The Surveillance, Epidemiology and End Results (SEER) Program and Pathology: Towards Strengthening the Critical Relationship

Máire A. Duggan, MD, FRCPC[1], **William F. Anderson, MD, MPH**[2], **Sean Altekruse, DVM, PhD**[3], **Lynne Penberthy, MD, MPH**[3], and **Mark E. Sherman, MD**[4]

[1]Department of Pathology and Laboratory Medicine, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

[2]Divisions of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA

[3]Cancer Control and Populations Sciences, National Cancer Institute, Bethesda, Maryland, USA

[4]Cancer Prevention, National Cancer Institute, Bethesda, Maryland, USA

Abstract

The Surveillance, Epidemiology and End Results (SEER) program of the National Cancer Institute collects data on cancer diagnoses, treatment and survival for approximately 30% of the United States (U.S.) population. To reflect advances in research and oncology practice, approaches to cancer control are evolving from simply enumerating the development of cancers by organ sites in populations to include monitoring of cancer occurrence by histopathologic and molecular subtype, as defined by driver mutations and other alterations. SEER is an important population-based resource for understanding the implications of pathology diagnoses across demographic groups, geographic regions, and time, and provides unique insights into the practice of oncology in the U.S that are not attainable from other sources. It provides incidence, survival and mortality data for histopathologic cancer subtypes, and data by molecular subtyping is expanding. The program is developing systems to capture additional biomarker data, results from special populations, and expand bio-specimen banking to enable cutting edge cancer research that can improve oncology practice. Pathology has always been central and critical to the effectiveness of SEER, and strengthening this relationship in this modern era of cancer diagnosis could be mutually beneficial. Achieving this goal requires close interactions between pathologists and the SEER program. This review provides a brief overview of SEER, focuses on facets relevant to pathology practice and research, and highlights the opportunities and challenges for pathologists to benefit from and enhance the value of SEER data.

Address for correspondence: Mark E. Sherman MD, Chief, Breast and Gynecologic Cancer Research Group, Division of Cancer Prevention, National Cancer Institute, 9609 Medical Center Drive, Room 5E334, Bethesda, MD 20892, USA. Phone: 240-276-7051, Fax: 240-276-7828, shermanm@exchange.nih.gov.

Conflicts of interest: None.

Introduction

Modern approaches to cancer control are evolving from simply counting cancer cases within populations at risk to monitoring the occurrence of cancer and cancer precursors, as defined by complex taxonomies that incorporate histopathology, molecular events, and clinical features. This modernization is an adaptation to improvements in clinical practice and is needed to inform cancer prevention, screening, diagnosis, and treatment in populations and individuals. Spurred by advances in molecular biology, including results from "The Cancer Genome Atlas" (cancergenome.nih.gov), investigations of cancer subtypes defined not only by the histopathology but also by molecular profile are an integral part of epidemiological and clinical cancer research[1]. Merging molecular data with histopathological diagnoses into meaningful cancer classifications is a central goal in cancer control and is redefining the practice of oncology.

To improve the accuracy and reproducibility of pathology diagnoses in the face of growing diagnostic complexity, the pathology specialties have developed enhanced post-graduate medical education curricula, and implemented regulated/mandatory, laboratory quality practices, which include the continuing medical education of pathologists[2, 3]. Curricula have expanded far beyond the fundamental principles of histopathology to include knowledge of molecular taxonomies of cancer and classifications that integrate multi-disciplinary data. For example, newer diagnostic classifications, such as those for hematologic and pulmonary malignancies, integrate histopathology, molecular biomarkers and clinical presentations[4, 5]. Pathologists today are expected to provide an accurate, state-of-the-science, quality cancer diagnosis, and to communicate its sometimes uncertain prognostic and predictive implications to multi-disciplinary management teams.

Information about the clinical implications of pathology diagnoses is obtained from sources such as clinical case series, observational epidemiological studies, and randomized clinical trials[6]. Case series reflect a selected experience at one or a few institutions, while even prospective cohorts and/or randomized clinical studies lack generalizability as they include only a fraction of the general population and are biased, including over-representation of healthier and younger participants with higher socioeconomic status (SES)[7]. These data are not representative of the general U.S. (United States of America) population nor reflective of U.S. oncology and pathology practice in the general medical community. The Surveillance, Epidemiology and End Results (SEER) program, however, is closest to that representativeness as it provides population-based data from 18 Cancer Registries from population-based catchment areas related to cancer diagnoses, treatment and survival for approximately 30% of the U.S. population[8]. It generates incidence and survival data for cancer subtypes, defined by histopathology, and increasingly by molecular characteristics. Thus, SEER is an important resource for understanding the implications of pathology diagnoses across demographic groups, geographic regions, and time, and provides unique

insights into the practice of oncology and pathology in the U.S not attainable from other sources.

Pathology has always been central and critical to the effectiveness of SEER, and strengthening this relationship in the modern era of cancer control could be mutually beneficial. SEER records pathology diagnoses rendered in routine practice; therefore, the quality of SEER data and pathology practice are inextricably linked. SEER in turn provides information and data which pathologists could use to further advance their roles as medical experts, scholars and managers. Further, through the Residual Tissue Repository (RTR) and the proposed Virtual Tissue Repository (VTR), SEER is working to set up bio-specimen banking which will be available as a research resource[9]. The goal of this article is to provide an overview of SEER for the pathology community, highlight aspects relevant to diagnostic pathology, present select examples for illustration, and stimulate interest amongst pathologists in working with registries such as SEER to enhance research and monitor practice.

Background

The SEER program was launched on January 1, 1973 by president Richard Nixon as part of the National Cancer Act[8]. SEER collects demographic, clinical and outcome information on all cancers diagnosed in representative geographic regions and subpopulations. Regions are included based on their ability to operate and maintain a high quality population-based cancer reporting system or Cancer Registry and to enhance the demographic and geographic diversity of the SEER data. Initially, 7 (SEER 7) registries with epidemiologically significant population subgroups of racial and ethnic minorities were included, and incrementally were expanded to the current 18 (SEER 18) Cancer Registries (Figure 1). The population covered by SEER is representative of the general U.S. population in regard to measures of poverty and education. However, the SEER population tends to have a higher proportion of foreign-born persons and urban dwellers and over-samples certain racial and ethnic minorities in order to improve diverse population representativeness.

