National Cancer Institute

Surveillance Research Program

Strategic Plan

2010

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

Division of Cancer Control and Population Sciences

The National Cancer Institute's Charge

"… collect, analyze, and disseminate all data useful in the prevention, diagnosis, and treatment of cancer …"

(The National Cancer Act of 1971 [P.L. 92-218]).

Contents

Executive Summary 1

Chapter 1: Introduction 7

The Surveillance Research Program 8

Development of the Vision for the NCI Surveillance Research Program 10

Chapter 2: NCI's Surveillance Research Program: Where Are We Now? 13

Collection and Management of Data 14

Developing Statistical Methods and Modeling 16

Methods To Analyze Population Registration Data 16

Modeling and Interpreting Trends 21

The Surveillance Research Program's Biostatistical Methodology Portfolio 21

Reporting and Interpreting Data 23

Chapter 3: The Strategy: Where Do We Go From Here? 25

Priority Area 1: Provide a more complete depiction of the burden of cancer in the United States. 28

Priority Area 2: Collect better quality data on the burden of cancer in the United States more efficiently through effective utilization of electronic tools. 31

Priority Area 3: Improve understanding of the differences and disparities in the burden of cancer in the United States. 33

Priority Area 4: Better understand the continuum in the burden of cancer in the United States from risk to prognosis. 35

Priority Area 5: Communicate cancer statistics more effectively to researchers and users and make the data more accessible and understandable to all. 37

Appendix A: Acknowledgements 41

Appendix B: NCI Resources and Tools for Cancer Surveillance Research 42

Appendix C: List of Acronyms 43

Executive Summary

Cancer accounts for nearly one-fourth of all deaths in the United States and ranks high among causes of mortality. Monitoring cancer trends and related factors is critical to reducing the burden of cancer. The National Cancer Institute (NCI) supports cancer surveillance research to study these trends and quantitatively measure cancer incidence, morbidity, survival, and mortality for persons with cancer in the United States. Cancer surveillance also assesses genetic predisposition, environmental and behavioral risk factors, screening practices, and the quality of care from prevention through palliation. In short, cancer surveillance measures progress made toward reducing the burden of cancer and provides a basis for advancing research and interventions across the cancer control continuum in prevention, early detection, diagnosis, treatment, and outcome. Appropriate decisions in science and public health depend on reliable data about the impact of efforts to control cancer, and cancer surveillance research provides a basis for the public to assess the efficacy of National efforts to detect and treat cancer in the general population.

NCI's Surveillance Research Program (SRP), housed within the Division of Cancer Control and Population Sciences (DCCPS), directs the collection and dissemination of information on cancer incidence, prevalence, mortality, and survival in the United States. SRP collaborates with other DCCPS Programs in these efforts, particularly with the Applied Research Program (ARP), which addresses risk factors, health behaviors, health care services, economics, and outcomes, including patient-reported outcomes. In addition, the Behavioral Research Program (BRP) complements SRP and ARP efforts in cancer surveillance research through a portfolio of behavioral interventions, such as tobacco use, screening, dietary behavior, and sun protection. The programs, priority areas, and action plans described in this Strategic Plan are focused on activities managed by SRP.

SRP also works with other entities within NCI, the National Institutes of Health (NIH), and the Centers for Disease Control and Prevention (CDC), and with other federal agencies, and has developed strong partnerships with cancer registry and surveillance organizations. SRP's rigorous quality standards have made it a model for cancer surveillance activities throughout the world.

One of SRP's premier projects is the Surveillance, Epidemiology, and End Results (SEER) Program, which collects and manages high-quality data from cancer registries in specific geographic areas that represent 28 percent of the U.S. population. These data are used increasingly to

answer questions about cancer etiology, prevention, treatment, and control. SEER collaborates with other agencies and programs to link with additional datasets and enhance cancer care and outcomes data. These include Medicare, Medical Health Outcomes Survey (MHOS), Consumer Assessment in Healthcare Providers and Systems (CAHPS), and the National Longitudinal Mortality Study (NLMS). The SEER Residual Tissue Repository (RTR) Program also plays an important role in data collection and management.

During the past decade, SRP has fostered standardization practices across the cancer surveillance field. The standardization of cancer registry data elements has allowed investigators to compare cases, treatments, and outcomes to support cancer research activities and promote public health. In addition, SRP has promoted the adoption and use of a standardized vocabulary across the surveillance arena. NCI is an active participant in the American Joint Committee on Cancer's (AJCC) collaborative staging efforts, works with numerous partners to standardize data items across the cancer registry community, and works to develop infrastructure and quantitative methodologies.

During the past 10 years, SRP also has built a substantial collection of methodologies for statistical analysis software, applications, and modeling, including the Cancer Intervention and Surveillance Modeling Network (CISNET), SEER*Stat, Joinpoint, and DevCan. These data and tools are used to distill cancer trends in populations, and to track health disparities among subpopulations. SRP's grant portfolio for biostatistical cancer research is available online through the StatFund resource.

Integrating SRP activities with other NCI and NIH efforts helps to advance the battle against cancer. For example, the SEER Data Management System (SEER*DMS), used by cancer registries across the United States, complies with the philosophy and guidelines of the cancer Biomedical Informatics Grid (caBIG®), which provides an information network that allows members of the cancer community to share data and knowledge openly, in a standardized manner, using a common vocabulary. In addition, NCI's Center for Bioinformatics and SRP collaborate on electronic data capture activities, including by co-sponsoring the Tools for Electronic Data request for proposals (RFP).

SRP communicates cancer statistics to a variety of data users through online tools and annual and biennial reports. Cancer Control P.L.A.N.E.T. (Plan, Link, Act, Network with Evidence-based Tools), State Cancer Profiles, *Cancer Statistics Review* (CSR), and FastStats are among the Web-based resources instituted by SRP during the past decade. The *Annual Report to the Nation* describes cancer trends in the United States, and the *Cancer Trends Progress Report* highlights progress made against cancer in relation to Healthy People 2010 initiative targets.

Ten years following the *Cancer Surveillance Research Implementation Plan of 1999*, SRP convened extramural scientists and NCI staff with expertise in cancer surveillance to develop a vision for the next 5 to 7 years. In 2008, staff from SRP, DCCPS' Applied Research Program (ARP), and SEER-supported registries began a process to determine the future direction for SEER through a series of meetings entitled "SEER: Visioning the Future." The consensus was that the SEER Program should assess its impact on the field of population surveillance, especially in terms of the quality and depth of its data, accessibility and usability of data, leadership in defining data elements, innovation in data collection, and development of analytical methods applied to surveillance data.

SRP also held a more comprehensive strategic planning workshop in October 2009 to address scientific opportunities, challenges, and gaps in the cancer surveillance arena, with an emphasis on incidence, prevalence, mortality, and survival. Extramural participants and NCI staff considered the scope and content of cancer surveillance, analytic methods and models, integrated health informatics, populations with regard to differences in risk and prognosis, and health communication.

Table 1. Cancer Surveillance Priority Areas and Action Plan, 2010

Priority Areas and Action Plan 2010	
PRIORITY AREA 1	Provide a more complete depiction of the burden of cancer in the United States, past, present, and future.
Action Plan	1. Link with other data sources.
	2. Develop processes to obtain and use other types of variables and data.
	3. Collect more detailed individual data from a limited number of registries.
PRIORITY AREA 2	Collect better quality data on the burden of cancer in the United States more efficiently through effective utilization of electronic tools.
Action Plan	4. Identify and evaluate relevant data sources.
	5. Automate the capture of information.
	6. Expand interoperability and standardization of surveillance data.
	7. Use data-capture techniques and statistical modeling to improve the timeliness of cancer statistics reporting.
PRIORITY AREA 3	Improve understanding of the differences and disparities in the burden of cancer in the United States.
Action Plan	8. Enhance measurements of health disparities.
	9. Expand data visualization and interpretation.
PRIORITY AREA 4	Better understand the continuum in the burden of cancer in the United States from risk to prognosis.
Action Plan	10. Evaluate the potential impact of prevention, early detection, and treatment on cancer statistics to inform cancer research and policy development.
	11. Identify and track a cohort of cancer patients to obtain data and monitor impacts at the population level.
PRIORITY AREA 5	Communicate cancer statistics more effectively to researchers and users and make the data more accessible and understandable to all.
Action Plan	12. Conduct needs assessment to identify data needs and audiences and develop a plan to communicate cancer statistics and interpret data for audiences.
	13. Produce standardized communication materials.
	14. Disseminate information on existing resources to researchers and evaluate their effectiveness.
	15. Develop and disseminate methods and software for the analysis and presentation of National cancer statistics.

Results from the SEER Visioning and SRP strategic planning efforts resulted in this Strategic Plan for the future direction of NCI-supported cancer surveillance research. This Executive Summary presents an overview of the Strategic Plan's four priority areas and 15 research opportunities in these areas (Table 1). The full report provides more detailed information about NCI-supported cancer surveillance research efforts and future directions.

PRIORITY AREA 1

Provide a more complete depiction of the burden of cancer in the United States, past, present, and future.

Action Plan 1: ***Link with other data sources.*** New technologies such as geographic information systems (GIS) and the electronic health record (EHR) provide an opportunity to increase data collection in a cost-effective manner. Expansion should be accomplished by establishing linkages between SEER, the NCI Cancer Human Biobank (caHUB®) and caBIG®, and the National Cancer Data Base (NCDB), as well as deeper data mining of Centers for Medicare and Medicaid Services' (CMS) Medicare data. Other data linkages might be established with larger insurance systems and professional societies. The relationship between reporting sources and the contents of the registries' summary records should be explored to maximize the benefits achieved from linked data.

Action Plan 2: ***Develop processes to obtain and use other types of variables and data.*** As understanding of risk, prognosis, and population differences advances, more individual data should be collected to support this progression. Processes that are developed to collect and use these data should be sensitive to privacy, patient consent, and legal issues, and be designed with the flexibility to accommodate changes in scientific advances. Better ways to support an expanded collection of patients' treatment history, disease status, and followup, as well as genomic data, should be established to facilitate access to multilevel data from multiple sources while maintaining patient data confidentiality. A process for using U.S. Census information to link surveillance data to other data types, such as biomarker data, should be developed.

Action Plan 3: ***Collect more detailed individual data from a limited number of registries.*** The collection of quality data should be expanded by capturing more detailed individual data and specimens on a limited basis. A select number of registries should focus on acquiring and maintaining a high-quality collection in a specific area for several years, with relevant information about the process applied when planning future collection efforts. Partnerships also could be established with other successful collection efforts.

PRIORITY AREA 2

Collect better quality data on the burden of cancer in the United States more efficiently through effective utilization of electronic tools.

Action Plan 4: ***Identify and evaluate relevant data sources.*** Data sources are needed that expand the current boundaries of the SEER dataset and support the emphasis on population-based cancer surveillance. Relevant sources should be examined for the availability of electronic data that can be linked. Ongoing collaborations, such as those with CMS for the SEER-Medicare dataset, should be evaluated for the potential to obtain further information, including treatment data.

Action Plan 5: ***Automate the capture of information.*** The nationwide movement toward adopting EHRs provides a prime occasion to further automate the capture of cancer-related information. In addition, use of natural language processing (NLP) should be explored to ensure greater accuracy in the extraction of information from hospital reports.

Action Plan 6: ***Expand the interoperability and standardization of surveillance data.*** Standards for the collection and coding of health data from a variety of sources are needed to ensure the consistent quality and interoperability of data. Cancer surveillance organizations, researchers, and clinicians should form mutually beneficial collaborations to impose a structure for data collected at the point of care. Standardization should help to simplify the process of recording medical information by clinicians.

Action Plan 7: ***Use data-capture techniques and statistical modeling to improve the timeliness of cancer statistics reporting.*** Methods should be developed to reduce the current 3-year time period between diagnosis, notification of the registry about a case, and reporting of incidence rates by the registry. New data-capture techniques and statistical modeling could provide a means to expedite the reporting of cancer statistics, such as through a delay-adjustment model. In addition, a data-alert system should be designed to inform registrars about new information added to a case chart and lessen the need to review data that were evaluated previously.

PRIORITY AREA 3

Improve understanding of the differences and disparities in the burden of cancer in the United States.

Action Plan 8: ***Enhance measurement of health disparities.*** Methods for measuring health disparities and encouraging the use of tools such as the health disparities calculator (HD*Calc) should be improved. Census tract data on area socioeconomic status and other disparities indicators should be evaluated for their potential to link to cancer surveillance data and use in analyzing cancer disparities.

Action Plan 9: ***Expand data visualization and interpretation.*** To better identify and quantify cancer risk and prognosis, data about subpopulations should be identified geographically and analyzed. Methods of data visualization and interpretation should be more fully explored, including geocoding techniques, spatial data, and spatial-temporal data, to determine relevant factors influencing risk or prognosis.

PRIORITY AREA 4

Better understand the continuum in the burden of cancer in the United States, from risk to prognosis.

Action Plan 10: ***Evaluate the potential impact of prevention, early detection, and treatment on cancer statistics to inform cancer research and policy development.*** Through the understanding of risk factor trends, use of early detection and treatment, and information on risks and efficacy from clinical trials, models can predict the impact on cancer incidence and mortality trends, as well as provide estimates of potential impact if interventions were more widely implemented in the population.

Action Plan 11: ***Identify and track a cohort of cancer patients to obtain data and monitor impacts at the population level.*** Tracking a cohort of cancer patients through the registry system will provide a comprehensive approach to population studies that is unavailable through multiple independent cancer registries and individual investigator-initiated studies. Guidance is needed for developing collaborative cohorts in the registry system, with access to cohort data available to investigators who are not participating in the cohort. The registry system would provide nationally representative samples to supplement data from more localized groups, allowing the opportunity to examine socioeconomic factors and other health disparities variables.

PRIORITY AREA 5

Communicate cancer statistics more effectively to researchers and users and make the data more accessible and understandable to all.

Action Plan 12: ***Conduct a needs assessment to identify data needs and audiences, and develop a plan to communicate cancer statistics and interpret data for audiences.*** A strategy should be developed to communicate cancer statistics and help researchers and other audiences interpret and report data with accuracy. The needs of various audiences, ranging from scientists to policymakers to the general public, should be evaluated and a plan to communicate cancer statistics should be developed that distinguishes approaches for these audiences. The needs assessment should identify existing products, ascertain the need for new materials, and recommend refinements to existing resources to encourage more effective use. The plan should outline steps to promote existing resources as well as advise on the translation of data and preparation of other communication materials for lay audiences.

