.g., eligibility criteria and size), the type of data collected at baseline, cancer sites, number of participants diagnosed with cancer, and biospecimen information. All data included in this database are aggregated for each cohort; there are no individual-level data. The CEDCD also provides information on how to contact investigators for collaborative research.

In addition, multiple research consortia, some organized by study design (e.g., the NCI Cohort Consortium or consortia of case-control studies such as Pancreatic Cancer Case-Control Consortium (PanC4)) and others by cancer type such as the Epidemiology of Endometrial Cancer Consortium (E2C2), have been formed. The creation of research consortia occurs independently from NCI and NIH funding mechanisms, although researchers participating in consortia may partner in writing research grant applications. Consortia are able to address scientific questions that cannot be addressed otherwise due to scope, resources, population size, or expertise. A listing of consortia, and whom to contact for additional information regarding collaboration and access to data can be found at https://epi.grants.cancer.gov/Consortia/. Many of the cohort and case-control studies have also contributed genotype and phenotype data to repositories such as dbGaP, which enables access to the data by qualified researchers.

Future initiatives funded through the Cancer Moonshot℠, which requires broad data sharing, will further enrich existing data resources, providing opportunities for a broad array of researchers to access the data, enhance our knowledge base, and further improve cancer control activities. With the ongoing conduct of observational studies supported through NIH funding, data resources will continue to grow. Study participants have given generously of their time and efforts by taking part in these observational studies. We are working with members of the NCI community to provide enhanced access by researchers to these rich data resources and maximize the impact of these studies on cancer control.

Improving the Delivery of Cancer-related Care

A major focus of activity in DCCPS is developing resources and initiatives designed to support innovative efforts to improve the delivery of cancer-related care. Advances in risk assessment, prevention, screening, and drug development have contributed to declining incidence and mortality rates for many common cancers. At the same time, numerous challenges have been identified that could limit additional progress. These challenges include the rapidly growing demand for care given the aging of the population, the increasing complexity of care, problems with care coordination, and sharply rising costs of care. To address these issues, DCCPS has adopted a multilevel perspective that focuses on the collection and use of data on patient experiences and health states, provider and system factors that influence care delivery, and patterns and outcomes of care. Consistent with this perspective, DCCPS provides access to a number of research resources for the extramural community.

At the patient level, there has been a major effort to facilitate the development of robust patient-centered health outcome measures and promote their use in clinical research. In collaboration with 14 NIH institutes and centers, NCI led an initiative that supported the development of HealthMeasures, an online resource that provides access to several patient-centered measurement systems that were originally developed and validated with NIH funding. Among these systems is PROMIS® (Patient Reported Outcomes Measurement Information System®), which can be used to assess multiple domains of physical, mental, and social health in children and adults. NCI has also supported the development and use of PRO-CTCAE™ (Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events). PRO-CTCAE was designed to collect patient perspectives on the experience of symptomatic adverse events in cancer clinical trials and is meant to be a companion to the clinician graded CTCAE. To advance the use of patient-reported outcome measures to improve cancer care, NCI responded to the Cancer Moonshot Blue Ribbon Panel recommendations by proposing and funding the IMPACT (Improving Management of symPtoms during And Following Cancer Treatment) initiative. This consortium is conducting research designed to accelerate the widespread adoption of electronic systems that routinely capture patient symptom reports and used that information to trigger clinical responses consistent with evidence-based guidelines.

At the provider and health system levels, DCCPS supports several initiatives to improve the delivery of cancer-related care. NCORP (NCI Community Oncology Research Program) is a national network that offers unique opportunities to conduct cancer care delivery research in community settings, where the majority of cancer patients receive their treatment. Some studies conducted within NCORP collect data on ways to improve clinical practice patterns and others develop and test promising interventions within health care delivery systems. Another initiative, PROSPR (Population-based Research to Optimize the Screening PRocess), is a research network that is investigating how to improve the cancer screening process in community health care settings. PROSPR grantees are currently conducting multilevel observational research across multiple health care systems to evaluate factors that affect the quality of the screening process for colorectal, cervical, and lung cancers. PROSPR will serve as a data resource that can be accessed by the extramural community to expand the depth and breadth of research conducted on cancer screening.

At the population level, DCCPS provides access to and support for use of several SEER-linked data resources and nationally representative surveys that can be used to address a wide range of research questions about the delivery of care across the cancer control continuum. SEER-linked data resources include SEER-Medicare, SEER-MHOS, and SEER-CAHPS and are described elsewhere in this document. Recent or planned enhancements to one or more of these SEER data resources include linkages to Medicare Part D prescription data, Medicaid enrollment data, and clinical assessment data for patients enrolled in nursing homes and home health care. DCCPS has also played a central role in conducting large-scale provider surveys designed to address specific issues in the delivery of cancer care. One of the most recent surveys of this kind described in this document is the National Survey of Precision Medicine in Cancer Treatment, the first nationally representative survey of oncologists about this topic. The survey was conducted to better understand how oncologists use genomic testing to inform treatment recommendations; their confidence in using genomic test results to guide care; and key factors that influence their use of genomic testing in clinical practice. Finally, DCCPS supports data collection through nationally representative surveys, including the National Health Interview Survey (NHIS) Cancer Control Supplement and the Medical Expenditure Panel Survey (MEPS) Experiences with Cancer Supplement. The NHIS supplement gathers information about cancer screening, sun protection, tobacco use, physical activity, diet, genetic counseling, and other behaviors germane to cancer control. The MEPS supplement includes a subsample of NHIS respondents and collects comprehensive data on health insurance, access to health care, utilization, and expenditures.