SEER currently captures 400,000 cancer cases annually and stores cancer data on approximately 30% of the U.S. population. The pathology report is an important source for abstracting SEER data, and for approximately 80% of cases, pathology reports are obtained electronically in real-time from approximately 360 laboratories. The abstracted pathology data are consolidated into a final case record along with data from other sources. These sources include patient medical records, reports gathered from freestanding diagnostic imaging and chemotherapy clinics, and death certificates. Traditionally cancer registry staff members abstract standard data items, and manually enter corresponding text into a data collection template. The use of electronic pathology reporting by nearly 80% of laboratories has the potential to enable the use of natural language processing (NLP) software to automatically code data fields. These abstracted records are then reviewed by the registry staff. Afterwards, all data are checked, edited and incorporated into the annual registry database, and submitted in a de-identified form to the National Cancer Institute (NCI). SEER submissions are checked for quality and completeness in the first November

following the final reporting year, and data are released for public use and access in April of the subsequent year.

SEER is administered and funded by the Division of Cancer Control and Population Sciences (DCCPS) at the NCI. Co-funding is provided for select SEER registries via the Centers for Disease Control and Prevention (CDC), National Program of Cancer Registries (NPCR), and participating state jurisdictions. Of note, SEER does not collect data on the entire U.S. population. However, it does coordinate with the North American Association of Central Cancer Registries (NAACCR) and NPCR to collect cancer data for the total U.S. population. The annual report of federal, cancer statistics is published as the "United States Cancer Statistics: Incidence and Mortality Report" and covers 96% of the U.S. population from 45 states, District of Columbia, Puerto Rico and the U.S. Pacific Islands. This collaboration is intended to inform and update health professionals, the U.S. public and political leaders on the effectiveness of the country's cancer control programs, strategies and initiatives.

SEER Data

Data collected for all primary invasive cancers and some other diagnoses e.g., in-situ carcinomas, include date of diagnosis, and demographic variables such as age at diagnosis, gender, race/ethnicity, and county of residence. Surgical management and/or radiation therapy data relating to the first course of treatment are extracted from health records. Detailed schemes of surgical excision were added in 1983 and by 1998 were completed for all tumor types. The program records the type of radiation therapy and whether delivery was neoadjuvant, adjuvant or intraoperative and data on chemotherapy use (yes, no or unknown) may also be assessed with a specific request. SEER also collects tumor data on anatomic site, laterality for paired organs, size, and histopathological type which is based on the 2000 International Classification of Diseases for Oncology version 3 or ICD-O-3 (www.who.int/ classifications/icd/adaptations/oncology). Tumor markers for some cancers, e.g., testis, breast, and prostate were added in 2004. As of 2010, tumor grade, extension/metastasis, site specific factors and stage were captured based on version 7 of the American Joint Committee on Cancer (AJCC. www.cancerstaging.org). Cancer data are updated annually to capture vital status, survival time, and cause of death. Follow up interval in SEER's original 7 Tumor Registries now exceeds 40 years. Vital status is confirmed by linkage to the National Death Index and with supplemental data on date of last known contact obtained by medical record abstraction.

SEER is considered the gold standard for data quality amongst cancer registries in the US and globally. Quality is maintained through contractual agreements with regional registries and SEER's standards must be met before the data are transmitted[8]. In particular, virtual editing of the individual data submissions and consolidated abstracts are routinely performed on cases ranging from 10% to 100% of all abstracts at the individual registries. SEER also has a quality program which includes ongoing education, training, and support for the regional registrars, quality control of the data to prevent and correct errors in coding and identify missing data, and scheduled monitoring and evaluation of the data to identify areas needing improvement. The quality of SEER data is critically dependent on the accuracy and

completeness of pathology reports. Laboratory use of standardized terminology and reporting templates e.g., College of American Pathologists (CAP) synoptics which reflect current biological and clinical thinking, and World Health Organization (WHO. www.whobluebooks.iarc.fr/) classifications of tumors would enhance and facilitate data abstraction[10].

Access to the SEER website (www.seer.cancer.gov) is unrestricted and SEER information may be reproduced or copied without permission. The "Cancer Statistics Review (CSR)" option provides summaries of all cancers and site specific cancers in easy to understand text, graphs and figures. The summaries include 5-year survival data, relative survival as compared with the general population, survival by tumor stage and race/ethnicity, the cancer's incidence ranking, risk factors for the cancer, and life time risk of acquiring the cancer. The "Researcher" option has information on the available data sets and software to analyze them. In addition to cancer data sets, other data sets in the SEER program are Standard Population Data for SEER areas, U.S. Mortality Data, and U.S. Population Data linked to a Census Tract Socio-Economic-Status (SES) Index or to County Attributes. These data can be used for matched analyses with SEER cancer data. Additionally, data sets are linked with other data bases to support a broader set of research including Medicare, Medicare Health Outcomes Survey (SEER-MHOS), National Longitudinal Mortality Study (NLMS), and linked bio-specimen collections. The online software includes SEER*Stat, SEER PREP, JointPoint, and the Health Disparities Calculator. Use of all data bases and software is free and can be accessed by completing the online application form (www.seer.cancer.gov/data/access_seer_data.pdf).

SEER Bio-specimen Pilot Programs

SEER places great importance on the availability of pathology materials for analyses such as immunohistochemical (IHC) testing and next generation sequencing. The RTR and the newly proposed VTR are recent pilot programs designed to scale up the "bio-banking" of pathology material from various cancer cases and to link (annotate) the tissues to the full SEER dataset[9]. The RTR maintains tissue from 3 cancer registries (Iowa, Hawaii and Los Angeles) and is comprised of formalin fixed paraffin embedded tissue blocks on all site specific cancers. By 2010, the estimated number of cancer tissues was 141,241, and the 4 largest cancer groups were lung, colon/rectum, breast and prostate. Tissue microarrays of some cancers e.g., breast, ovary, and colon/rectum are also available. Researchers can access this population-based material by submitting the online application (www.seer.cancer.gov/biospecimen/application.html), and providing a brief summary of the proposed study.

The VTR is a pilot project involving 7 SEER registries. The initial pilot is designed to provide information on costs and best practices to scale this process to a larger set of SEER registries. It will specifically explore the ability to annotate tissues from pancreatic ductal adenocarcinoma patients who survived at least 5 years, and localized node-negative, female breast cancer patients who died of their cancer in a short time interval. Cases will be matched to controls with more typical patterns of survival and will be based on tumor and demographic attributes identified in logistic regression models. The initial pilot project will define best practices for population-based bio-specimen acquisition. Custom annotated

information will be collected including co-morbidities, detailed chemotherapy, time-to-recurrence, and outcomes. Laboratory surveys will collect information on the location of tissues, retrieval costs, and requirements for release of de-identified data to investigators. The pilot will also explore best practices for acquisition of materials, and linkage with digital images and pathology review. The goal is to scale-up a future SEER VTRs that can support a broad range of current cancer research questions.