Action Plan 13: ***Produce standardized communications materials.*** To present a coherent message about cancer rates and help reduce the misinterpretation of statistical data, standardized communication materials are needed. Standards should be adopted by leading organizations regarding a clear and unified message about understanding and using surveillance statistics. Consensus about common data elements and the creation of standardized templates that present data to different audiences should be initial steps toward this goal. Explanation of data limitations and cross-links between the SEER Web site and other credible Web resources should be provided to help researchers use data more effectively.

Action Plan 14: ***Disseminate information on existing resources to researchers and evaluate their effectiveness.*** Better methods to promote existing resources to researchers are needed. A systematic, heuristic review of existing resources and how they are promoted and accessed should be conducted to provide a structured approach that promotes these resources to the wider research community. SEER-Medicare has been used widely by the research community and provides a model of how research resources can be used to address a range of research questions. NCI also could facilitate access to local cancer surveillance data and make data from SEER and other registries more readily available.

Action Plan 15: ***Develop and disseminate methods and software for the analysis and presentation of national cancer statistics.*** Development of new methods and software to analyze and present national cancer statistics will help invigorate the field and inform research conducted on earlier stages of the cancer control continuum to increase the impact of research on cancer mortality in the United States. One such tool is a cancer survival query system that could help researchers and other audiences obtain easier access to surveillance survival data.

Chapter 1
Introduction

The National Cancer Institute's charge:

"… collect, analyze, and disseminate all data useful in the prevention, diagnosis, and treatment of cancer …"
(The National Cancer Act of 1971 [P.L. 92-218]).

Cancer surveillance provides a portrait of cancer over time across a population. It offers a means to measure progress made against a disease that is estimated will affect one in every two people in the United States during their lifetime (*http://seer.cancer.gov/statfacts/html/all.html#risk*). Decisions in science and public health depend on reliable information about the impact of efforts to control cancer. Cancer surveillance research describes how well cancer has been controlled through the identification of risk factors and their prevention, disease detection and diagnosis, delivery of interventions to populations, and the effectiveness of those interventions.

Monitoring trends in the national cancer burden is important to determine what research is needed and to identify the best ways to build knowledge and translate results of cancer etiology and molecular science. Tracking cancer rates and trends also helps policymakers make sound decisions that promote public health and reduce the burden of cancer, and informs the U.S. public about the impact of cancer across the country.

Reductions in cancer incidence and mortality during the past 40 years have been realized through the amalgamated effect of a multifaceted campaign: programs to educate the public about adopting healthy lifestyles and the importance of screening for tumors; significant improvements in the treatment of cancers; and advances in biological, molecular, and genetic sciences that have provided the foundation for the development of effective agents and other therapies. Gene-environment interactions, genome mapping, proteomics, imaging, and bioinformatics are but a few fields that have burgeoned in the past 10 years to spur new discoveries, novel interventions, and sophisticated technologies, all in pursuit of helping cancer patients attain healthy status and preventing cancer. Cancer surveillance research has supported these advances by capturing data and analyzing rates for cancer incidence, prevalence, mortality, and survivorship in the United States. Researchers, physicians, and policymakers alike rely on this information to refine their investigations, treatment of patients, and health policies.

Cancer Surveillance and Research

Surveillance of disease trends is an important component of public health research.

Surveillance research about these trends, including prevention, treatment, and outcome of disease, offers a means to interpret and understand surveillance trends. It provides an avenue to create and validate new ways of conducting surveillance.

NCI has built a premier cancer surveillance research enterprise since the passage of the National Cancer Act of 1971, nearly 40 years ago. Based on a population science approach, NCI's surveillance research programs and activities serve as the most authoritative sources of information about cancer incidence, prevalence, mortality, and survival in the United States.

Cancer surveillance data show overall reduction or increases in cancer incidence and mortality in the United States. Societal effects—such as the effect of tobacco smoking bans in public locations on lung cancer incidence or the reduction of hormone replacement therapy and ensuing decline in breast cancer—can be tracked in tandem with peaks and valleys in incidence and mortality rates. Effects due to diagnosis of celebrities' cancers followed by a dramatic rise in screening also are documented, as seen in the case of President Ronald Reagan and prostate cancer. The *Annual Report to the Nation* provides cancer trends in the United States. For example, the 2009 report described the decrease for mortality rates overall—and specifically for male lung, male bronchus, cervical, and colorectal cancers—between 1975 and 2006 for all races/ethnicities monitored.

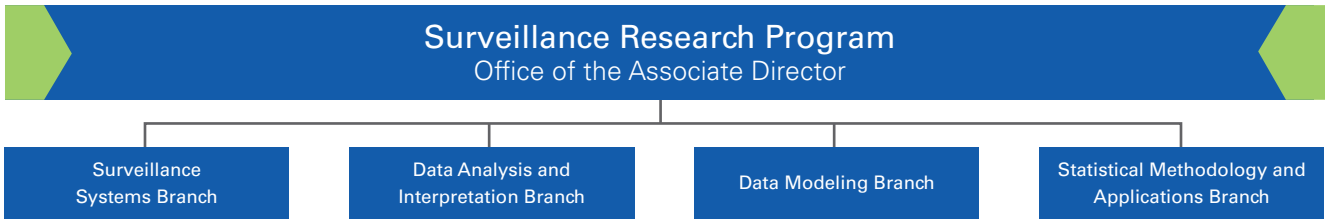
The Surveillance Research Program

The National Cancer Institute's (NCI) Surveillance Research Program (SRP) is one of the most authoritative sources of information on cancer incidence, prevalence, mortality, and survival in the United States. Housed within NCI's Division of Cancer Control and Population Sciences (DCCPS), SRP characterizes the cancer burden borne in the United States over time by integrating medical record data on persons diagnosed with cancer with a range of other relevant data from organizations such as the Social Security Administration, National Center for Health Statistics (NCHS), Centers for Medicare and Medicaid Services (CMS), and the U.S. Census Bureau. Its best-known project is the Surveillance, Epidemiology, and End Results (SEER) Program, which collects and manages high-quality data that provide a means to measure the progress made in reducing the Nation's cancer burden. These data are used by researchers in the areas of cancer prevention, treatment, and control. Through its statistical and modeling activities, particularly the Cancer Intervention and Surveillance Modeling Network (CISNET), SRP provides information and feedback about trends over time and location across the cancer control continuum. Other statistical methodologies and tools developed by SRP staff to analyze trends and evaluate the impact of cancer control interventions and other factors on the cancer burden include SEER*Stat and Joinpoint. SRP's data provide a snapshot that illustrates to the public how the war on cancer is faring, particularly with regard to health disparities and the effect of cancer on minority populations. Appropriate decisions in science and public health depend on reliable data about the impact of efforts to control cancer.

To accomplish its mission, SRP is organized into four branches that address surveillance systems, data analysis and interpretation, data modeling, and statistical methodology research (**Figure 1**). Each branch enables the growth of surveillance research through the management of grant portfolios and development of research infrastructure through cooperative agreements. Since 1999, SRP's portfolio has doubled in size. SRP also responds to more than 2,000 requests from investigators for data each year.

SRP collaborates with other extramural programs within DCCPS and other NCI Divisions to collect and analyze population-based data across the cancer continuum. Synergistic

Figure 1. SRP Organization



SRP Mission Statement

SRP directs the collection and analysis of pertinent data in order to answer key questions about cancer incidence, morbidity, mortality, and cancer-related health status in diverse regions and populations in the United States.

collaborations occur particularly between DCCPS' Applied Research Program (ARP) and Behavioral Research Program (BRP), which focus on risk factors, health behaviors, health care services, economics, and outcomes, including patient-reported outcomes (ARP); and behavioral interventions, such as tobacco use, screening, dietary behavior, and sun protection (BRP).

SRP has developed close partnerships with many cancer surveillance organizations, including other federal agencies, and private sector and academic institutions, in all aspects of cancer data and statistics. In turn, partners and researchers alike depend on SRP for SEER's high-quality data as well as for assistance in analyzing data and interpreting cancer statistics. As a comprehensive, population-based reporting system, SEER routinely collects cancer data from designated population-based cancer registries in various areas of the country (see Figure 2). Areas not covered directly by SEER are covered by the Centers for Disease Control and Prevention's (CDC) National Program of Cancer Registries (NPCR). NPCR and SEER data are integrated and presented in summary form in periodic reports produced by the CDC and International Association of Cancer Registries (IACR). In addition, these data are presented in the *Annual Report to the Nation,* produced through collaboration between NCI, the American Cancer Society (ACS), and the North American Association of Central Cancer Registries, Inc. (NAACCR). Investigators in the United States as well as worldwide rely on the SEER's high-quality cancer data and statistics.

SRP's Major Partners

Federal Agencies

- CDC's National Program of Cancer Registries (NPCR)
- CDC's National Center for Health Statistics (NCHS)
- Centers for Medicare and Medicaid Services (CMS)
- U.S. Census Bureau

Professional and Private Organizations

- North American Association of Central Cancer Registries, Inc. (NAACCR)
- National Cancer Registrars Association (NCRA)
- American College of Surgeons (ACoS)
- Commission on Cancer (COC)
- American Cancer Society (ACS)
- National Coordinating Council for Cancer Surveillance (NCCCS)
- International Association of Cancer Registries (IACR)

Development of the Vision for the NCI Surveillance Research Program

Cancer Surveillance Research Implementation Plan of 1999

In 1999, NCI's Cancer Surveillance Research Program (CSRP) prepared NCI's *Cancer Surveillance Research Implementation Plan of 1999* (hereafter the *1999 Strategic Plan*). The Surveillance Implementation Group (SIG) led the planning process, identified research directions and priorities, and produced an implementation plan for a comprehensive, focused, and coherent vision for NCI-funded surveillance research. Composed of 42 leading scientists and experts representing NCI, other federal agencies, and the extramural community, the SIG identified research opportunities within five priority areas (**Table 2**) to expand and enhance cancer surveillance research supporting the control of cancer.

Based on the Plan's recommendations, additional data on risk factors, screening, quality of life, patterns of care, treatment effectiveness, and biologic samples have been collected, and the *Annual Report to the Nation* has been published in collaboration with the CDC, ACS, and NAACCR, providing a narrative and visual overview of the state of cancer trends each year.

2005 National Framework for Cancer Surveillance

NCI's 2005 *National Framework for Cancer Surveillance* provided additional direction to the cancer surveillance field, describing the cancer burden from diagnosis to dying with cancer at the national, state, and local levels.

SEER Visioning Process

In April 2008, NCI staff from SRP and ARP met with staff from 15 SEER registries to discuss a vision for SEER, including new directions and strategies. Participants concluded that the SEER mission should continue to focus on the collection of high-quality, timely, population-based data that describe the cancer burden—specifically, cancer incidence, outcomes, trends, and patterns of care. New directions for SEER included an expansion of data collection to cover the entire cancer continuum (prevention through survivorship, including quality of life), and dissemination of those data to a wider range of stakeholders in the area of cancer control and prevention. Visioning participants also emphasized the need for SEER to support the collection and analysis of cancer data over time and location, and to support infrastructure for surveillance research.

The SEER Visioning effort generated four priority areas:

- Timeliness, quality, and relevance of SEER data.
- Efficiency and flexibility of SEER registry processes.
- Improvement of the visibility of SEER activities and services. Visioning participants agreed that SEER needed to better define its unique contribution to the field of population surveillance.
- Expansion of SEER data sources and users.

Four work groups were formed to tackle these priority areas:

- The Core Data Timeliness Work Group, charged with (1) examining SEER data to determine elements that might be eliminated, and (2) identifying strategies for making SEER data available to the public sooner.
- The Registry Operations Work Group, charged with examining the operations of each registry to determine options for reducing complexity in data processing as well as unnecessary activities.
- The Data Use and Marketing Work Group, charged with examining ways to improve SEER's relevance and visibility, with the long-term goal of maximizing data utility/value to the community and other stakeholders.
- The Health Informatics Work Group, charged with (1) assessing the state of automation across registries, and (2) assessing ways to use automation to streamline processes and reduce operating costs.

Table 2. Cancer Surveillance Research Implementation Plan March 1999–2009

Research Opportunities	Implementation: Accomplishments*
PRIORITY AREA 1 Expand the scope of surveillance research through additional data collection and methods development.	
1. Collect data on patterns of care, quality of life, and morbidity in cohorts of registered cancer patients.	• SEER Patterns of Care Studies yearly • HEAL (breast cancer cohort) • Rapid Response Surveillance Studies
2. Collect data on risk factors and screening in defined populations.	• Rapid Response Surveillance Studies • National Health Interview Survey • California Health Interview Survey • Breast Cancer Surveillance Consortium
3. Develop research methods to improve measurement of cancer burden trends.	• CISNET (breast, prostate, colorectal, lung) • SEER-DMS • NLMS linkages • Combining survey for small area estimation • Prevalence measures
4. Explore geographic information system (GIS) methods.	• State Cancer Profiles (P.L.A.N.E.T.) • Geospatial and data visualization methods
PRIORITY AREA 2 Expand the scope of surveillance to improve the representativeness of cancer burden estimates.	
5. Expand to improve representation of ethnic minority and underserved populations.	• 2000 SEER expansions increase coverage to 26% • Publications: Hispanics, AI/AN, API; survival • Link to Indian Health Service
6. Explore methods for improved national estimates of cancer burden.	• Estimation of # new cancer cases in U.S. this year; projections; delay adjustment; Joinpoint; cure models; prevalence by phase of care • HD*Calc
7. Work with partners on National Cancer Surveillance Plan.	• National Coordinating Council for Cancer Surveillance (framework); NAACCR; CDC; ACS; C Change
PRIORITY AREA 3 Produce and disseminate a national report card on the cancer burden.	
8. Proceed with production of Report Card for Year 2000.	• *Annual Report to the Nation* (1998–2009) • Cancer Progress Trends Report
9. Develop plan to improve dissemination of surveillance data.	• Web-based, interactive SEER reports • Cancer Fact Sheets updated annually • 2005 Workshop for Advocates • 2000+ requests for SEER research file
PRIORITY AREA 4 Support molecular and genetics research for surveillance.	
10. Develop tools to assess family history and prevalence of familial cancers.	• Survey on family history
11. Explore the feasibility of collecting biospecimens to support population-based molecular and genetic research.	• Residual Tissue Repository
PRIORITY AREA 5 Develop a training strategy for cancer surveillance research.	
12. Identify training needs and develop a plan to incorporate surveillance sciences into training mechanisms.	• Surveillance Institute 2005 • 1-week summer course (Johns Hopkins University) • Workshops at NAACCR • SEER*Stat training

* This list includes a number of initiatives that are managed by Programs across DCCPS, notably ARP.