Informing Implementation Science

DCCPS has had a long-standing interest in implementation science, defined as "the study of methods to promote the adoption and integration of evidence-based practices, interventions, and policies into routine health care and public health settings to improve our impact on population health." The scientific community's effort to bridge the gap between research and cancer control relies on the ability to assess the degree to which evidence-based cancer control interventions are adopted, implemented, and sustained within clinical and community settings, which requires access to and analysis of high-quality observational data. Implementation scientists rely on health care claims and EHR data to determine spread of effective screening, diagnostic, and therapeutic interventions, both to identify implementation gaps and to assess the impact of strategies used to improve uptake and sustainment of effective interventions. For cancer prevention interventions, implementation scientists must rely on data from communities and state health departments reporting quality and reach of intervention delivery.

NCI, in collaboration with 19 other NIH institutes, centers, and offices, recently reissued the standing program announcements for extramural funding in Implementation Science, Dissemination, and Implementation Research in Health (R01/R03/R21). Investigators are encouraged to consider the use of observational data to answer the full range of questions related to implementation science in cancer, including understanding factors associated with improvements in implementation and sustained use of evidence-based interventions, as well as strategies through which ineffective and harmful practices can be de-implemented.

The Cancer MoonshotSM has given researchers in the cancer implementation science community a number of additional opportunities to use observational data to chart progress in implementing effective interventions for cancer screening, diagnosis, treatment, and follow-up. Initiatives such as Accelerating Colorectal Cancer Screening and follow-up through Implementation Sciences (ACCSIS); Research for Improving the Management of Symptoms during and following Cancer Treatment (IMPACT); and Approaches to Identify and Care for Individuals with Inherited Cancer Syndromes all rely on the existence of administrative and EHR data to demonstrate improvements in care across the cancer continuum. The newest Cancer Moonshot RFAs in Implementation Science, RFA CA-19-005/6: Implementation Science Centers in Cancer Control, are focused on building infrastructure to support scale-up of efforts to grow the field. At the center of this initiative is the goal of developing "implementation laboratories" in clinical and community settings that can carry out a range of innovative implementation studies. We anticipate that the initiative will pave the way, using observational data, to establish an implementation science data ecosystem that can more rapidly identify opportunities for clinical and community care improvement and report on the impact of a variety of implementation efforts nationwide.

Scanning the Observational Data Collection Environment

As we identify gaps in the post-diagnosis observational data currently collected, as well as the opportunities to close those gaps through collaboration and enhancement, we recognize that we must consider—and leverage—complementary efforts under way in the private sector, avoiding duplication. Here we provide a sampling of such efforts by some prominent companies and organizations:*

- Flatiron Health, founded in 2012, collects community oncology data through their electronic health records (EHR) system, OncoEMR. The data collected includes patient and practice demographics, cancer diagnosis and treatment, and patient outcomes. Flatiron Health is partnered with over 265 community oncology clinics and has 2 million active patient records available for research.

- CancerLinQ is an initiative of the American Society of Clinical Oncology. To improve the health care delivery quality for cancer patients, CancerLinQ collects and analyzes data from all patient-physician interactions. CancerLinQ is unique in that it is the only not-for-profit, physician-led, and professional society-driven big data effort. In 2017, CancerLinQ LLC and NCI announced a national data sharing collaboration to facilitate the exchange of information between CancerLinQ®-participating oncology practices and NCI's SEER Program.

- The National Cancer Database (NCDB), is sponsored by the partnership of the American College of Surgeons and the American Cancer Society. The NCDB obtains clinical oncology data from over 1,500 Commission on Cancer-accredited hospitals. These data are used to track and analyze the treatments and outcomes of patients who suffer from malignant neoplastic disease. The NCDB has information on approximately 70% of newly diagnosed cancer cases in the US.

- The Oncology Research Information Exchange Network (ORIEN) was founded by the Moffitt Cancer Center in Tampa, FL, and the Ohio State University Comprehensive Cancer Center. With prominent collaborating members, ORIEN collects data from some of the nation's leading cancer centers. The database analyzes these data to match patients with personalized treatments. ORIEN members have access to information from over 100,000 patients' molecular, clinical, and epidemiological data.