SEER Data Analyses

The SEER program consists of a consortium of 18 regional cancer registries with meticulous and consistent data collection and standards. The SEER program provides annual frequency distributions, incidence, prevalence and mortality rates over time on all cancers and site specific cancers in its 18 well-defined catchment areas[8]. SEER cancer rates are age-standardized which adjusts the age distributions within or among the populations at risk to a standard population. Age standardization or adjustment enables comparisons of cancer rates among different racial groups and geographic locations. For example, age-standardized cancer incidence rates can be compared between Hawaii and Utah even though the general population is older in Hawaii than Utah. Age-adjustment also enables comparisons of incidence rates by calendar period irrespective of changing age structure of the populations at risk over time.

Because SEER provides data suitable for comparative analyses of cases within populations by defined characteristics, it can be used to answer critical questions about racial disparities, effects of new medical practices and changes in etiologic exposures. Age-adjusted incidence or mortality rates of cancer represent absolute risks of cancer, usually expressed as the number of newly diagnosed cancer cases or deaths per 100,000 persons per year (i.e., per 100,000 person-years). Cancer-specific survival is the percentage of individuals with cancer alive following a specified interval after diagnosis. Survival data are often considered in pathology case series when the numerator of mortality (number of deceased persons or cases) is complete but the denominator (the population at risk) is not; i.e., mortality data are generally only available in population-based datasets such as SEER when the number of individuals in the entire population at risk of developing cancer is recorded.

SEER also generates age-specific incidence rates for carcinoma in-situ and invasive carcinomas by stage, which informs (generally) the average times between various disease states. For example, the large difference in the average age at diagnosis of cervical carcinoma *in-situ* and invasive cervical squamous cell carcinoma and invasive adenocarcinoma supports the value of screening and intervention to prevent these cancers, and provides value in considering clinical guidelines. As the SEER program includes increasingly more refined information on the more complex characterization of cancer cases, including information such as biomarkers and molecular profiles, these data will increasingly enable computation of refined rates and trends - moving towards "precision surveillance".

SEER data resources are extensively analyzed by researchers worldwide and provide critical insights about cancer and the practice of oncology in the U.S (Table 1)[11]. The 40,031

citations in a PubMed (www.ncbi.nlm.nih.gov/pubmed) search from 1973 to 2015 using key words Surveillance, Epidemiology and End Results confirm, exemplify and underscore the research productivity originating from the program (Figure 2). This research is typically observational and examines the distribution of cancer in populations and groups and how demographic, clinico-pathological and treatment variables affect cancer burden. Some studies link SEER data sets to others that are more richly annotated. These composite data sets can then be analyzed to identify possible risk factors and thereby generate hypotheses to be tested via experimental study designs.

The large numbers of cases in SEER are a major strength of these analyses since studies may be amply powered to evaluate associations or precisely estimate rates by finely sub-classified diagnostic categories and/or rare diseases. Large case numbers also lessen the impact of sporadic or random misclassification of pathology diagnoses. These are expected since the pathology data in SEER reflect usual U.S. practice and cannot be eliminated as SEER's quality processes are not designed to check the accuracy of the pathology diagnosis. SEER is reliant on the laboratories quality practices to minimize such errors[3]. In a large sample size however, such random misdiagnoses will have non-differential impact upon the results as they tend to regress the mean to a null result and therefore, i.e., will blunt rather than create a signal in the data.

Absolute Risk of a Cancer

Importantly, SEER rates yield insights about cancer that are not attainable through a simple estimation of numbers of cases or frequency distributions (percentages). Rates inform us about an individual's absolute risk of developing cancer since the calculation is based on the complete enumeration of all cancer cases in a given population at risk within a well-defined catchment area. Percentages on the other hand merely inform us of the number of specific cancers expressed as a proportion of all cancers at a point in time. Even in studies that collect every cancer from a population, the percentage of cancers by subtype may be misleading because the rate of cancers in the at-risk population is not considered. For example, Horne et al, using SEER data showed U.S. breast cancer rates are extremely high overall, largely because of the high incidence rate of estrogen receptor (ER)-positive cancers among older women[12]. However, rates of ER-negative triple negative cancers in the U.S. are also higher than in many other countries e.g., Malaysia, even though the percentage or relative frequency distribution of triple negative cancer is higher in Malaysia than in the US. In contrast, studies performed in SEER have confirmed results from epidemiologic studies showing that African Americans have notably higher population-based incidence rates of triple negative breast cancer[13].

Rare Cancers and Cancers in Minority Groups

Much of the knowledge on the epidemiology of rare cancers such as dermatofibrosarcoma protuberans (DFSP) or the childhood and atypical teratoid/rhabdoid (ART) tumor of the central nervous system stems from studies utilizing SEER data. Criscione and Weinstock in their study of 2,885 DFSP found the incidence was higher among women although case series from regional U.S. treatment centers had reported a higher incidence amongst men[14]. This contradiction exposed the importance of understanding that a case series reflects the

pattern of disease in a population serviced by a treatment center and may not be representative of the general population. Knowledge about ART which is a rare, locally aggressive childhood malignancy was limited and based on case reports or small retrospective, observational series. In a study of 174 ART registered between 1973 and 2010, Lau and colleagues were able to demonstrate the incidence was higher in males, Whites and children less than 3 years, and that it most frequently presented as a loco-regional 4 cm or more tumor involving the cerebellum, ventricles or frontal lobes[15].

Survival can reflect early detection, better treatment or other health factors, such as improved access to care or possible disparities in access by age, race/ethnicity, minority group or region. The over-sampling of certain racial and ethnic minorities by SEER enables evaluation of these factors. For example, age-standardized, breast cancer incidence rates overall tend to be higher among non-Hispanic White women in comparison to Black women. However, age-specific breast cancer incidence rates appear to be higher among younger Black than younger non-Hispanic White women[16]. Additionally, breast cancer-specific mortality and survival also are worse among Black compared to non-Hispanic White women. These observations suggest Black women, and in particular younger Black women, have challenges in accessing health care because in randomized clinical trials where access to care is more equal, racial survival disparities are minimized and perhaps non-existent[17,18].

Birth Cohort- Effect and Calendar-Period Effect

Changes in cancer incidence over a time period may reflect a change in the prevalence of an important risk factor (birth-cohort or exposure effect) and/or the effects of screening or case ascertainment (calendar-period effect). Analyses of SEER data reaching back for over 4 decades, provide an important context in which to understand temporal trends in cancer rates. Ravdin et al reported the age-adjusted incidence rate of breast cancer in the U.S. in 2003 fell sharply by 6.7%, as compared with the rate in 2002[19]. The decrease was evident only in women who were 50 plus years of age and more particularly evident in ER-positive cancers. The decrease seemed temporally related to a decline in the use of hormone-replacement therapy amongst postmenopausal women in the U.S. The observed conclusion was supported in subsequent epidemiological studies which found that post-menopausal women who developed breast cancer were more likely to be current or recent users of combined estrogen plus progestin hormone replacement therapy[20].