Each of these groups developed several recommendations for SEER within the assigned focus areas. Specific activities of the work groups to date include:

- Development of a process diagram for evaluating data items.
- Analysis of SEER cost study data to determine common and best practices across registries.
- Working with a communications specialist to assess the brand awareness of SEER.
- Conducting an online survey of registry principal investigators and managers to determine levels and types of automation at the 18 SEER registries.

2010 Strategic Planning Process for NCI-Supported Cancer Surveillance Research

On October 22–23, 2009, SRP held a Cancer Surveillance Research Workshop to examine potential directions for surveillance research during the next 3 to 7 years, building on achievements from the past decade and from the strategic planning efforts outlined in the *1999 Strategic Plan*.

The workshop covered five topic areas: (1) scope and content of cancer surveillance; (2) development of analytic methods and models to support surveillance research; (3) integrated health informatics; (4) population subgroups to understand differences in risk and prognosis; and (5) communications and dissemination of data to users. NCI staff and guest speakers presented on each topic, and participants broke into small groups to develop, discuss, and prioritize recommendations. Discussions covered the upsurge in information available and its growing complexity, as well as challenges presented in an era of collaborative efforts, pooled data, and shared resources, with greater uncertainties related to funding and staffing.

Specific considerations during the workshop included: data items for specific purposes; building and analyzing multilevel data systems; analysis and interpretation of integrated information; development and application of analysis tools for new data submissions and incomplete datasets; and application of advanced statistical methods, including classical and Bayesian.

Based on discussions from the October 2009 workshop and from ongoing conversations between NCI staff and SEER registries that commenced with the SEER Visioning meeting in 2008, as well as dialogues with other principal investigators and colleagues, priority areas and research opportunities have been identified as future directions for NCI-supported cancer surveillance research. These are described in detail in Chapter 3.

Chapter 2

NCI's Surveillance Research Program: Where Are We Now?

Nearly 40 years after the passage of the National Cancer Act of 1971, SRP serves as a leader in the Nation's robust cancer surveillance enterprise. At the heart of its efforts, SRP measures the impact made on the cancer burden through shifts in cancer incidence, morbidity, survival, and mortality, particularly by evaluating data about genetic predisposition, environmental and behavioral risk factors, screening practices, and the patterns and quality of care provided throughout the cancer disease process, from prevention through outcome.

SRP has evolved over the years as the demand for new and expanded data and more sophisticated analyses increases, refinements to the collaborative staging of cancers result in additional data elements to be collected, advances are made in cancer biology and etiology, and populations desire more information about the public health impact of science research. SRP's approach is to anticipate research needs and cancer trends, integrate data into the SEER dataset collection, and remain both focused and flexible in terms of resources, infrastructures, and priorities. Health disparities figure largely in monitoring the cancer burden as cancers differ among racial/ethnic groups, and population diversity is increasing in the United States.

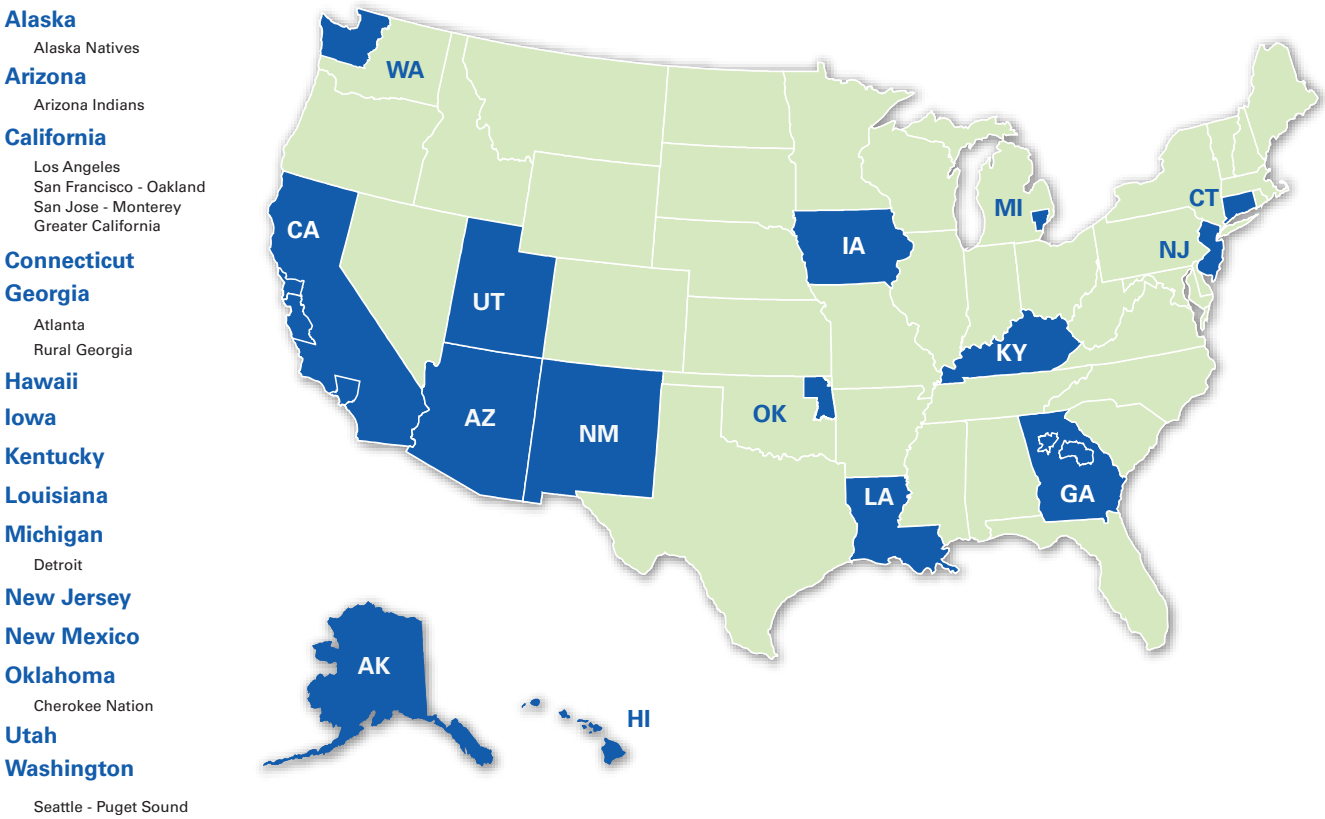
Uses of SEER

SEER is used in many research studies to understand disease trends and as a source for etiologic cancer research. It serves as a major pillar for many research infrastructures (e.g., SEER-Medicare) that enable a wide range of research studies on cancer-related topics and SRP statistical methods research techniques used for many other diseases. SEER provides a gold-standard cancer surveillance methodology for the United States and the world.

Researchers rely on SEER to:

- Track population trends.
- Understand factors influencing trends.
- Analyze cancer incidence, prevalence, survival, and mortality at the state/local level.
- Understand factors related to cancer incidence, prevalence, survival, and mortality.
- Describe health disparities among subpopulations.
- Validate cancer statistics from other sources.
- Identify cancer patients for approved studies.
- Compile cancer statistics for policy decision making.

Figure 2. SEER Coverage Across the United States



SRP's cancer surveillance research programs and activities are best described in three broad activity areas: (1) collecting and managing data; (2) developing statistical methods and modeling; and (3) reporting and sharing data.

Collection and Management of Data

The SEER Program (*http://seer.cancer.gov*) serves as the primary vehicle for compiling data about incidence, prevalence, mortality, and survival, which researchers who are interested in cancer control, etiology, molecular genetics, health services, rates/trends, and other areas can use. SEER supports multiple population-based cancer registries, encompassing 26 percent of the U.S. population. During the past decade, SEER significantly expanded its data collection activities through its registries and through partnerships with the CDC and other data-collection efforts to capture incidence data across the United States (**Figure 2**). NAACCR data from 2002 to 2006 encompass 85 percent of the Nation; the U.S. Cancer Statistics (USCS) data in 2005 cover 96 percent of the United States.

Links with other datasets have enhanced cancer care and outcomes information. These include: SEER-Medicare data linkage (*http://healthservices.cancer.gov/seermedicare*); SEER-Medical Health Outcomes Study (MHOS) (*http://outcomes.cancer.gov/surveys/seer3-mhos*); and SEER-Consumer Assessment in Healthcare Providers and Systems (CAHPS), a database designed to improve understanding of how Medicare beneficiaries with cancer perceive the quality of their health care. Additional partnerships that have been pivotal in SEER's data collection efforts and in advancing cancer surveillance research include the SEER-National Longitudinal Mortality Study (NLMS) database and Residual Tissue Repository (RTR) Program.

Specific programs and activities that facilitate data collection and management include:

SEER-Medicare. A linkage between the SEER database and Medicare claims database provides a large, population-based source of information for cancer-related epidemiologic and health services research, particularly relevant for cancer care and outcomes (*http://healthservices.cancer.gov/seermedicare*). The SEER-Medicare data include information about all patients reported to the SEER registries who are Medicare eligible. The data have expanded from a focus on assessing the costs of care in 1992 to now including large numbers of cases, spanning from initial Medicare coverage until death, encompassing most clinical areas, representing a diversity of geographic areas, being population-based, and including data on multiple disease conditions. To ensure that patient/provider confidentiality is safeguarded, SEER-Medicare data are de-identified, and investigators are required to obtain approval before accessing or acquiring data.

NLMS Database and SEER-NLMS Linked Dataset. Data from SEER (1973–2003) have been linked with data from the NLMS dataset (1979–2002) to provide additional individual-level socioeconomic and demographic information on the SEER cancer cases, including cancer incidence, survival, and tumor characteristics based on self-reported factors (*http://surveillance.cancer.gov/disparities/nlms/*). The NLMS database includes records on approximately 3 million people as well as cause of death on more than 250,000 in the study population. Individual-level socioeconomic data collected by the U.S. Census Bureau through annual Current Population Surveys (CPS) are combined with cause of death information from official state death certificates for the period 1979–2002. The NLMS database includes information about: race/ethnicity, income, industry, nativity/immigrant status, marital status, employment status, veteran status, smoking, education, occupation, household size, and health status. Analyses of the data can be used to investigate mortality differentials across various socioeconomic and demographic groups.

RTR Program. The SEER RTR Program (*http://seer.cancer.gov/biospecimen*), established in 2003, maintains biospecimens obtained from three of SEER's population-based cancer registries (Iowa, Hawaii, and Los Angeles). RTR covers more than 60,000 cancer cases from diverse geographic regions (urban/rural) and racial/ethnic groups. It enables studies on rare cancers for which no single registry has enough cases to allow statistically valid conclusions to be drawn, allows validation studies on specimens from population-based registries, and allows comparisons of cases with biospecimens from all cases in the registry for the purpose of assessing population-based representativeness. RTR also allows analysis of trends in incidence and the potential for correlation with treatment trends based on the history and diversity of the SEER registries. Updates to survival data are possible after a tissue microarray is formed without violating confidentiality and privacy protections.

Rapid Response Surveillance Studies (RRSS). This vehicle allows studies to move expeditiously from the initial concept through completion in approximately 2 years. SRP has used RRSS to address various cancer surveillance issues, beginning with the SEER Patterns of Care Study in 1987 and extending to a broad range of topics related to cancer prevention and control (*http://seer.cancer.gov/rapidresponse*). Research areas addressed by these studies include evaluation of cancer surveillance methodologies and cancer treatment outcomes, and monitoring of screening practices linked to cancer outcomes. In addition, these studies have supported pilot and feasibility studies that address the development of procedures for enhanced monitoring of health behaviors and risk factors, expanded utility of SEER data through linkage with other databases, and improved cancer registry operations. RRSS studies have provided a foundation for larger research initiatives funded by other federal and private mechanisms.

SEER*DMS. The SEER Data Management System (*http://seer.cancer.gov/seerdms/*) is a hardware/software application that provides support for all core cancer registry functions. Centralized development has proved cost effective and resulted in improved data quality, quicker data reporting, and expansion of collected fields. Refinements are distributed

simultaneously to all registry locations, and the program can be customized at the local level, such as site-specific data fields and choice of matching algorithms. SEER*DMS reduces duplication of effort and ensures data quality and consistency among registries by facilitating the collection and reporting of high-quality cancer incidence, treatment, and survival data, and by enabling data sharing among users.

E-Path. Electronic Pathology (E-Path) software (*http://www.aim.on.ca/products/ePath.jsp*) provides a means to collect complete and accurate patient data from hospital and other pathology laboratories that report to SEER's cancer registries. E-Path reports cancers that are diagnosed through histology and comprise approximately 95 percent of all cancers. E-Path helps reduce errors and provides cost and labor savings by standardizing and automating the registries' work in identifying new cases and determining key characteristics required for classification standards of a cancer. NCI began installing E-Path within SEER registries in 2004. Currently, the software has been implemented in more than 200 reporting hospitals and laboratories throughout the United States.

Developing Statistical Methods and Modeling

Research methods have been developed and modeling tools designed and refined to improve national estimates and measurement of cancer burden trends. Cancer statistical data have been incorporated into models to help synthesize, interpret, and assess current and potential trends in cancer rates. SRP's efforts during the past decade encompass:

- Developing methods to analyze population registration data.
- Modeling and interpreting trends.
- Promoting biostatistical cancer research.

Methods To Analyze Population Registration Data

SRP has developed a number of statistical tools for population surveillance researchers to use in analyzing cancer data. These methods help scientists generate hypotheses that form the basis for cancer research and interventions for cancer prevention and control. The results of these analyses also are used to re-inform and redirect earlier stages of cancer research.