- OPTUMIQ is OPTUM's health care intelligence database comprising more than 26,000 expert opinions, data, and analytics from EHRs, medical claims, and prescriptions. Together, this intelligence lends itself to solutions including population analytics, pharmacy care services, and life sciences data and tools.

- IBM Watson Health's MarketScan Database has information on more than 240 million patients, over the last 23 years. MarketScan collects information on medical claims, EMRs, hospital databases, and patient clinical information to deliver physicians a complete look at their patient's disease, treatment options, and outcomes data.

- The Cancer Research Network (CRN), first funded in 1999 by NCI, is a coalition of non-profit organizations focused on research involving integrated health care delivery systems. The CRN promotes innovative and novel cancer research through collaboration and transparency between members. The CRN consists of nine health care institutions, cumulatively delivering care to almost 9 million patients.

- As one of the largest commercially insured population databases in North America, HealthCore Integrated Research Database (HIRD) is able to connect its data with clinical reports. HIRD has information regarding patient eligibility, prescription data, physician and facility claims data, and lab test results data. HIRD is beneficial for evaluating outcomes in distinct populations, due to its broad geographic representation. HealthCore does not work exclusively with clinical oncology data—they have data on a broad array of health care specialties. Their database includes information on approximately 65 million individuals, sourced from many commercial health care plans in the United States.

- IQVIA (previously IMS Health) uses Human Data Science, the integration of human science data and data science, to provide partners with evidence networks and analytics. IQVIA has information on more than 15 million cancer patients, globally, over the past 10 years. Using these data, IQVIA provides consultation on several oncological interests such as clinical trials, health care delivery, and commercial optimization.

- Founded in 2010, Syapse was created to make sure all cancer patients had access to precision treatment. Their software enables health care providers to deliver patient-centered cancer care by collecting clinical, molecular, treatment, and outcomes data to a single platform. Syapse's network of data allows partners access to health outcomes data for analysis and examination.

- Bristol-Myers Squibb announced that it is using Flatiron Health's real-world data to accelerate its research and development efforts, as well as to improve its ability to generate additional evidence on the use of its cancer medicines outside of clinical trials. The two companies intend to form a joint scientific advisory board to advance the use of real-world evidence for regulatory decision making.

- COTA, Inc., announced a 2-year research collaboration agreement with the US Food and Drug Administration (FDA) Information Exchange and Data Transformation (INFORMED) Program, to establish a study protocol with an initial focus on breast cancer. This effort will provide the FDA with information on the evolving treatment landscape, including insight on treatment variation within defined subpopulations of patients with the disease. As the project advances, the research may expand to other cancer types and explore disease characteristics, treatment patterns, and patient outcomes for the broader application of precision medicine.

The National Cancer Institute does not endorse or recommend any commercial products, processes, or services. Therefore, mention of commercial products, processes, or services in this report cannot be construed as an endorsement or recommendation.

Opportunities for Researchers

In addition to encouraging scientific ideas for researchers through investigator-initiated applications and omnibus solicitations, DCCPS develops and participates in NIH funding opportunities aimed at stimulating new directions to examine, discover, and test methodologies to improve public health. The following are examples of recent Funding Opportunity Announcements to encourage research projects in a variety of priority areas and are appropriate for projects that include collection and sharing of post-diagnosis observational data.

- Modular R01s in Cancer Control and Population Sciences (PAR-18-869)
- Core Infrastructure and Methodological Research for Cancer Epidemiology Cohorts (PAR-17-233)
- Integration of Individual Residential Histories into Cancer Research (PA-17-298, PA-17-295)
- Leveraging Population-based Cancer Registry Data to Study Health Disparities (PA-17-289, PA-17-288)
- New Informatics Tools and Methods to Enhance US Cancer Surveillance Research (PAR-16-349)

More information about funding opportunities can be found at cancercontrol.cancer.gov/funding.html.

Looking Ahead

Cancer care has been transitioning gradually to systems that put greater focus on outcomes, value, and patient choice. At the same time, there has been a growing realization that post-diagnosis observational data, or real-world data, can provide critical insights that can be difficult—or even impossible—to gather from more traditional evidence-generation approaches such as clinical trials. For these reasons, researchers, health care providers, commercial entities, and nonprofit organizations are working ever more closely together to ensure that high-quality, integrated real-world data from multiple sources are findable, accessible, interoperable, and reusable for the cancer research community.

> **"What types of data do we still need to answer critical questions about cancer care and outcomes?"**

The evidence garnered from post-diagnosis observational data analyses are most impactful when the types of data analyzed are as complete and comprehensive as possible. With that understanding, our next steps in DCCPS are to continue to assess the types of data collected in our current inventory of data sets, and then to work with our research community, partners, and stakeholders to identify critical missing elements in those data sets by asking, "What types of data do we still need to answer critical questions about cancer care and outcomes?"

In identifying such gaps, we will be better positioned to consider opportunities for data enhancement and cross-linking that will allow us to capture a more robust, fully representative picture of the disparate factors, alone and in concert, that affect cancer care outcomes.