Improved survival may reflect benefits of early detection through screening, or highlight the potential harms of screening, such as the detection of cancers that may have limited clinical consequences. Earlier detection may lead to falsely reassuring estimates of survival because of lead time bias (identified sooner without changing the date of death) and/or length bias (e.g., increased detection of indolent compared to aggressive tumors). Therefore, lengthy follow up, as provided by the SEER program is important for relating stage at detection to mortality rates, although interpretation of such data with regard to screening requires caution[21].

Cancer Risk Prediction Modelling

SEER can be viewed as an open cohort or group of individuals that share a common experience, with possibilities of entering or leaving at any time. Within SEER, sub-cohorts may be defined by year of birth, demographics and date of follow-up and/or diagnosis. The "Age-Period-Cohort" statistical model allows researchers to analyze one of these parameters, adjusted for the others and to independently assess patterns of rates of diagnosis by age at diagnosis, calendar period or year of birth. In many risk prediction models, age-specific SEER incidence data contribute to the baseline risk which is modified by an individual's additional risk factors so as to develop a prediction of the absolute risk of developing a particular cancer at a certain age over a defined period. One of the best known and used is the "Breast Cancer Risk Assessment Tool" or "Gail Model" for estimating risk of developing breast cancer[22]. The model was originally developed for White women and then extended to apply to non-White women (www.cancer.gov.bcrisktool). It is routinely updated to reflect changes in breast cancer incidence[23].

Oncology Practice and Biomarker Utilization

SEER data alone or in conjunction with bio-specimens, biomarkers, and other resources have contributed importantly to understanding the clinical implications of pathology diagnoses. To provide an overview of the richness of possibilities for using SEER to evaluate the extent of biomarker use in oncology practice, a PubMed search using the phrase: "SEER" AND "BIOMARKERS" for the most recent 5-year period was conducted. From 191 references, titles and abstracts of 41 published within a year of the search were briefly reviewed. Many have implications for the clinical interpretation of pathology diagnoses, including assessing prognosis, underscoring reasons for racial disparities and understanding the effects of screening on cancer related incidence and outcomes. Further, some studies may impact pathology practice directly.

One report evaluating a 9-gene signature for lung cancer prognosis found that combining results with basic patient information from SEER led to improved prediction, suggesting that similar to several other cancer types, tumor taxonomies may increasingly incorporate histopathologic classification, molecular testing and clinical factors to guide management[24]. Other studies considered how pathologic under-staging of colon or prostate cancer might be reduced by combining pathology, clinical studies and measurement of circulating markers, such as CEA (Carcino-Embryonic Antigen) or PSA (Prostatic Specific Antigen), respectively[25, 26]. An analysis of mucinous colorectal carcinoma highlighted the clinical implications of this diagnosis with regard to worse outcomes and provided early evidence for prognostic value in testing for expression of *PINCH* (Particularly Interesting New Cysteine-Histidine rich protein) and *RAD50*[27]. Others explored how reporting of prognostic measures may influence treatment, which in turn can influence cancer outcomes; examples included relationships between a 21-gene prognostic signature for breast cancer and frequency of chemotherapy treatment and the intensity of follow-up for low-risk prostate adenocarcinomas managed with "watchful waiting"[28, 29]. Thus, SEER may have value in corroborating results of clinical studies and documenting clinical practices and their relationships to management guidelines. Other studies found that reporting Paget's Disease in conjunction with invasive breast cancer may be related to nodal status, but not survival,

and that lymph node size in breast cancer may have prognostic value, independent of the number of positive nodes[30, 31].

SEER Program Limitations

The best use of SEER data leverages the programs strengths with regard to representativeness and generalizability to the U.S. population, lengthy period of data collection, large numbers of cases, and collection of cancer specific outcomes. Limitations include incompleteness of individual-level data collected on specific cancer risk and treatment, and inaccuracies and incompleteness of the data collected from the source registries[32]. Further, data related to SES, co-morbidities and other health information are lacking, although SES data by censual segment are available. SEER does not collect data on risk-reducing procedures (e.g. risk-reducing salpingo-oophorectomy) or organ removal for non-cancer indications (e.g. hysterectomy); although, it is possible to adjust for such factors based on other information sources.

Tumor recurrence data are currently not collected and therefore progression free survival, correlates of local, regional and distant control, and the effectiveness of salvage therapy cannot be assessed. Moreover, survival outcome cannot be fully evaluated as the collected data does not distinguish the intent of therapy as curative or palliative. Specific details on the type, dose and duration of chemo- and radiation therapy, and the use of other oral pharmaceuticals are not collected. Information gaps in treatment and follow up occur when individuals migrate into and out of SEER and non-SEER regions and could bias conclusions about a cancer behavior, particularly if re-location and outcome are linked. Such interruptions in data collection also occur when individuals within a SEER region have a procedure in a facility which has no contractual obligation to transmit data to SEER or have prescriptions in pharmacies or diagnostic/predictive molecular tests in commercial laboratories who are not obligated to transmit their data to SEER.

Inaccuracies in the source data can occur either due to miscoding of the data transmitted to NCI by the regional registries or the data made available to the regional registrars for coding is not correct. The SEER quality program is designed to identify miscoding of the transmitted data and acts quickly to rectify any identified error. As an example, NCI withdrew access to the PSA data after a routine quality control check discovered a number of registrars were miscoding the decimal point within the 3-digit field[33]. Following review, SEER data for 2009–2013 have been corrected and re-released in keeping with the special commitment of the program to maintaining high quality data for long-term future use. Nonetheless, factors such as assignment of race/ethnicity categories and pathology diagnoses evolve, creating considerations for performing studies over extended calendar periods. A large proportion of the U.S. population does not self-identify as a single race and would now be classified using genetic studies as mixed race. Due to advances in tumor classification, SEER pathology terminology is ever changing and this creates secular issues regarding the comparability of SEER data that are captured over 40 years. Recent SEER pathology diagnoses are expected to be more precise than those from nearly 40 years ago due to advances in diagnostic testing, pathologists training and quality improvement practices such as external consultation and mandatory second reviews[2, 3]. The frequency of

"Poorly Specified Malignancies including Carcinomas Not Otherwise Specified (ICD-O-8000-8010)" for all organ sites is a relatively small category within SEER. It declined from 6 to 5% in the 12-year period 2001 to 2013 (Figure 3). When analyzed by site, the greatest declines occurred in pancreatic cancers. This highlights greater advances in understanding the pathogenesis of this cancer with resultant new diagnostic criteria and a revised tumor classification system[34]. However, the accuracy and reproducibility of specific histotypes and cancer grades in the SEER data are mostly unknown as they have not been studied to any extent. Based on the few studies published, it appears that histotyping accuracy and agreement for some cancers e.g., Hodgkin's diseases may be very good, but agreement in histotyping and grading for others e.g., ovarian carcinoma are variable and organ site/cancer dependent (Table 2)[35–38].