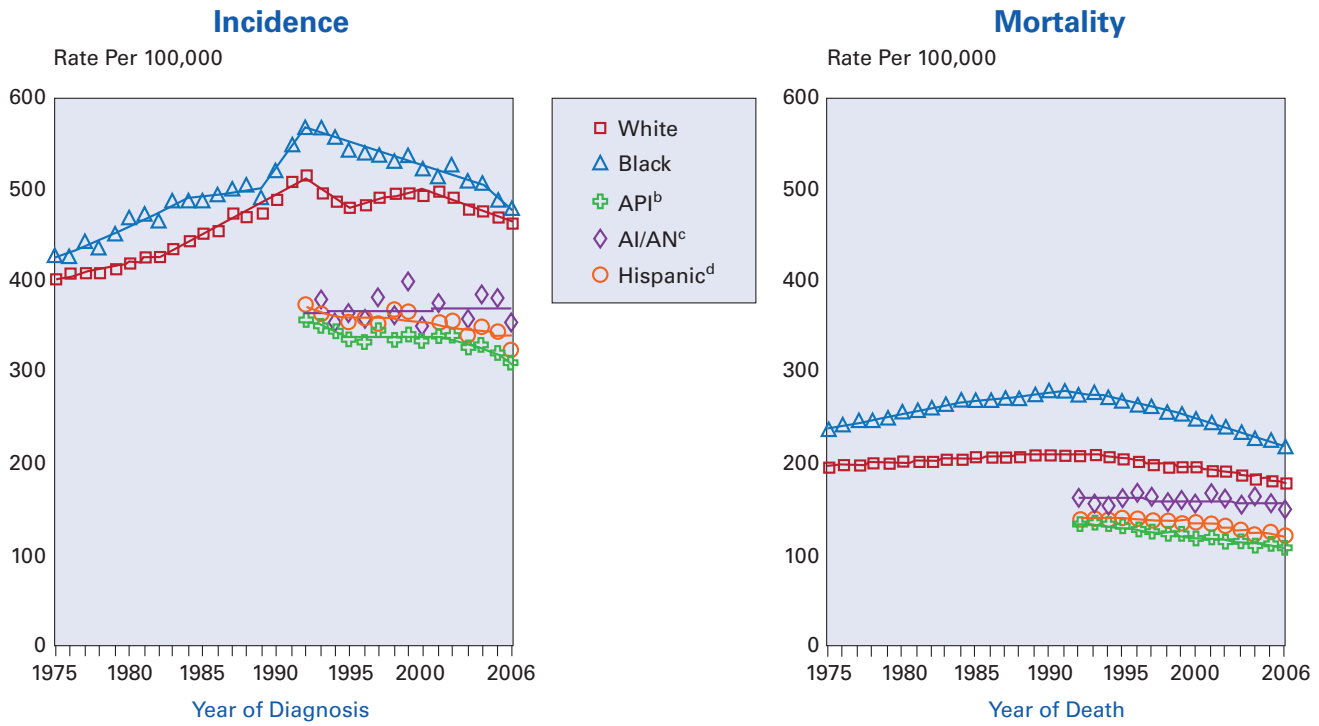
Many of the methods developed are included in SEER*Stat, a standard software program used for analyzing registry data, and other publically available software programs that work in conjunction with SEER*Stat, which has approximately 5,000 users nationally and internationally. Other statistical applications software programs have been designed to interface with SEER*Stat for ease of use with registry data. **Figure 4** highlights the major areas of methods developed to analyze population registry data and how software developed to apply these methods work together. Three examples that play a critical role in reporting annual statistics to the Nation are outlined below.

SEER*Stat. SEER*Stat (*http://seer.cancer.gov/seerstat/*) is a PC-based statistical software that provides a convenient, intuitive mechanism for the analysis of SEER and other cancer-related databases. It is a powerful tool for viewing individual cancer records and for producing statistics to study the impact of cancer on a population. SEER*Stat software is distributed with SEER Limited-Use Data and access to the data is necessary before using the software.

Joinpoint. Joinpoint (*http://srab.cancer.gov/joinpoint/*) is a statistical software program that analyzes trends using joinpoint models, that is, models where several different lines are connected together at the "joinpoints." Cancer trends reported in NCI publications are calculated using the Joinpoint Regression Program to analyze rates calculated by SEER*Stat software. The software takes trend data (e.g., cancer rates) and fits the simplest joinpoint model that the data allow. **Figure 3** provides an example of cancer incidence and mortality analyses resulting from Joinpoint software.

DevCan. DevCan (*http://srab.cancer.gov/devcan/*) is a statistical modeling software program that computes the probability of developing and dying of cancer from birth or conditional

Figure 3. Sample Joinpoint Analyses



SEER Incidence and US Death Rates[a]

All Cancer Sites, Both Sexes

Joinpoint Analyses for Whites and Blacks from 1975 -2006

Source: Incidence data for whites and blacks are from the SEER 9 areas (San Francisco, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, Atlanta).

Incidence data for Asian/Pacific Islanders, American Indians/Alaskan Natives and Hispanics are from the SEER 13 Areas (SEER 9 Areas. San Jose-Monterey, Los Angeles, Alaska Native Registry and Rural Georgia). Mortality data are from US Mortality Files. National Center for Health Statistics, CDC.

[a] Rates are age-adjusted to the 2000 US Std Population (19 age groups – Census P25-1103).

Regression lines are calculated using the Joinpoint Regression Program Version 3.3.1. April 2008, National Cancer Institute. Joinpoint analyses for Whites and Blacks during the 1975-2006 period allow a maximum of 4 joinpoints. Analyses for other ethnic groups during the period 1992-2006 allow a maximum of 2 joinpoints.

[b] API = Asian/Pacific Islander.

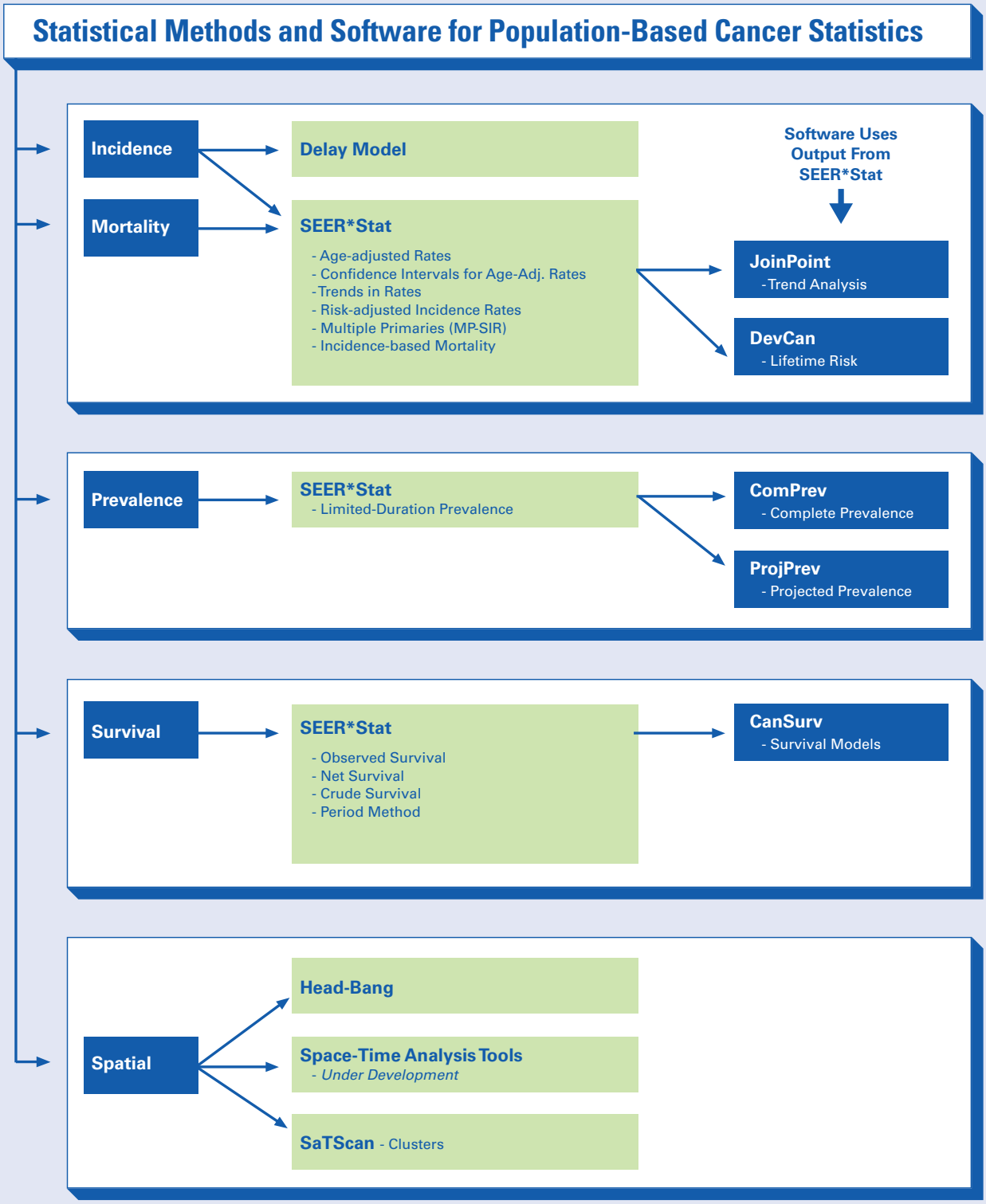
[c] AI/AN = American Indian/Alaskan Native. Rates for American Indian/Alaskan Native are based on the CHSDA (Contract Health Service Delivery Area) counties.

[d] Hispanic is not mutually exclusive from whites, blacks, Asian/Pacific Islanders, and American Indians/Alaska Natives. Incidence data for Hispanics are based on NIHA and Exclude cases from the Alaska Native Registry. Mortality data for Hispanics exclude cases from Connecticut, the District of Columbia, Main, Maryland, Minnesota, New Hampshire, New York, North Dakota, Oklahoma, and Vermont.

on a certain age. DevCan takes cross-sectional counts of incident cases from the standard areas of SEER plus mortality counts for the same areas from data collected by the NCHS and uses them to calculate the probabilities of developing or dying from cancer for a hypothetical population.

Because the SEER program is the authoritative source of information on cancer survival for the U.S. population, SRP has focused on developing methods to describe survival and estimate prevalence of cancer survivors. **Boxes 2.1** and **2.2** highlight accomplishments to date in these areas of research.

Figure 4. NCI Statistical Methods and Software for Cancer Surveillance Research



Box 2.1. Cancer Survival Statistics

Cancer survival statistics are concerned with the proportion of patients alive at some point subsequent to the diagnosis of their cancer. Relative survival is an estimate of the percentage of patients who would be expected to survive the effects of their cancer. Observed survival is the actual percentage of patients still alive at some specified time after diagnosis of cancer. A number of tools and metrics are available to estimate population-based cancer survival, depending on the purpose for using the data.

SEER*Stat Statistical Methods

- There are three measures of cancer survival that can be calculated in SEER*Stat software, including:
- Observed all cause survival—survival using as end point all causes of death.
- Net cancer-specific survival—cancer survival in the absence of other causes of death (the confounding effects of death from other causes are removed).
- Crude probability of death—probability of death from cancer and the probability of death from other causes, each estimated in the presence of the other.

Each of these measures can be calculated using relative survival by using the cause of death information listed on the death certificate. An ongoing area of research has been to investigate how to best use the information on the cause of death to minimize biases associated with misclassification or missing information.

Cancer Survival Analysis Software (Cansurv)

Cansurv is statistical software to analyze population-based survival data using mixture cure models and provide graphs and tests for inference and diagnosis. It can fit parametric (cure) survival models to grouped data or individually-listed data. Cansurv uses population-based survival data extracted from a SEER*Stat survival session.

Competing Risk Modeling

SRP is working with a variety of competing risk models to develop approaches to model the probability of dying from cancer and other causes that includes information on the cancer diagnosis and existing comorbidities at the time of diagnosis.These estimates are being developed as part of a Cancer Survival Query System (CSQS), which will provide up-to-date estimates of survival as a function of factors influencing both cancer and other causes of mortality for recently diagnosed patients.

Providing Up-To-Date Estimates of Survival

Survival estimates from cancer registry data usually are dated measures of current-year survival because of the time needed to observe survival and the time lapse between available data and the current year. Four approaches are used for grouping survival experience: (1) cohort, which uses the observed survival for a cohort of patients diagnosed in a single calendar year; (2) multiple-year cohort, which includes all patients diagnosed in the most recent years spanning the maximum duration to be estimated; (3) period, which uses only the most recent interval survival estimate of cases diagnosed in different calendar years; and (4) projection, which involves projecting survival trends to the current calendar year.

Box 2.2. Cancer Prevalence

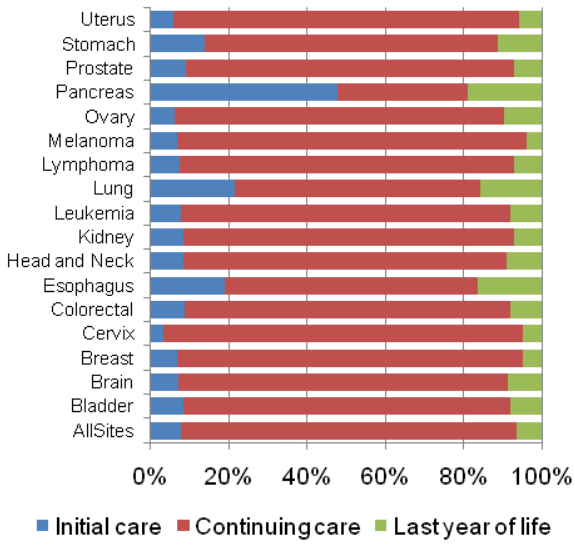
Cancer prevalence is the measure of the total number of people alive in a population ever diagnosed with cancer at a given time (referred to as complete prevalence).

Uses

Prevalence is used by government agencies, local health administrators, the research community, and cancer advocacy groups to:

- Describe the health of populations.
- Establish public health goals.
- Measure the burden of the disease.
- Understand the cancer survivor population.
- Evaluate and plan health resources allocations.
- Describe how rare a disease must be in order to be used for the orphan drug development process.
- Estimate costs of cancer care.

Figure 1. Estimates of the proportion of survivors in 2010 by cancer site and phase of care



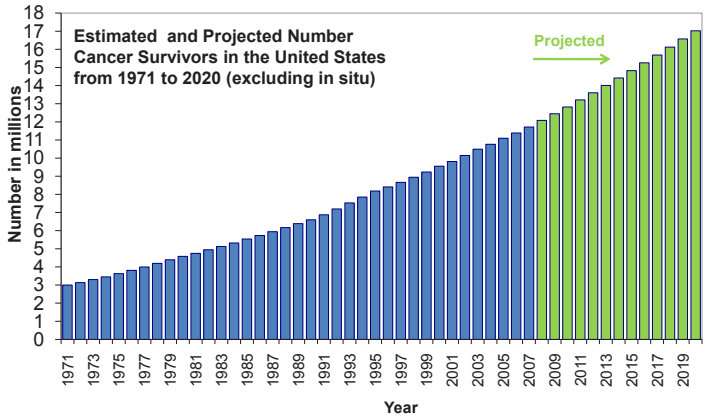
Methods of Estimation

- SEER*Stat software includes prevalence methods that estimate limited duration prevalence from cancer registry data. Limited duration prevalence represents the prevalence of people diagnosed with the disease for the maximum number of years that the registry has been collecting incidence data.
- COMPREV and PROJPREV are other software that allow researchers to estimate complete prevalence and the prevalence of people ever diagnosed, and to extrapolate prevalence to different populations.