SEER Future and Opportunities for Pathology

Future plans for the SEER program include expanded collection of biomarkers and treatment, custom annotation, linkage with other data bases for complete capture of relevant information, harmonization of coding systems over time, and expanded bio-specimen resources. The use of annotation may be broadened to deliver key variables to address current cancer research questions. Electronic record linkage with pharmacy and commercial biomarker laboratory databases offers the prospect of more detailed and complete treatment and biomarker data. Use of NLP may minimize missing data and misclassification during data abstraction. Variables collected as part of the pathology cancer synoptic report and IHC expressed biomarkers may also be added. Access to registry cancer bio-specimens would increase the number and population-based representativeness of bio-specimens available for cancer research. These changes will increase the role and value of SEER as a cancer research tool and the bio-specimen based research will be of particular interest and value to pathology. Developing access to large collections of digital images has also been considered and could have important value for pathology practice and cancer surveillance research. Pathology collaboration would offer added expertise in SEER studies, even when based purely on data analysis without histopathologic review. Pathologists may aid SEER analysts in converting outmoded pathology terminology/classifications to current versions or review diagnostic slides as part of a centralized standardized review to confirm the diagnosis and add detailed pathology annotation.

SEER can provide valid and reliable information on how diagnostic, predictive and prognostic laboratory tests are used nationally and could be a useful resource for defining essential services provided by laboratories and estimating the necessary resources needed to provide them. The selected examples illustrate how SEER analyses reflect pathology research and practice, and how trends in these arenas can influence oncologists future expectations of pathology. Together with changes in cancer incidence, and projection of future rates and burden, SEER could be used to plan changes in the size and skill set of the laboratory workforce. For example, age-specific incidence rates of cervical cancer now indicate that cervical cytology screening has had a greater impact on lowering incidence of squamous cell carcinoma than adenocarcinoma, highlighting an area for practice improvement[39]. Further, recent analyses of incidence rates adjusted for benign hysterectomy procedures suggest that cervical cancer incidence may remain higher than previously

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

supposed at older ages, providing impetus for re-considering ages at which to discontinue screening which will affect pathology workloads[40]. With increases in SEER annotation, linkage across databases and access to bio- specimens, the value of the program will continue to increase. Strengthening collaboration between pathologists and the SEER program can ensure this success and identify top priorities for expanding the program in practical, highly useful ways.

Acknowledgments

Glossary of acronyms

AJCC	American Joint Commission on Cancer
CAP	College of American Pathologists
CDC	Center for Disease Control and Prevention
CEA	CarcinoEmbryonic Antigen
DCCPS	Division of Cancer Control and Population Services
ER	Estrogen Receptor
ICD-O	International Classification of Diseases for Oncology
NAACCR	North American Association of Central Cancer Registries
NCI	National Cancer Institute
NLMS	National Longitudinal Mortality Study
NLP	Natural Language Processing
NPCR	National Program of Cancer Registries
PINCH	Particularly Interesting New Cysteine-Histidine-rich protein
PSA	Prostatic Specific Antigen
RAD50	DNA repair protein
RTR	Residual Tumor Repository
SES	Socio-Economic Status
SEER	Surveillance, Epidemiology and End Results
SEERMHOS	Surveillance, Epidemiology and End Results Medicare Health Outcomes Survey
U.S.	United States of America
VTR	Virtual Tissue Repository

WHO   World Health Organization

References

1. Horlings HM, Flanagan AM, Huntsman DG. Categorization of cancer through genomic complexity could guide research and management strategies. J Pathol. 2015 Aug.236:397–402. [PubMed: 25864408]

2. Alexander CB. Pathology graduate medical education (overview from 2006–2010). Hum Pathol. 2011; 42:763–769. [PubMed: 21333325]

3. Nakhleh RE. Core components of a comprehensive quality assurance program in anatomic pathology. Adv Anat Pathol. 2009; 16:418–423. [PubMed: 19851132]

4. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization (WHO) classification of lymphoid neoplasm. Blood. 2016; 127:2375–2390. [PubMed: 26980727]

5. Lindeman NI, Cagle PT, Beasley MB, et al. Molecular Testing Guideline for Selection of Lung Cancer Patients for EGFR and ALK Tyrosine Kinase Inhibitors. Guideline from the College of American Pathologists, International Association for the Study of Lung Cancer and the Association for Molecular Pathology. J Molecular Diagnostics. 2013; 15:416–453.

6. Hall KN, Kothari RU. Research fundamentals: IV. Choosing a research design. Acad Emerg Med. 1999; 6:67–74. [PubMed: 9928979]

7. Jordan S, Watkins A, Storey M, et al. Volunteer bias in recruitment, retention, and blood sample donation in a randomised controlled trial involving mothers and their children at six months and two years: a longitudinal analysis. PLoS One. 2013; 8:1–17.

8. Park HS, Lloyd S, Decker RH, et al. Overview of the Surveillance, Epidemiology, and End Results database: evolution, data variables, and quality assurance. Curr Probl Cancer. 2012; 36:183–190. [PubMed: 22481006]

9. Altekruse SF, Rosenfeld GE, Carrick DM, et al. SEER cancer registry biospecimen research: yesterday and tomorrow. Cancer Epidemiol Biomarkers Prev. 2014; 23:2681–2687. [PubMed: 25472677]

10. Srigley JR, McGowan T, Maclean A, et al. Standardized synoptic cancer pathology reporting: a population-based approach. J Surg Oncol. 2009; 99:517–524. [PubMed: 19466743]

11. Yu JB, Gross CP, Wilson LD, et al. SEER public-use data: applications and limitations in oncology research. Oncology (Williston Park). 2009; 23:288–295. [PubMed: 19418830]

12. Horne HN, Beena Devi CR, Sung H, et al. Greater absolute risk for all subtypes of breast cancer in the US than Malaysia. Breast Cancer Res Treat. 2015; 149:285–291. [PubMed: 25537643]

13. Howlader N, Altekruse SF, Li CI. US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. J Natl Cancer Inst. 2014:106.

14. Criscione VD, Weinstock MA. Descriptive epidemiology of dermatofibrosarcoma protruberans in the United States, 1973 to 2002. J Am Acad Dermatol. 2007; 56:968–973. [PubMed: 17141362]

15. Lau CS, Mahendraraj K, Chamberlain RS. Atypical teratoid rhabdoid tumors: a population-based clinical outcomes study involving 174 patients from the Surveillance, Epidemiology, and End Results database (1973–2010). Cancer Manag Res. 2015; 7:301–309. [PubMed: 26425106]

16. Menashe I, Anderson WF, Jatoi I, et al. Underlying causes of the black-white racial disparity in breast cancer mortality: a population based analysis. J Natl Cancer Inst. 2009; 101:993–1000. [PubMed: 19584327]

17. Albain KS, Unger JM, Crowley JJ, et al. Racial disparities in cancer survival among randomized clinical trials patients in the Southwest Oncology Group. J Natl Cancer Inst. 2009; 101:984–992. [PubMed: 19584328]

18. Rosenberg PS, Menashe I, Jatoi I, et al. Re: Racial disparities in cancer survival among randomized clinical trials patients in the Southwest Oncology Group. J Natl Cancer Inst. 2010; 102:277.

19. Ravdin PM, Cronin KA, Howlader N, et al. The decrease in breast-cancer incidence in 2003 in the United States. N Engl J Med. 2007; 356:1670–1674. [PubMed: 17442911]

20. Beral V, Reeves G, Bull D, et al. Breast cancer risk in relation to the interval between menopause and starting hormone replacement therapy. J Natl Cancer Inst. 2011; 103:296–305. [PubMed: 21278356]

21. Elmore JG, Etzioni R. Effect of screening mammography on cancer incidence and mortality. JAMA Intern Med. 2015; 175:1490–1. [PubMed: 26147329]

22. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst. 1989; 81:1879–1886. [PubMed: 2593165]

23. Schonfeld SJ, Pee D, Greenlee RT, et al. Effect of changing breast cancer incidence rates on the calibration of the Gail model. J Clin Oncol. 2010; 28:2411–2417. [PubMed: 20368565]

24. Gentles AJ, Bratman SV, Lee LJ, et al. Integrating Tumor and Stromal Gene Expression Signatures With Clinical Indices for Survival Stratification of Early-Stage Non-Small Cell Lung Cancer. J Natl Cancer Inst. 2015; 107:1–11.

25. Thirunavukarasu P, Talati C, Munjal S, et al. Effect of Incorporation of Pretreatment Serum Carcinoembryonic Antigen Levels Into AJCC Staging for Colon Cancer on 5-Year Survival. JAMA Surg. 2015; 150:747–755. [PubMed: 26083632]

26. Dinh KT, Mahal BA, Ziehr DR, et al. Incidence and Predictors of Upgrading and Up Staging among 10,000 Contemporary Patients with Low Risk Prostate Cancer. J Urol. 2015; 194:343–349. [PubMed: 25681290]

27. Wang MJ, Ping J, Li Y, et al. Prognostic Significance and Molecular Features of Colorectal Mucinous Adenocarcinomas: A Strobe-Compliant Study. Medicine (Baltimore). 2015; 94:1–8.

28. Dinan MA, Mi X, Reed SD, et al. Association Between Use of the 21-Gene Recurrence Score Assay and Receipt of Chemotherapy Among Medicare Beneficiaries With Early-Stage Breast Cancer, 2005–2009. JAMA Oncol. 2015; 1:1098–1109. [PubMed: 26313372]

29. Chamie K, Williams SB, Hershman DL, et al. Population-based assessment of determining predictors for quality of prostate cancer surveillance. Cancer. 2015; 121:4150–4157. [PubMed: 26307939]

30. Wong SM, Freedman RA, Sagara Y, et al. The effect of Paget disease on axillary lymph node metastases and survival in invasive ductal carcinoma. Cancer. 2015; 121:4333–4340. [PubMed: 26376021]

31. Rose BS, Jiang W, Punglia RS. Effect of lymph node metastasis size on breast cancer-specific and overall survival in women with node-positive breast cancer. Breast Cancer Res Treat. 2015; 152:209–216. [PubMed: 26041688]

32. Park HS, Lloyd S, Decker RH, et al. Limitations and biases of the Surveillance, Epidemiology, and End Results Database. Curr Probl Cancer. 2012; 36:216–224. [PubMed: 22481009]

33. Sun M, Trinh QD. A Surveillance, Epidemiology and End Results (SEER) database malfunction: perceptions, pitfalls and verities. BJU Int. 2016; 117:551–5522. [PubMed: 26190064]

34. Hruban RH, Iacobuzio-Donahue C, Wilentz RE, et al. Molecular pathology of pancreatic cancer. Cancer J. 2001; 7:251–258. [PubMed: 11561601]

35. Shuch B, Hofmann JN, Merino MJ, et al. Pathologic validation of renal cell carcinoma histology in the Surveillance, Epidemiology, and End Results Program. Urol Oncol. 2014; 32:1–13. [PubMed: 24360658]

36. Field RW, Smith BJ, Platz CE, et al. Lung cancer histologic type in the Surveillance Epidemiology, and End Results registry versus independent review. J Natl Cancer Inst. 2004; 96:1105–1107. [PubMed: 15265973]

37. Glaser SL, Dorfman RF, Clarke CA. Expert review of the diagnosis and histologic classification of Hodgkin Disease in a population-based cancer registry: interobserver reliability and impact on incidence and survival rates. Cancer. 2001; 92:218–224. [PubMed: 11466672]

38. Matsuno RK, Sherman ME, Visvanathan K, et al. Agreement for tumor grade of ovarian carcinoma: analysis of archival tissues from the surveillance, epidemiology and end results residual tissue repository. Cancer Causes and Control. 2013; 24:749–757. [PubMed: 23378140]

39. Adegoke O, Kulasingam S, Virnig B. Cervical cancer trends in the United States: a 35-year population-based analysis. J Womens Health (Larchmt). 2012; 21:1031–1037. [PubMed: 22816437]