Research

Prevalence across the cancer continuum is more useful; however, it is difficult to estimate because SEER collects information at the time of diagnosis and life status at the end of follow-up.

- Research is being conducted to estimate:
  - Survivors who will never die from the cancer (Cure prevalence).
  - Survivors who have recurrent cancer (Recurrence prevalence).
  - Survivors by phase of care: initial, last year of life, and phase in between.
  - Projections of cancer prevalence and costs of cancer care by different scenario assumptions of incidence and survival (*http://costprojections.cancer.gov/*).
  - Methods to estimate complete prevalence from cancer registries with 5 years of incidence data.



Data sources: NCI SEER November 2006 Submission, U.S. Estimated Prevalence counts were estimated by applying U.S. populations to SEER-9 and historical Connecticut Limited Duration Prevalence proportions. Populations from January 2004 were based on the average of the July 2003 and July 2004 population estimates from the U.S. Bureau of Census.

Modeling and Interpreting Trends

Modeling has played an important role in SRP's efforts to analyze data and interpret cancer trends. Mathematical and simulation modeling increase the usefulness of data for assessing progress in cancer control. SRP's primary modeling efforts to better understand and project patterns in cancer data have been made through the CISNET program.

CISNET. CISNET (*http://cisnet.cancer.gov/*) serves as SRP's principal vehicle for the statistical modeling of cancer control interventions to better understand interventions and population trends, particularly the impact of prevention, screening, and treatment on cancer incidence and mortality. CISNET's models have been used since 2000 to guide public health research and priorities by projecting future trends and determining optimal cancer control strategies for breast, prostate, colorectal, and lung cancers. CISNET was re-funded in 2010 and now also includes esophageal cancer. Risk factors, screening behaviors, and the diffusion of new treatments are analyzed to simulate potential effects and consider the costs versus benefits of interventions.

CISNET uses simulation modeling of disease processes and the impact of interventions on these processes to bring the most sophisticated evidence-based planning tools to the areas of population health and public policy. CISNET models can translate evidence from randomized trials and epidemiologic studies to the population health setting by extrapolating evidence beyond study protocols to the general population, accounting for actual usage in less controlled settings. Modeling real and hypothetical scenarios allows for the identification of key factors influencing outcomes and efficient cancer control strategies. The consortium's work informs clinical practice and recommended guidelines by synthesizing existing, albeit often incomplete, information to model gaps in available knowledge.

CISNET provides a suite of models to meet the challenges of the increasing pace of scientific discovery that are poised to address emerging questions. Collaborative work on key questions promotes efficient collecting and sharing of the most important data resources and critical evaluation of the strengths and weaknesses of each data source. A systematic comparative modeling approach brings transparency to the modeling process. Providing a range of models, rather than a single estimate from one model, brings credibility to the process and reassures the users of the models' results that the results are reproducible. CISNET investigators collaborate with groups such as the United States Preventive Services Task Force (USPSTF), CMS, CDC, ACS, and the American College of Radiology Imaging Network (ACRIN).

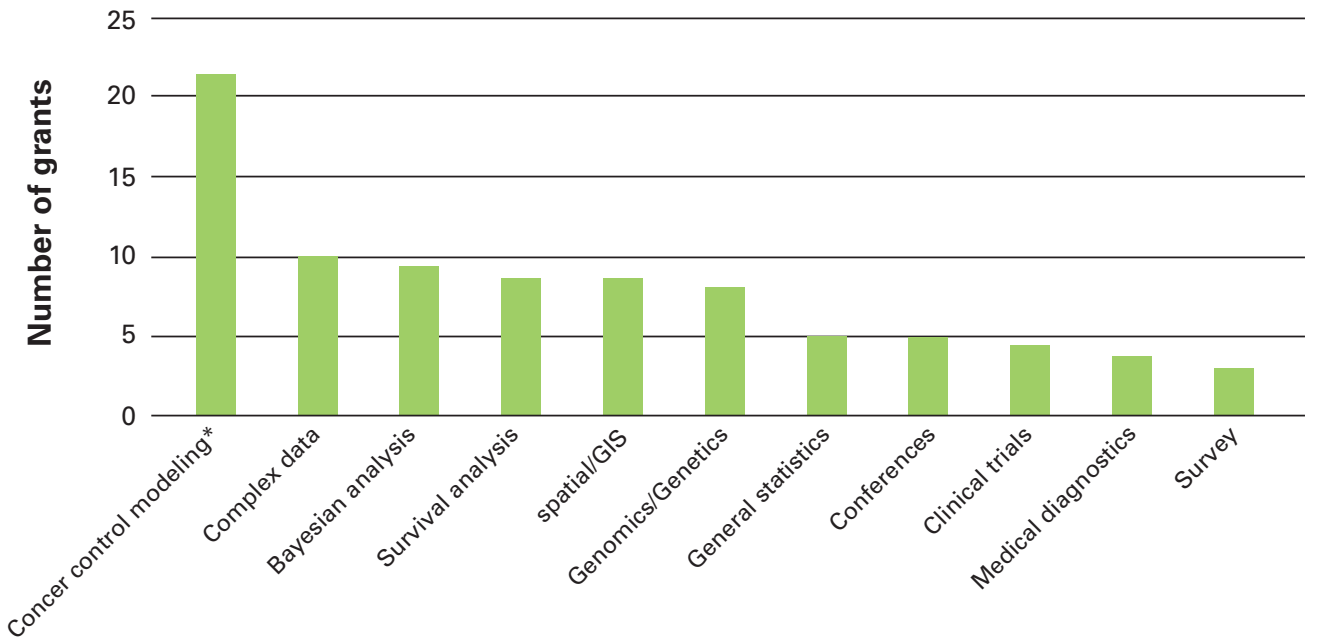
The Surveillance Research Program's Biostatistical Methodology Portfolio

SRP has a biostatistical grant portfolio currently consisting of 83 active grants (**Figure 5**). Most of these grants are research program grants (47 R01s), but also represented are cooperative agreements (15 U01s), small grants (10 R03s), exploratory grants (2 R21s), MERIT awards (2 R37s), conference grants (5 R13s), and Pathway to Independence Awards (1 K99/R00); AREA grants (R15s) previously were represented. Most of the Principal Investigators (PIs) are senior statisticians in statistics or biostatistics departments.

Many biostatistical methodology grants at NIH are cancer related, and the majority of these are administered in SRP. Some of these grants develop new statistical methodologies arising from collaborations with clinical or laboratory scientists, whereas others are translational. Examples of statistical methodology work in the SRP portfolio are:

- Development of methodologies for missing data in longitudinal models.
- Functional mixed models using wavelets to model images.
- Use of biomarkers as surrogate endpoints.
- Missing data in survival models.

Figure 5. SRP Grant Portfolio†



* includes CISNET Cooperative Agreements
† Data are from March 2010

SRP supports translational research, which aims to make the best use of already developed methodologies to benefit areas of health research other than its original purpose. One way that SRP supports translational research is by encouraging statisticians who are developing methodologies to create user-friendly software as well. Development of user-friendly software is supported through administrative supplements and Small Business Innovation Research (SBIR) contracts.

SRP is a strong supporter of the statistical community in ways other than administering statistical grants. SRP program directors regularly attend statistics meetings (such as the Joint Statistical Meetings and the spring meeting of the Eastern North American Region of the International Biometric Society), providing public service by educating new researchers about the NIH grant-funding processes. This is accomplished through presentations, organizing sessions, leading roundtables, and hosting exhibit booths. In addition, SRP communicates information about funding through overview articles published in bulletins such as *Amstat News* and through its Web site, *www.statfund.cancer.gov*, which contains funding information particularly useful for statisticians. StatFund provides notices about funding opportunities, NIH policies, and application procedures, as well as listing statistical grants that are administered by SRP.

StatFund. Information about biostatistical funding opportunities is available at *http://statfund.cancer.gov*. StatFund provides notices about SRP statistical grants, funding opportunities, and application procedures. NCI's statistical grant portfolio related to cancer control and population sciences encompasses many areas of research, including cancer treatment, epidemiology, prevention, outcomes research, dietary assessment, and cancer surveillance. Statistical topics range from longitudinal and spatial models to fixed, mixed, or random effects models, as well as survival and sequential analyses and Bayesian inference. Research results can be interpreted across the cancer control continuum and include prevention, detection, diagnosis, treatment, and survivorship. Many National Institutes of Health's (NIH) statistical grant applications are cancer related.

Figure 5 illustrates the topics covered in the SRP grant portfolio. The largest single category is cancer control modeling. The 44 grants classified under statistical methods represent a range of topic areas, such as survey analysis, analysis of complex datasets, survival analysis, Bayesian applications, and general statistical methodologies.

Reporting and Interpreting Data

In addition to collecting and analyzing data, a primary mission of SRP is to communicate cancer statistics to a variety of audiences. The form of the communication reflects the various needs of data users, ranging from evaluating progress toward meeting health objectives to resource allocation and planning. Cancer registry data provide insight into the identification of emerging trends through evaluation of cancer registry data. They also can indicate the need for additional interventions to provide evidence of the success or failure of previously implemented health policy. The usefulness of the data collected hinges on SRP's ability to communicate the information to potential users, including the public.

Cancer Control P.L.A.N.E.T. The Plan, Link, Act, Network with Evidence-based Tools portal (*http://cancercontrolplanet.cancer.gov/*) provides access to comprehensive cancer control data and resources for public health professionals to assist with the design, implementation, and evaluation of evidence-based cancer control programs.

State Cancer Profiles. State Cancer Profiles (*http://statecancerprofiles.cancer.gov/*) is a collaborative effort between NCI and CDC that provides graphic portraits of cancer statistics for prioritizing cancer control efforts in the Nation, states, and counties. Interactive maps, graphs, comparison tables, and support data are available for incidence rates, death rates, rate/trend comparisons, and prevalence projections to characterize the cancer burden in a standardized manner. Historical trends and interactive maps also are available.

Annual Report to the Nation. The *Annual Report to the Nation* (*http://seer.cancer.gov/report_to_nation/*) describes cancer trends in the United States with data analyzed by anatomical site, population, gender, and other categories. The 2009 report devotes special attention to colorectal cancer and the use of microsimulation modeling as a tool for interpreting past and future trends that affect cancer control planning and policy decisions.

Cancer Trends Progress Report. The *Cancer Trends Progress Report* (*http://progressreport.cancer.gov/*) highlights progress against cancer in relation to Healthy People 2010 targets. The report includes key measures of progress along the cancer control continuum and uses national trend data to illustrate where advances have been made.

Cancer Statistics Review. The annual *Cancer Statistics Review* (CSR) (*http://seer.cancer.gov/csr/1975_2006/index.html*) is an online resource that reports cancer incidence, mortality, survival, prevalence, and lifetime risk statistics from 1975 through the most recent year for which data are

available. The current report presents statistics on 27 primary sites and subsites, organized into site-specific chapters, and includes delay-adjusted cancer incidence rates. Features recently added to the CSR include detailed histology breakdowns for lymphomas and for cancers of the oral cavity and pharynx, soft tissue, and pancreas; cause-specific survival by expanded race and ethnic groups; SEER 13 delay-adjustment; and adjustments for Veterans' Administration (VA) underreporting.

Fast Stats. Fast Stats (*http://seer.cancer.gov/faststats/*) is an interactive data query Web system that provides easy, rapid access to key SEER and U.S. cancer statistics at various sites. Data may be selected by age group, sex, race, and year of diagnosis. Incidence is the default statistic displayed; survival, prevalence, and lifetime risk also may be available, depending on the data site.

Chapter 3

The Strategy: Where Do We Go From Here?

SRP leads cancer surveillance research science with strong programs in individual- and group-level data collection, statistical methods, and modeling tools that enhance the understanding of risk, prognosis, and population differences (**Figure 6**). SRP works with surveillance partners to set standards for data collection, use, and communication.