40. Rositch AF, Nowak RG, Gravitt PE. Increased age and race-specific incidence of cervical cancer after correction for hysterectomy prevalence in the United States from 2000 to 2009. Cancer. 2014; 120:2032–2038. [PubMed: 24821088]
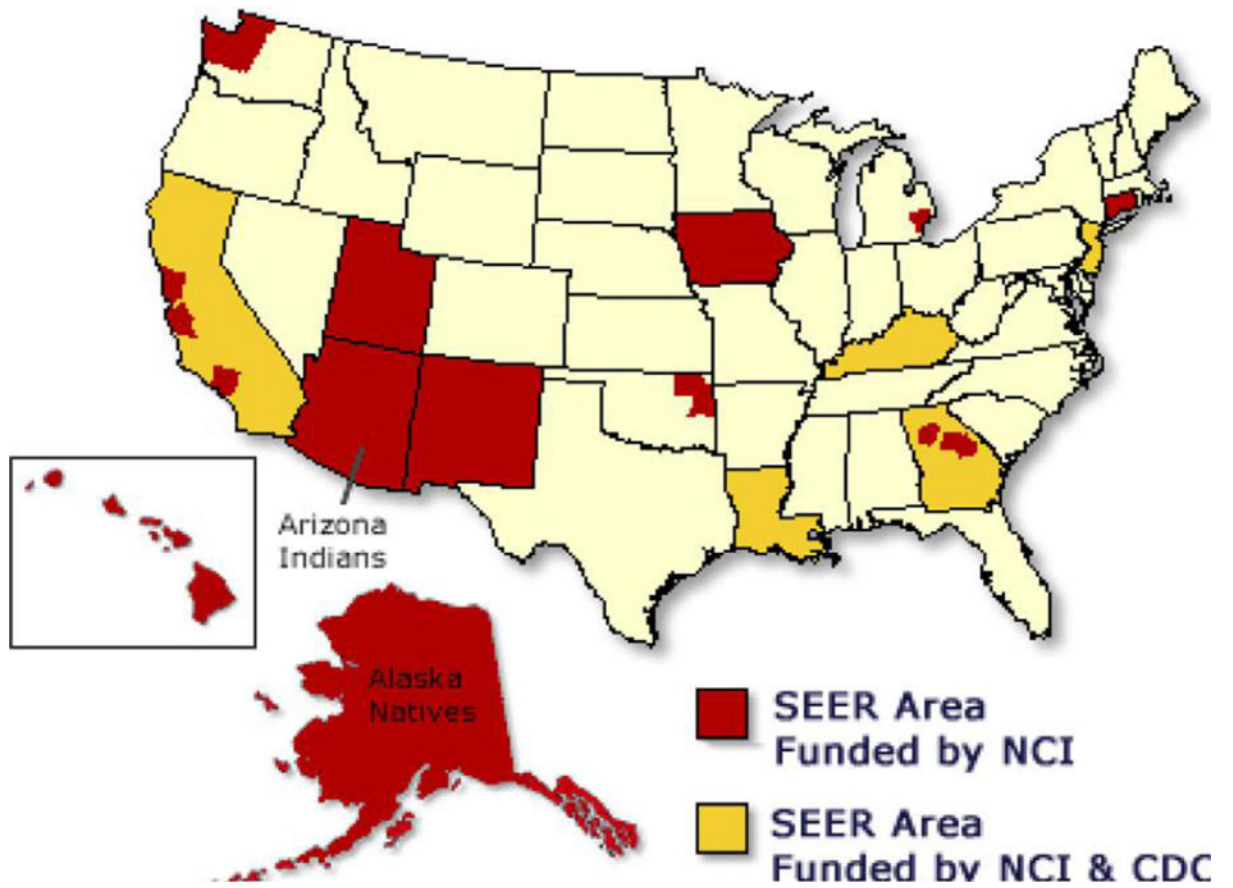
Figure 1. Map of Surveillance, Epidemiology, and End Results registries in the United States of America
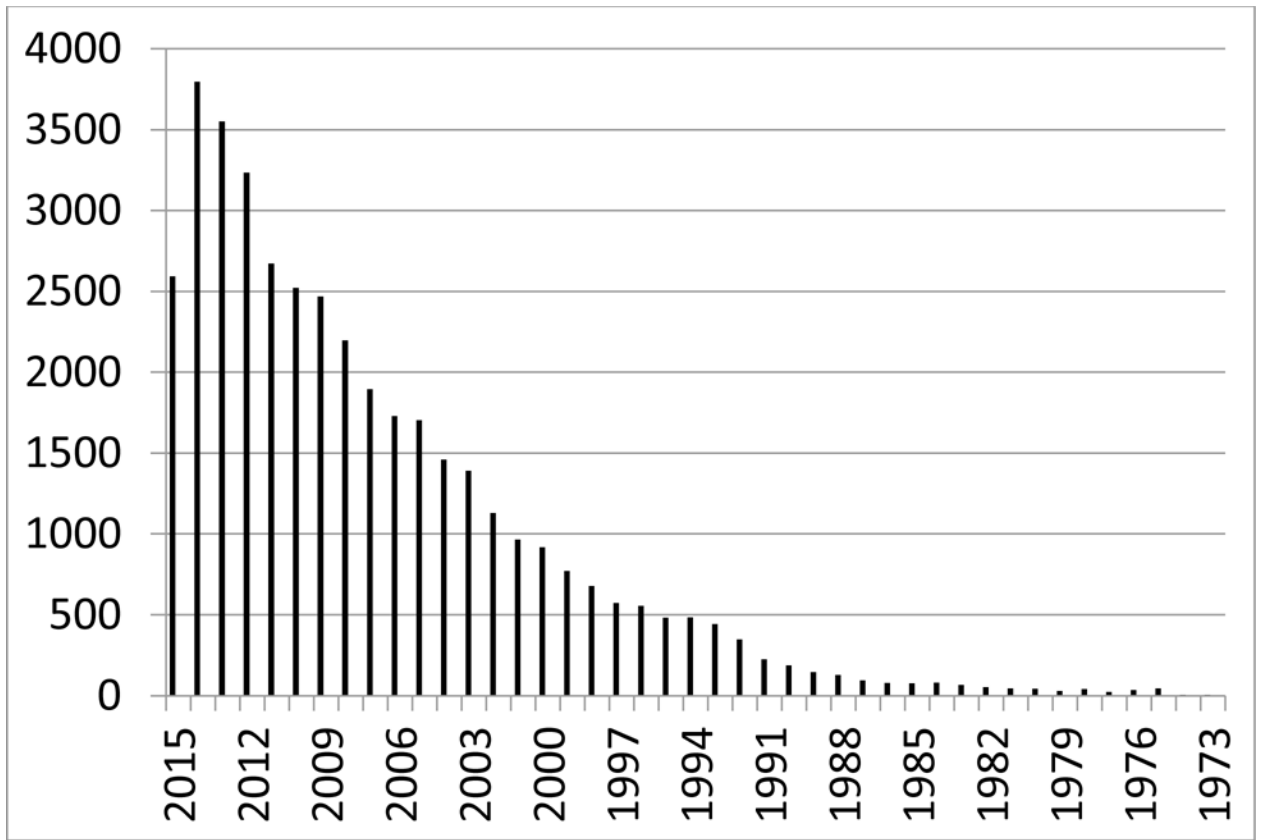Geographic distribution of SEER cancer registries.