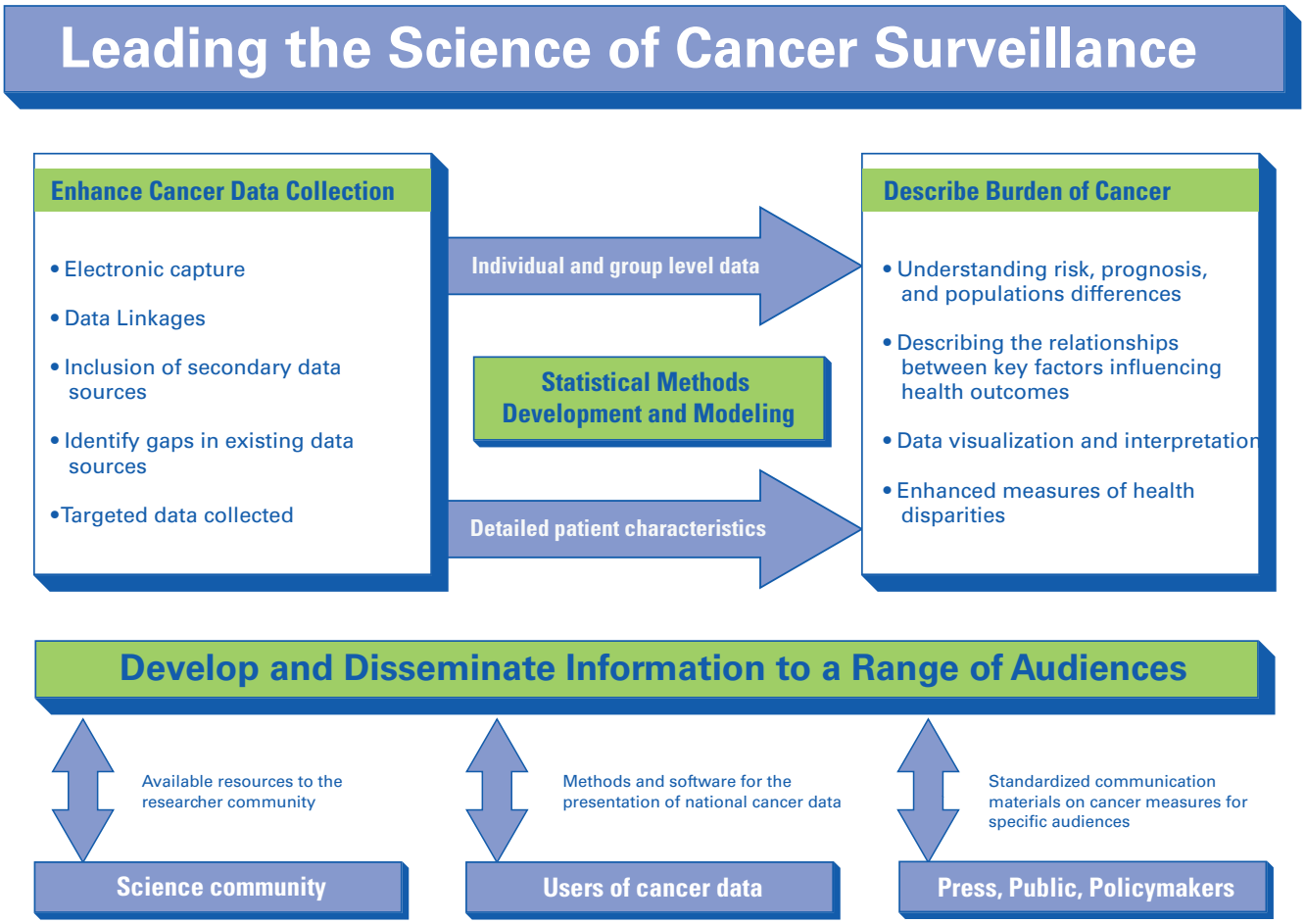
The October 2009 workshop participants identified major scientific and societal trends that affect the burden of cancer in the United States:

- Cancer survivors are living longer. This success brings a consequent surge in long-term health effects and cancer recurrences.
- There is a lessening of health disparities in some health areas, but constant or growing health disparities in other areas.
- The research community recognizes greater clinical and tumor heterogeneity within major cancer sites. The movement to develop therapies specific to tumor molecular and proteomic characteristics is anticipated to continue to expand.
- Broad societal factors at multiple levels of analysis increasingly are being shown to have substantial impacts on the burden of cancer in the United States.
- Consumers' and patient advocates' interest in cancer statistics is burgeoning. Scientists have a better understanding of how to disseminate information and how people understand numbers and statistics.
- The growth of electronic health records (EHRs) has enabled more efficient and more comprehensive collection of data about the burden of cancer. Interoperability among data systems is facilitating opportunities for data linkages to provide more comprehensive data on factors not currently obtained in cancer registration systems.

The priorities and action plans for cancer surveillance research for the next 3 to 7 years aim to enhance cancer data collection efforts to include detailed patient characteristics using electronic and other tools that help refine the understanding of population differences and health disparities, and communicate these cancer statistics to researchers and the public. This chapter provides a brief rationale of each of the priority areas and action plans to meet urgent needs of public health and biomedical science research.

The informatics infrastructure needed to expand data collection in a timely and efficient manner currently is not in place. Major advances in the development of EHRs and electronic capture of patient information are ongoing;

Figure 6. SRP's Approach to Cancer Surveillance



however, work remains to define how cancer surveillance systems will tie into the nationwide efforts to build up EHRs. As the electronic exchange of information and linkages of data from various sources expands, concerns regarding data privacy and confidentiality become more prominent and need to be addressed. Utilizing cancer surveillance data requires knowledge of informatics, demographics, statistical concepts, and disease characteristics. Increasing the number of researchers trained to use surveillance systems is critical in moving cancer surveillance forward.

Researchers who use SEER data and analytic tools developed for registry data require training across all areas of the cancer surveillance research spectrum, including gathering, compiling, analyzing, and disseminating data. For data to be used effectively and widely, NCI needs to make information more readily available and explain how to access and use it. A comprehensive approach to training for surveillance researchers is needed.

Table 3. Cancer Surveillance Priority Areas and Action Plan, 2010

Priority Areas and Action Plan 2010	
Priority Area 1	Provide a more complete depiction of the burden of cancer in the United States, past, present, and future.
Action Plan	1. Link with other data sources.
	2. Develop processes to obtain and use other types of variables and data.
	3. Collect more detailed individual data from a limited number of registries.
Priority Area 2	Collect better quality data on the burden of cancer in the United States more efficiently through effective utilization of electronic tools.
Action Plan	4. Identify and evaluate relevant data sources.
	5. Automate the capture of information.
	6. Expand interoperability and standardization of surveillance data.
	7. Use data-capture techniques and statistical modeling to improve the timeliness of cancer statistics reporting.
Priority Area 3	Improve understanding of the differences and disparities in the burden of cancer in the United States.
Action Plan	8. Enhance measurements of health disparities.
	9. Expand data visualization and interpretation.
Priority Area 4	Better understand the continuum in the burden of cancer in the United States from risk to prognosis.
Action Plan	10. Evaluate the potential impact of prevention, early detection, and treatment on cancer statistics to inform cancer research and policy development.
	11. Identify and track a cohort of cancer patients to obtain data and monitor impacts at the population level.
Priority Area 5	Communicate cancer statistics more effectively to researchers and users and make the data more accessible and understandable to all.
Action Plan	12. Conduct needs assessment to identify data needs and audiences and develop a plan to communicate cancer statistics and interpret data for audiences.
	13. Produce standardized communication materials.
	14. Disseminate information on existing resources to researchers and evaluate their effectiveness.
	15. Develop and disseminate methods and software for the analysis and presentation of National cancer statistics.

PRIORITY AREA 1

Provide a more complete depiction of the burden of cancer in the United States.

As epidemiologic and genetic research facilitates more personalized medical treatment, greater attention is being devoted to the progression of cancer in patients, from the onset of the disease through survivorship, as well as the identification of risk factors and biomarkers for cancer. Cancer surveillance must capture these trends and assess how they affect the burden of cancer by addressing data needs along the length of the cancer continuum, from screening and early detection through diagnosis, treatment, and outcome (survival, disease recurrence, or death).

Concerted efforts during the past decade to collect information on care patterns, quality of life, and morbidity in cancer patients include the annual SEER Patterns of Care and RRSS studies, the Health, Eating, Activity, and Lifestyle (HEAL) study, Cancer Care Outcomes Research and Surveillance Consortium (CanCORS), and the NLMS-SEER and SEER-Medicare datasets. Because of these activities, data currently collected by SEER cover incidence and survival information, with detailed characteristics of the cancer at diagnosis, some information on patient characteristics, and initial treatment data. No single initiative, however, provides a comprehensive look at a patient from diagnosis to death or long-term survival. With advances in technologies, broadening focus on the EHR, and the surge in biological and molecular information, a significant opportunity exists to expand data collection and management to include complete treatment data; recurrence, disease status, and trajectory data; comorbidities; and late effects. Calls have been made for the creation of a patient treatment summary and survivorship plan to support the move toward personalized medicine for cancer patients. Data on the existence of such summaries and plans should be captured by future surveillance systems.

Monitoring treatment on a population level allows for a basis for understanding the usage and impact of treatment on outcomes such as survival, mortality, and treatment-related comorbidities. It is an essential step in assessing if all cancer patients have access to the latest treatments and benefit from medical advances.

Data collection should be increased along the cancer continuum through coordination with other programs to identify common goals, and the provision of data from other sources to SEER to complete the information or expand the variables collected, both on an episodic and a routine basis. Statistical methodologies that can assist with data expansion efforts are described in **Box 3.1**. Additional opportunities to broaden data collection include identifying gaps in population coverage, developing approaches to collect information from individuals not covered by existing sources, and automating linkages between SEER and other data sources. NCI should supply the infrastructure and expertise to develop cancer survivor cohorts that can provide missing data for specific geographic areas, and also should support research to better understand regional differences, which will help generalize information to the U.S. population.

Current SRP surveillance systems rely on population data from the census and other sources, cancer registry data, and mortality data systems to estimate risk of cancer, prevalence of cancer, and survival. A systematic approach is used to obtain statistics about recurrences; use of long-term cancer-directed therapies; and physical, social, and emotional sequelae. It depends on systematic registry efforts to capture and relate these events to population data, and thence to enable the computation of rates; in some instances, observational cohorts can provide this information. Extensive research is needed to develop standardized definitions for recurrences that are sufficiently practical to be applied in large populations that can be generalized to U.S. cancer patients. In addition, it may be useful to establish cohorts that provide this information. For example, information obtained via a population-based sample from a cohort of patients from the registry system could be

Box 3.1. Statistical Methods: Promoting Data Collection

Statistical methodologies and applications can be used to facilitate the collection of data from a wider range of sources. Research opportunities include:

- **Development of imputation methods for missing data.** As we expand the information collected it will become increasingly important to account for missing information and adjust for related biases in population estimates.
- **Development of competing risk survival methods that incorporate comorbidities.** Survival models that predict outcomes for cancer patients will need to include information on the whole individual not just the cancer diagnosed. Competing risk models that adjust for existing health condition when estimating prognosis will customize results for individuals.
- **Data mining techniques to identify adverse outcomes associated with treatments.** Identifying rare outcomes associated with increased risk or cancer treatments through population data will require the advanced data mining techniques and procedures to distinguish false alarms from signals that would warrant further investigation.

used to assess or compare morbidity or survivorship for a wide range of subpopulations.

Three specific research opportunities to advance data collection along the patient continuum are:

- Link with other data sources.
- Develop processes to obtain and use other types of variables and data.
- Collect more detailed individual data within a limited number of registries.

Action Plan 1: Link with other data sources.

The cancer surveillance system should be improved to better handle data on the growing number of cancer survivors and in preparation for the expected increase in the aging population in the United States. A key ingredient for success in this effort is to enhance the system without incurring additional costs. Newer technologies and more data linkages may allow expansion of data with cost savings. Important operational processes to consider in developing such a system include utility of data linkages to allow collection of additional data needed in a cost-effective manner; EHRs and structured, standardized formats for collecting clinical data; and collaborations across governments and with clinical data networks.

Specific research opportunities in this area include:

Coordinate with existing collaborators and initiatives or other data sources to gather missing data; expand the variables collected; or promote a standardized, cancer-focused EHR that incorporates data into the SEER database routinely, particularly through automated collection processes and linkages. Such organizations and initiatives could include: CMS' Medicare data, larger insurance systems (HMO Cancer Research Network [CRN]), National Comprehensive Cancer Network (NCCN), the American Society of Clinical Oncology (ASCO), NCI's caHUB™, AJCC's National Cancer Data Base (NCDB), and pharmacosurveillance groups.

- Explore the relationship between reporting sources and other sources of information and the contents of the registries' summary records to maximize the benefits achieved from collected data. The existing two-way data flow involves, for example, making data from RRSS available through the SEER system, and bringing data from hospital special studies into the SEER system from local registries into the central registry. Such efforts require that legal requirements in this area be identified and satisfied.
- Develop a plan to address Institutional Review Boards' (IRB) concerns about sharing data and samples.
- Establish a "biospecimen library" to house blood or tissue samples to help address anticipated changes driven by continued discoveries in basic and clinical research and their eventual incorporation into care delivery. Analyze the types of data that are collected and determine how registries might serve as a bridge between discovery research and clinical practice. RTR's experiences with maintaining biospecimens from several SEER registries could assist in the development of a library.

Action Plan 2: Develop processes to obtain and use other types of variables and data.

SRP surveillance systems rely on population data from a variety of primary and secondary sources, including the census, cancer registry data, and mortality data systems. Each of these sources has a unique informatics infrastructure and modes of operating and collaborating. A systematic approach to obtaining and using other types of variables and data would facilitate the capture of cancer data and expedite the computation of population-based cancer statistics.

Potential models for data access include SEER-Medicare and the California Community Health System. These systems have created public files while providing security layers and additional precautions regarding degrees of access to individual private data; they use a firewall to protect sensitive information. The Tools for Electronic Data initiative and the Cancer Research Network (CRN)-SEER data transfer pilot project also serve as data expansion models for developing processes to obtain and use other information.

Data needs include: socioeconomic status and education; height/weight; address, including past addresses; race/ethnicity (better quality data); family history; tumor subtypes based on molecular signature; environmental risk factors (e.g., satellite imagery); treatment and interventions; and tobacco use, comorbidities, insurance information, and mode of detection. Other variables include geographic/built environment, census tract, education, BMI, family history, and environmental or other risk factors. Genetic/biomarkers for prognosis are a statistical priority.

Specific research opportunities in this area include:

- Develop processes to capture more individual data through both primary data collection and increasing linkages with secondary sources. Processes should be sensitive to issues of privacy, statutory regulations, and patient consent, and should incorporate flexibility and adaptability to accommodate changes in scientific advances and data collection as methods of quality care or assessments of risk change in the future.
- Examine methodological issues to determine the best way to: support the expanded collection of patients' treatment history, disease status, and followup; prepare for the collection, access, and use of genomic data (germline and somatic DNA); and facilitate collection of and access to multilevel data from multiple sources. Missing denominator data also should be imputed from residual tissue repositories.
- Explore linkages with the SEER dataset in understanding health disparities, social determinants of health, and genetic and geographic population subgroups, in tandem with Priority Area 3. Develop pilot studies to determine the efficacy of using SEER data at the census level and to investigate additional linkages at this level. Expand SEER data to provide a resource for researchers through custom population data requests and targeted studies.

Action Plan 3: Collect more detailed individual data from a limited number of registries.

The factors influencing the burden of cancer are extremely complex, and more detailed individual data could strengthen the understanding and prognosis of cancer. Despite the

increasing availability of EHRs and genomic data, there are serious obstacles to collecting all of the information needed. For example, it is frequently difficult to capture data when patients suffer from cancer recurrences.

Specific research opportunities in this area include:

- Develop a surveillance system to capture systematically the extent of: use of long-term cancer-directed therapies and recurrences; as well as to enhance the physical, social, and emotional sequelae available through SEER-MHOS. An intensive, rigorous data-collection process could be used to extrapolate regionally specific data to a broader application across U.S. populations. An alternative approach is the "sentinel registry" model, in which a few cancer registries concentrate on collecting more detailed individual data and specimens on a limited basis from a specific area(s). A "sentinel registry" approach is appropriate for times when widespread adoption is not feasible because of the costs associated with collecting additional variables. The model could encompass one disease site covered by one registry for a 2- to 3-year followup as a demonstration, either population-based or sampling.
- Consider cost-effective alternatives to the "sentinel registry" model, such as establishing partnerships with other data-collection efforts that collect samples (e.g., or linking with large clinical or observational trials (e.g., Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial [PLCO], which has collected specimens from individuals for the past 15 years).