Figure 2. Numbers of Surveillance Epidemiology and End Results Pubmed* Citations: 1973–2015
Citations in literature referencing SEER by year.
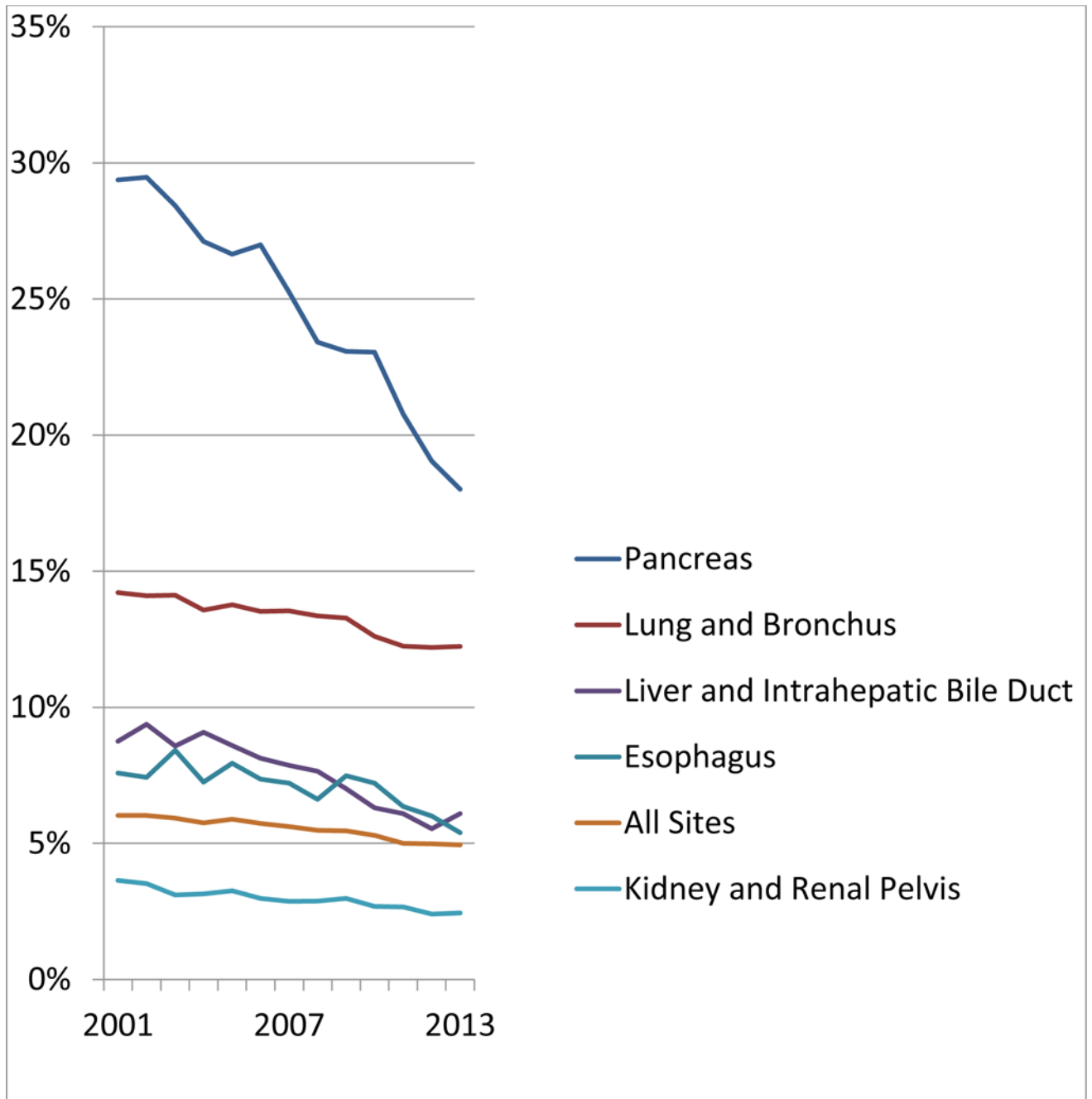
*www.pubmed.gov

Figure 3. Percentage of cases in SEER diagnosed with poorly specified histologies* for select cancer sites: 2001–2013
Trends in poorly define histopathologic subtypes of cancer by site and year.
*ICD-O-8000-8100

Table 1

Surveillance, Epidemiology and End Results Program: Unique Analyses and Critical Insights

Analyses		**Critical Insights**	
1	Population-based cancer rates	**1**	Absolute risk of cancer occurrence
2	Rare cancer rates	**2**	Precise and comprehensive description
3	Cancers rates in minority groups	**3**	Health disparity assessments
4	Birth-cohort effect	**4**	Risk factor exposure assessments
5	Calendar-period effect	**5**	Benefits/harms of screening
6	Risk modelling	**6**	Individual's risk of a particular cancer
7	Oncology practice and biomarker utilization	**7**	Therapeutic targets and diagnostic, predictive and prognostic markers.

Table 2

Accuracy/reproducibility of select SEER pathology diagnoses in comparison to a standard review

Organ site and/or cancer	Pathology standard	Accuracy/reproducibility
Kidney/renal cell carcinoma[35] Clear cell carcinoma Papillary carcinoma Chromophobe carcinoma	1 pathologist	Agreement=78.2% (kappa=0.55: Moderate) Sensitivity/Specificity=79.1/88.1% Sensitivity/Specificity=73.5/97.5% Sensitivity/Specificity=72.4/97.6%
Lung[36] Squamous cell carcinoma Small cell carcinoma Large cell carcinoma Adenocarcinoma	2 pathologists	Agreement=91% Sensitivity/Specificity=70.9/96.2% Agreement=98%. Sensitivity/Specificity=94.1/98.8% Agreement=87.9%. Sensitivity/Specificity=21.9/93.7% Agreement=82.9%. Sensitivity/Specificity=80.8/84.4%
Hodgkin's disease[37]	1 pathologist	Agreement=68% (kappa=0.66: Substantial)
Ovarian carcinoma[38] *3 tier grade* Serous carcinoma – (2 tier grade) Mucinous carcinoma Endometrioid carcinoma Clear cell carcinoma	1 pathologist	Agreement=57% (kappa=0.21: Fair) – Agreement=64% (kappa=0.10: Slight) Agreement=44% (kappa=0.26: Fair) Agreement=46% (kappa=0.26: Fair) Agreement=24% (kappa=0.00: Chance)