PRIORITY AREA 2

Collect better quality data on the burden of cancer in the United States more efficiently through effective utilization of electronic tools.

Integrated health informatics is a central component in the contemporary health care system setting, particularly in cancer surveillance.

NCI has developed and refined several premier tools to capture electronic data, including: SEER*DMS and caBIG®, as well as SEER Abstracting Software and NAACCR Interoperability Working Groups. SEER*DMS has played a significant role in transforming population-based research by helping registries access, process, and submit high-quality data to NCI. E-Path screens pathology reports for cancer-related reports in 239 facilities at 13 SEER central cancer registries, and ongoing refinements include auto-coding and use of synoptic guideline formats. caBIG® provides a conduit between institutional infrastructures to allow information sharing while respecting both patient and organizational confidentiality. Cancer surveillance informatics also involves linkages between datasets, such as SEER-Medicare and SEER-NLMS, as well as the development of software and dissemination of results.

Advances in technological and computing fields, the development of more sophisticated electronic platforms and tools, and the concerted movement toward widespread adoption of EHRs present a prime opportunity to develop a strategy to use electronic tools to expand the collection of surveillance data. As new data sources are found and mined, however, patient privacy and confidentiality must be safeguarded as tools are used to access electronic data and auto-populate SEER data fields. NCI will continue to participate actively in the national Health Information Technology effort and will help to develop standards on electronic reporting. Through an RFP on Tools for Electronic Data, NCI will expand its capabilities to collect population-based cancer surveillance data from electronic pathology reports beyond the current SEER dataset. Statistical methods that can help to capture electronic data are suggested in **Box 3.2**. Additional challenges include obtaining future research consent from patients, developing new models for e-data acquisition, creating real-time data feeds in the data-capture process, and reducing the timeframe for generating and reporting cancer incidence rates.

Box 3.2. Statistical Methods: Capturing Electronic Data

Statistical methodologies and applications can be used to facilitate capture of electronic data. Research opportunities include:

- Development of matching algorithms and learning algorithms to efficiently merge datasets.
- Use of Natural Language Processing (NLP) to help obtain defined data elements from free text in medical records.
- Development of approaches for linkage to other databases.

Specific research opportunities to help capture existing data and add to the portfolio of surveillance information include:

- Identify and evaluate relevant data sources.
- Automate the capture of information.
- Expand the interoperability and standardization of surveillance data.
- Use data-capture techniques and statistical modeling to improve the timeliness of reporting cancer statistics.

Action Plan 4: Identify and evaluate relevant data sources.

SEER's surveillance data currently covers a range of geographic areas and select populations, from which extrapolations are made and applied across a broader population. Collaborators provide further data, but data sources beyond SEER and its collaborators are needed. These sources should provide information that both expands the current boundaries of the SEER dataset and supports the emphasis on population-based cancer surveillance.

Specific research opportunities in this area include:

- Evaluate ongoing collaborations for their potential to obtain further data, including further synergy between SRP and CMS to collect more treatment data using the SEER-Medicare dataset.
- Develop new algorithms or tools to feed treatment data back to the registries from SEER-Medicare.
- Address treatment definitions and coding rules that limit SEER's ability to link to treatment data from other sources. HL7 extraction processes may be helpful in providing links to other clinical data sources.
- Develop an approach that leverages the caBIG® network to link cancer surveillance data and data from other sources.

Action Plan 5: Automate the capture of information.

An informatics approach is needed on a broad scale to automate the capture of information. SEER*DMS currently provides a good model for the automation of processes that make data submission to NCI as seamless as possible. The widespread adoption of EHRs slated for 2015, however, is a prime opportunity to expand automation processes to capture more cancer-related information. A key element in this automation is NLP, a nexus of computer science and linguistics focused on the interactions between computers and human languages. The NLP field is rapidly improving the capacity to identify and extract discrete data items from a wide range of free text documents that may contain grammatical and other errors.

Specific research opportunities in this area include:

- Explore NLP's capability to extract information from reports produced by hospitals and physicians' offices.
- Determine NLP's capability to use auto-coding and transformation of data to synoptic formats (Synoptix) to increase recognition sensitivity and accuracy.

Action Plan 6: Expand the interoperability and standardization of surveillance data.

Cancer surveillance organizations and their partners often struggle with sharing data because of the myriad of software and hardware systems used. In addition, the data elements collected often vary widely. As the quantity of data burgeons and caBIG® provides a framework within which researchers can share data, the capability to exchange information within SEER as well as with hospitals and other data providers likewise should be expanded. Mutually beneficial collaborations should involve a structure for data collected at the point of care. A key element in the success of these interactions will be collegiality among organizations and widespread adoption of bioinformatic and surveillance data standards.

Specific research opportunities in this area include:

- Establish standards for the collection and coding of health data to ensure the consistent quality and interoperability of data from a variety of sources and to facilitate the integration of information for use by registries, researchers, and other entities. Cancer surveillance organizations and their partners should work toward agreement on standards.

Action Plan 7: Use data-capture techniques and statistical modeling to improve the timeliness of reporting cancer statistics.

Cancer statistics currently are reported several years following the year of diagnosis. With the public accustomed to receiving information quickly, the current 3-year time lag presents challenges in meeting this expectation. Decreasing the time between diagnosis, notification to the registry about a case, and reporting from the registry would assist in computation, help ensure more accurate estimation, and improve the timeliness in reporting cancer statistics.

At the present time, Medicare is an important source of treatment data, but EHRs eventually may provide more current and complete treatment data for cases of all ages. Efforts should begin now to understand and refine approaches to capture and analyze this information.

Specific research opportunities in this area include:

- Develop data-capture techniques, including E-Path, to expedite the reporting process.
- Design a delay adjustment model to collect incidence data in a short timeframe. Cancer surveillance programs should determine: (1) What is needed for all cancer registries to have the capability of capturing a high percentage of incident case information electronically? (2) How can technology similar to E-Path be extended to all diagnostic resources? (3) What data components are precoded, and what components need to be transformed? A system should be designed to alert registrars when more information becomes available in a case's chart, or to extract new information automatically.
- Accelerate reporting of data to CMS and improve access to more current Medicare treatment data.

PRIORITY AREA 3

Improve understanding of the differences and disparities in the burden of cancer in the United States.

A major goal of the public health system is to improve understanding of the difference and disparities in the U.S. cancer burden. Broad societal factors at multiple levels of analysis increasingly are being shown to have substantial impacts on the burden of cancer in the United States. Measuring and monitoring the effectiveness of treatments across populations and subpopulations provides indications of where planning and programmatic shifts may be helpful in reducing the cancer burden. As significant changes occur in the use of tumor markers, a strategy should be developed to improve the characterizing of subgroups of cancer patients, particularly in terms of treatment, recurrence, and comorbidities. Statistical methods and tools also could help to improve understanding of the differences and disparities in the U.S. cancer burden (**Box 3.3**).

Box 3.3. Statistical Methods: Improving the Understanding of the Differences and Disparities in the U.S. Cancer Burden

Statistical methodologies and applications can be applied to further the understanding of differences and disparities in cancer patients. Research opportunities include:

- **Multilevel modeling**—Modeling to capture the complex interrelated factors contributing to health disparities.

Specific research opportunities to help better understand differences and disparities in the burden of cancer in the United States include:

- Enhance measurements of health disparities.
- Expand data visualization and interpretation.

Action Plan 8: Enhance measurements of health disparities.

SRP plays a leading role in the development of health disparities measures that promote the cancer-related objectives of *Healthy People 2020*, including the elimination of health disparities across multiple categories. There is a lessening of health disparities in some health areas, but constant or increasing health disparities in other areas. Methods to measure cancer disparities should continue to be developed, with the recognition that (1) a single measure generally will not provide adequate information to inform policy decisions, (2) the choice of measures (e.g., absolute versus relative) makes a substantive difference in the results and interpretation of data, and (3) the size of population groups and magnitude of disparities should be considered when determining the impact on the public health system. SRP has encouraged the use of measures through HD*Calc (*http://seer.cancer.gov/hdcalc/*) and should continue to lead in this area of research by developing geographic and area-socioeconomic status measures that capture disparities, and by promoting HD*Calc as a resource for researchers.

Specific research opportunities in this area include:

- Consider ways to link SEER registry data to geographic variables such as county of residence and census tract of residence at the time of diagnosis.
- Evaluate cancer cases linked to area-socioeconomic status/census tract variables, with the ultimate goal of making this data available to the research community.
- Evaluate linkages between SEER data and other available data through geographic variables such as census tract of primary residence to expand the set of measures that may describe disparities.

Action Plan 9: Expand data visualization and interpretation.

Geographic methods, such as data visualization and interpretation, provide a unique perspective on the identification and quantification of cancer risk and prognosis. The value of geography goes beyond cancer etiology and can influence prevention, detection, diagnosis, treatment, and survivorship. Geography often serves as a surrogate for other factors, including environmental risk, behavioral risk, demographics, race/ethnicity and ancestry, and socioeconomic factors. However, more complete spatial and spatial-temporal data are needed about geographic subgroups of cancer patients.

Although access to spatial data is improving through clearinghouses, Web-based access, and improved technologies, challenges remain concerning the availability of data, pace

of change, and spatial data linkage through geocoding. Geocoding techniques are improving in terms of match rates, accuracy, and measures of uncertainty, but spatial-temporal analysis is time consuming and involves daily travel and access to care as well as residential history and lifetime exposure.

Specific research opportunities in this area include:

- Develop new geocoding techniques to impute missing data.
- Develop methods with communications researchers to display the communication of uncertainty in spatial analysis and visualization of data.

PRIORITY AREA 4

Better understand the continuum in the burden of cancer in the United States from risk to prognosis.

Cancer risk and prognosis address both sides of a cancer diagnosis: prevention and detection (before diagnosis) and treatment and survivorship (after diagnosis). Surveillance research within SRP currently approaches risk and prognosis in populations by understanding health disparities and social determinants of health, as well as genetic and geographic population subgroups. Researchers and clinicians are studying and using tumor markers along the disease continuum, including in terms of treatment, recurrence, and comorbidities. The cancer surveillance research community can help with these efforts by devising better ways to characterize subgroups of cancer patients.

During the past decade, NCI has built a multi-faceted portfolio to understand population differences that includes the RTR, monographs on health disparities, HD*Calc, CISNET, and grants on data display/presentation.

Challenges remain, however, in identifying and obtaining relevant denominator data for race/ethnicity, socioeconomic data and social networking, the pace of change in genomic discoveries about genetic subgroups, and spatial and temporal data about cancer patients. To advance population sciences research about risk and prognosis, pilot studies should be developed to determine the usefulness of analyzing SEER data at the census level, promote the use of HD*Calc to researchers, help CISNET expand cancer sites, and interact with other research and policy groups. Other opportunities to apply statistical methods to better understand risk, prognosis, and population differences are provided in **Box 3.4**. Alternative formats also should be developed to display annual cancer statistics. Additional opportunities include multilevel modeling to link population and molecular models and use of expanded data collection and cancer cohorts to gain a more comprehensive understanding of survival in populations and subpopulations.

Specific research opportunities to advance understanding of risk, prognosis, and populations differences are:

- Evaluate the potential impact of prevention, early detection, and treatment on cancer statistics to inform cancer research and policy development. Develop a plan to better understand characterizations of cancer populations and subpopulations, in unison with efforts described in Priority Area 3 above.
- Identify and monitor a cohort of cancer patients by sampling a cohort from the registry system, in which patients represent the entire population and provide the ability to examine health disparities across many levels.

Action Plan 10: ***Evaluate the potential impact of prevention, early detection, and treatment on cancer statistics to inform cancer research and policy development.***

The ultimate goal of prevention, early detection, and treatment is to reduce the rate of cancer incidence and mortality. As new information, technologies, and treatment advances

Box 3.4. Statistical Methods: Advancing the Understanding of Risk, Prognosis, and Population Differences

Statistical methodologies and applications can be applied to further the understanding of risk, prognosis, and population differences. Research opportunities include:

- **Understanding risk of disclosure to ensure patient confidentiality.** Methods need to be put in place to ensure patient privacy when combining data sources.
- **Small-area estimation methods and measures of uncertainty.** Interest in producing estimates for small geographic areas or small subpopulations continues to grow, leading to results with large levels of uncertainty. Understanding the limitations and communicating the uncertainty are essential to avoiding the misinterpretation of results.
- **Classification methods to identify cases with similar prognosis.** As biological markers for risk and prognosis continue to be discovered, grouping patients with similar prognoses will become more complex. This is the first step to estimating survival and other outcomes that are patient relevant.

are disseminated, the surveillance program should assess the effect on the cancer burden in the population. Through the understanding of risk factor trends, use of early detection and treatment, and information on risks and efficacy from clinical trials, models can predict the impact on cancer incidence and mortality trends. Modeling also can provide estimates of potential impact if interventions were more widely implemented in the population. This type of analysis provides both the feedback needed to evaluate the success of particular advances in early detection and treatment and the input needed by policymakers and health planners when deciding where to focus their efforts. Surveillance can evaluate differences between results observed in a tightly controlled clinical setting and what occurs in the population setting. CISNET has provided an example of the contribution that population modeling can make in assessing the benefits of early detection and treatment advances, comparing the benefits associated with different technologies and treatments, and informing policy decisions.

Specific research opportunities in this area include:

- Expand surveillance modeling to cover a wider range of applications and cancer sites.
- Incorporate genomic and family history risk profiles.
- Upstream modeling—determinants of trends in risk factors, screening, and treatment.
- Issues in comparative effectiveness research (CER).
- Optimizing biomarker development strategies.
- Multiscale modeling, ranging from the molecular to population level.
- Evaluation of diagnostic tests.
- Optimal routes to reducing health disparities.
- Translation of trial results into clinical guidelines and public health policy.
- Develop interactive policy-level decision-making tools.

Action Plan 11: Identify and track a cohort of cancer patients to obtain data and monitor impacts at the population level.

As cancer patients enjoy increased longevity, health providers and biomedical researchers remain concerned about cancer recurrence and linking the effects of comorbidities on recurrence. The definition of recurrence is not well established or standardized for most cancers, and it relies on a battery of tests that are not applied consistently to all patients. In addition to recurrence, concerns exist about

patient-reported outcomes and biospecimens; data about these currently are not captured by SEER datasets, although other NCI program efforts to link patient-reported outcomes to SEER data, such as SEER-MHOS and SEER-CAHPS, are beginning to provide national surveillance data on patient-reported outcomes. Other areas in which data are limited include: complete treatment data, trajectory data, clinically relevant and evolving data, comorbidities, late effects, patient-reported outcomes, multilevel analyses, health behaviors, and patient treatment summary and survivorship plans.

Newer cancer-directed therapies may be having a substantial impact on the cancer burden in the United States. In addition, therapies can be prolonged, even over many years, with long-term or indefinite physical, social, and emotional consequences. The effect of these newer cancer-directed therapies and of prolonged treatment regimens has not been established at the population level. Knowledge of this impact may help guide medical practice and health policies.

A holistic approach to population studies involving investigators collaborating across registries would build a base of knowledge more effectively than a selective, single-aspect focus brought by an investigator working within one cancer registry. The registry system also would provide nationally representative samples to supplement data from more localized groups, allowing the opportunity to examine socioeconomic factors and other health disparities variables.

Specific research opportunities in this area include:

- Identify and follow a cohort of cancer survivors who are identified through the registry system to supplement data from more localized groups and examine health disparities factors, such as socioeconomic status. Develop guidance for groups that wish to develop such a cohort, including data consolidation and data collection aimed at filling gaps for geographically defined areas. Design a plan to provide access to the data by investigators who are not participating in the cohort.

PRIORITY AREA 5

Communicate cancer statistics more effectively to researchers and users and make the data more accessible and understandable to all.

The availability to use cancer statistics by a broad range of audiences has expanded significantly during the past decade. Consumers and patient advocates are increasingly interested in cancer statistics; scientists have a better understanding of how to disseminate information and how people understand numbers and statistics. Each type of audience—from researchers, registry staff, tumor boards, and NCI staff, to journalists, patients and patient advocates, policymakers, and the public—seeks data for different purposes, and each time data are used provides an opportunity for data misinterpretation.

NCI has built a portfolio to address communication needs, including an ongoing study of SEER branding, CSR, cancer fact sheets, the *Annual Report to the Nation*, state cancer profiles, P.L.A.N.E.T., short courses on surveillance survival methods, and web sites.

The need for a more comprehensive, unified approach to communications about cancer statistics data to these audiences is growing. A needs assessment to better help audiences, the promotion and revision of existing software and products, and a cancer survival query system all would help fine-tune NCI's communications. Statistical methods and tools also could help improve the communication of cancer data (**Box 3.5**). A National Cancer Index could provide a centralized resource for cancer incidence information. Other opportunities include the introduction of alternative formats to display cancer statistics, assistance to registries in their communications with NCI and between registries, expanded communication regarding surveillance statistical methods, and periodic evaluations of communications materials.

Research opportunities to communicate cancer statistics and resources available to the research community include:

- Conduct a needs assessment to identify data needs and audiences and develop a plan to communicate cancer statistics and interpret data for audiences.
- Produce standardized communication materials that are informed by research advances in the communication and interpretation of health data.
- Disseminate information on existing resources to researchers and evaluate their effectiveness.
- Develop and disseminate methods and software for the analysis and presentation of national cancer statistics.

Action Plan 12: Conduct a needs assessment to identify data needs and audiences and develop a plan to communicate cancer statistics and interpret data for audiences.

As the broad swath of data about cancer incidence, mortality, morbidity, and survival continues to be gathered and used by many groups, there is a need to communicate cancer statistics and help researchers and other audiences interpret and report data with accuracy. Information about cancer sought by scientists, however, differs greatly from that requested by lay audiences. Different communications media are needed to reach a plethora of audiences, such as researchers, registry staff, tumor boards, journalists, patients and their families, patient advocates, policymakers, journalists, and the public. A plan to communicate cancer statistics should be developed that distinguishes approaches for these audiences.

Specific research opportunities in this area include:

- Conduct a needs assessment to determine the type of information sought by investigators. Identify those questions currently asked of national and local SEER staff and examine how SEER data are used. Conduct a bibliometric analysis of publications that use or otherwise reference SEER data to obtain additional information. Content that should be examined includes statements of data limitations, materials targeting local populations, and cross-links. Review the Health Information National Trends Survey (HINTS) and other needs assessments to extract useful approaches and suggestions in the process of conducting the assessment. Also identify existing products and determine whether they should be expanded or new ones developed.
- Delineate ways to promote existing products and resources, such as Fast*Stats and SEER*Stat, to the scientific community (e.g., expanded training and dissemination, short courses, conferences, and Webinars). Develop recommendations regarding how to translate data for journalists and the public, including information about data limitations, as well as how locally relevant cancer data could be provided to researchers. Accentuate the need to evaluate communications materials periodically, including the effectiveness of formats used.
- Survey users of State Cancer Profiles and other data on Cancer Control P.L.A.N.E.T. to help refine the resource to promote more efficient use of the data by researchers preparing grant applications and those developing presentations to Congress.
- Consider the public's need for cancer surveillance data and whether to incorporate the lay audience in the assessment and develop a separate plan to meet this need. Identify preferred formats for communicating data and perform usability testing for each type of audience—including via the Web, print, and Webinars; specific formats that should be tested include accessible language, various formats that facilitate rapid access to desired information, and segmented Web pages. Developing templates for presenting cancer surveillance data to different audiences would expedite the process. Revise standardized fact sheets to be more descriptive, present risk in the context of other cancers and other major diseases, and show risks by age or other approaches that have been found to enhance comprehension of such data.

Action Plan 13: Produce standardized communication materials.

A number of national efforts currently disseminate cancer surveillance data to interested parties. Collaboration among leading organizations, including NCI, CDC, and ACS, has facilitated and guided some of this communication.

Box 3.5. Statistical Methods: Enhancing the Communication of Cancer Data to the Research Community

Statistical methodologies and applications can be used to facilitate the communication of data to the cancer research community. Research opportunities include:

- **Development of absolute measures of risk (in the presence of competing risks)**. Cancer survival typically is reported in the absence of competing risks. This measure currently provides the best progress measure in the fight against cancer because it is not influenced by changes in other diseases. Better measures are needed, however, to help the individual patient who considers their overall heath when making decisions.

However, much of the information available about surveillance data is fragmented or in some cases inconsistent in presentation. Miscommunication and misinterpretation of prognosis information is a particular problem. The adoption of standardized messages and templates for communicating cancer surveillance data will present a cogent, coherent voice about cancer trends to the health community at large, dissipate confusion, and help reduce data misinterpretation.

The development of standardized templates that present data to different audiences could help advance this effort significantly. In particular, fact sheets that contain the same type of information for existing data (e.g., caveats/descriptors, or good news/bad news) could be prepared with existing resources in a relatively short timeframe.

Specific research opportunities in this area include:

- In a collaborative setting that includes NCI, CDC, ACS, and their registries, define data communication and develop strategies and standards for data communication by reaching consensus about common data elements, while remaining sensitive to issues related to a potential overload of cancer surveillance information.
- Consider ways to communicate data limitations to researchers to avoid misinterpretation, such as the inclusion of caveats that acknowledge data limitations.
- Identify possible cross-links between the SEER Web site and other credible Web resources (e.g., risk calculators, screening recommendations), particularly to help users answer questions that SEER data cannot answer. Review other federal Web sites (e.g., NIH, Department of Health and Human Services [HHS]). Use a structured approach to weigh the benefits and challenges of linking SEER to nonfederal Web sites, recognizing the lack of consensus on standards for health information Web sites, as well as the difficulty in obtaining government approval for such linkage.
- Develop standardized, accessible communication materials for lay audiences, as directed by the needs assessment and communication plan described in Research Opportunity 12.

Action Plan 14: Disseminate information on existing resources to researchers and evaluate their effectiveness.

The SEER cancer registry system serves as a platform for research studies in the field of cancer. A wide variety of resources are available currently, and more are being developed. As available resources are developed, SRP needs a structured approach to promote available resources and a system to make these resources available to the wider research community to maximize their potential. SEER-Medicare has been used widely by the research community and provides a model of how research resources can be used to address a range of research questions. The RTR and the linked SEER-NLMS databases are examples of resources that have potential to be more widely disseminated to the larger research community. SRP should develop a

structured approach to communicating available resources and facilitating the acquisition and use of these resources.

Specific research opportunities in this area include:

- Determine the needs of local registries in disseminating local-level cancer surveillance data. Define the steps required to access cancer surveillance data that is not publically available and facilitate the process.
- Conduct a systematic, heuristic review of existing resources and how they are promoted and accessed to better understand the needs of the extramural research community and determine the strengths, weaknesses, and user satisfaction with existing resources available in SRP.
- Investigate methods of making data from SEER and other registries more readily available to researchers for a variety of purposes

Action Plan 15: ***Develop and disseminate methods and software for the analysis and presentation of national cancer statistics.***

NCI's surveillance program has developed statistical methods and software for the analysis of surveillance data in the past and continues to develop and refine analysis tools. These tools allow researchers with diverse backgrounds to easily access surveillance data and use the data for particular purposes. Although SEER*STAT and related software programs are widely used, NCI should further promote the use of available analysis methods and tools and provide training on the use of available software. Dissemination and targeted training are expected to help invigorate the field and inform research conducted on the cancer control continuum to enhance the impact on cancer mortality in the United States.

Specific research opportunities in this area include:

- Determine better ways to communicate new and existing surveillance statistical methods. Communications should be expanded through through the use of monographs, special journal issues, a textbook on surveillance methods, or other materials as the quantity of cancer data increases.
- Expand training efforts on the use of statistical methods and software to facilitate wider use of the tools developed at NCI.

Appendices

Appendix A:

Acknowledgements

This Strategic Plan would not have been possible without the collaboration and efforts of many individuals and organizations. Each made significant contributions throughout the process of preparing this document. We thank Brenda K. Edwards and Stacey Vandor for their invaluable support and comments throughout the process of developing this Plan.

2009 Cancer Surveillance Research Workshop Planning Committee

Kathleen A. Cronin, Chair
Rachel M. Ballard-Barbash
Martin Brown
Brenda K. Edwards
Eric (Rocky) J. Feuer
Sally Glaser
Bradford W. Hesse
Muin Khoury
Marsha Reichman
Julia Rowland
David G. Stinchcomb
Stacey Vandor
Deborah M. Winn

Contract Support

The Scientific Consulting Group, Inc.
Marcia Feinleib, Project Director
Darlene Summers, Principal Writer
Charles Wallace, Graphic Designer

Appendix B:

NCI Resources and Tools for Cancer Surveillance Research

Annual Report to the Nation
http://seer.cancer.gov/report_to_nation/

Cancer Statistics Review
http://seer.cancer.gov/csr/1975_2006/index.html

Cancer Trends Progress Report
http://progressreport.cancer.gov/

Cancer Intervention and Surveillance Modeling Network (CISNET)
http://cisnet.cancer.gov/

DevCan
http://srab.cancer.gov/devcan/

Electronic Pathology (E-Path) software
http://www.aim.on.ca/products/ePath.jsp

Fast Stats
http://seer.cancer.gov/faststats/

Health Disparities Calculator (HD*Calc)
http://seer.cancer.gov/hdcalc/index.html

Joinpoint
http://srab.cancer.gov/joinpoint/

Plan, Link, Act, Network with Evidence-based Tools (P.L.A.N.E.T.)
http://cancercontrolplanet.cancer.gov/

Rapid Response Surveillance Studies (RRSS)
http://seer.cancer.gov/rapidresponse

Residual Tissue Repository (RTR) Program
http://seer.cancer.gov/biospecimen

SEER
http://seer.cancer.gov/

SEER Data Management System (SEER*DMS)
http://seer.cancer.gov/seerdms/

SEER-Medicare data
http://healthservices.cancer.gov/seermedicare

SEER-Medical Health Outcomes Survey (SEER-MHOS)
http://outcomes.cancer.gov/surveys/seer3-mhos)

SEER-National Longitudinal Mortality Study (SEER-NLMS) database
http://surveillance.cancer.gov/disparities/nlms/

SEER*Prep
http://seer.cancer.gov/seerprep/index.html

SEER*Stat
http://seer.cancer.gov/seerstat/

StatFund
http://statfund.cancer.gov

State Cancer Profiles
http://statecancerprofiles.cancer.gov/

Appendix C:

List of Acronyms

ACoS	American College of Surgeons
ACS	American Cancer Society
AJCC	American Joint Committee on Cancer
ARP	Applied Research Program
ASCO	American Society of Clinical Oncology
caBIG ®	cancer Biomedical Informatics Grid®
CAHPS	Consumer Assessment in Healthcare Providers and Systems
caHUB™	Cancer Human Biobank
CanCORS	Cancer Care Outcomes Research and Surveillance Consortium
CDC	Centers for Disease Control and Prevention
CISNET	Cancer Intervention and Surveillance Modeling Network
CMS	Centers for Medicare and Medicaid Services
COC	Commission on Cancer (ACoS)
CRN	Cancer Research Network
CSR	Cancer Statistics Review
CSRP	Cancer Surveillance Research Program
DCCPS	Division of Cancer Control and Population Sciences
EHR	electronic health record
E-Path	Electronic Pathology software
GIS	geographic information systems
HD*Calc	health disparities calculator
HEAL	Health, Eating, Activity, and Lifestyle Study
HHS	Department of Health and Human Services
HINTS	Health Information National Trends Survey
IRB	Institutional Review Board
IACR	International Association of Cancer Registries
MHOS	Medical Health Outcomes Survey
NAACCR	North American Association of Central Cancer Registries, Inc.
NCCCS	National Coordinating Council for Cancer Surveillance
NCCN	National Comprehensive Cancer Network
NCDB	National Cancer Data Base
NCHS	National Center for Health Statistics (CDC)
NCI	National Cancer Institute
NCRA	National Cancer Registrars Association
NHANES	National Health and Nutrition Examination Survey
NLMS	National Longitudinal Mortality Study
NLP	natural language processing
NPCR	National Program of Cancer Registries (CDC)
P.L.A.N.E.T.	Plan, Link, Act, Network with Evidence-based Tools
PLCO	Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial
RRSS	Rapid Response Surveillance Studies
RTR	Residual Tissue Repository Program
SEER	Surveillance, Epidemiology, and End Results Program
SEER*DMS	SEER Data Management System
SIG	Surveillance Implementation Group
SRP	Surveillance Research Program
USCS	U.S. Cancer Statistics

DEPARTMENT OF HEALTH & HUMAN SERVICES · USA
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

March 2011