HHS Public Access

Author manuscript

Ann Surg Oncol. Author manuscript; available in PMC 2024 September 01.

Published in final edited form as:

Ann Surg Oncol. 2024 September ; 31(9): 5546–5559. doi:10.1245/s10434-024-15393-8.

National Cancer Database Conforms with Standardized Framework for Registry and Data Quality

Bryan Palis, MA[1], **Lauren Janczewski, MD MS**[1], **Amanda E. Browner, MS**[1], **Joseph Cotler, PhD**[1], **Leticia Nogueira, PhD MPH**[2], **Lisa C. Richardson, MD MPH**[3], **Vicki Benard, PhD**[3], **Reda J. Wilson, MPH CTR**[3], **Nadine Walker, MS CTR**[4], **Daniel Boffa, MD MBA FACS**[5], **Heidi Nelson, MD FACS**[1]

[1]American College of Surgeons

[2]American Cancer Society

[3]Centers for Disease Control and Prevention

[4]National Cancer Registrars Association

[5]Yale School of Medicine

Abstract

Purpose—Standardization of practices and procedures for data abstraction by cancer registries is fundamental for cancer surveillance, clinical and policy decision making, hospital benchmarking, and research efforts. The objective of the present study was to evaluate adherence to the 4 components (completeness, comparability, timeliness, and validity) defined by Bray and Parkin that determine registries' ability to carry out these activities to the hospital-based National Cancer Database (NCDB).

Methods—We used data from US Cancer Statistics, the official federal cancer statistics and joint effort between CDC and NCI, which includes data from NPCR and SEER, to evaluate NCDB completeness between 2016 through 2020. We evaluated comparability of case identification and coding procedures. We utilized Commission on Cancer (CoC) Standards from 2022 to assess timeliness and validity.

Results—Completeness was demonstrated with a total of 6,828,507 cases identified within the NCDB, representing 73.7% of all cancer cases nationwide. Comparability was followed by the use of standardized and international guidelines on coding and classification procedures. For timeliness, hospital compliance with timely data submission was 92.7%. Validity criteria for re-abstracting, recording, and reliability procedures across hospitals demonstrated 94.2% compliance. Additionally, data validity was shown by a 99.1% compliance with histological verification standards, 93.6% assessment of pathological synoptic reporting, and 99.1% internal consistency of staff credentials.

Conclusion—The NCDB is characterized by a high level of case completeness and comparability with uniform standards for data collection and by hospitals with high compliance,

Corresponding Author: Bryan Palis, American College of Surgeons, 633 North Saint Clair St., Chicago, IL 60611-3295, bpalis@facs.org.

timely data submission, and high rates of compliance with validity standards to registry and data quality evaluation.

Keywords

NCDB; Registry; Quality; Standardization; Coverage; Comparability; Timeliness; Validity

Introduction

Medical practices and advances in healthcare are information dependent, and both rely on high quality data. In recent years, the availability of healthcare data and analytic platforms has grown exponentially with increasing use of electronic medical records and insurance claims. However, just as the evidence generated by clinical trials is rigorously tested through a set of preexisting data quality procedures,[1,2] other sources of data could also be graded in a uniformly defined and regulated manner. The usability of all data sources is crucial to understanding strengths and limitations. With new data sources becoming more accessible among clinicians and researchers to help shape the future of healthcare, ensuring data quality through a standardized evaluation plays an increasingly critical role. One such standardized approach to assessing the quality of data collected by cancer registries is the framework described by Bray and Parkin in 2009.[3,4]

The Bray and Parkin registry and data quality framework was developed with 4 unique domains: completeness, comparability, timeliness, and validity.[3,4] Completeness represents the extent to which all the incident cancers occurring in the population are included in a registry.[3,4] Completeness is crucial for ensuring estimates approximate the true value in the population.[3,4] Comparability represents the extent to which statistics generated for different populations, using data from different sources and over time, can be compared.[3,4] Comparability is achieved using standardized guidelines on classification procedures, maintaining consistency for coding cancer cases.[3,4] Timeliness relates to the rapidity through which a registry can abstract and report reliable cancer data, which is crucial for decision making.[3,4] Validity represents the proportion of cases in a dataset with a given characteristic that truly has that attribute, which is crucial for relevant interpretation of estimates calculated using the data.[3,4] Importantly, this framework has been applied across numerous cancer registries worldwide, demonstrating its ability to affirm, document, and benchmark data quality.[5–7]

The processes that assure data quality of both population and hospital-based cancer registries in the USA have been consistent for several decades and include standardization of data field definitions, quality checks executed during data abstraction, and case monitoring following submission (Figure 1). The principal aim of a population-based cancer registry is to record all new cases in a geographical area or state with an emphasis on epidemiology and public health.[8,9] A hospital-based registry by contrast is designed to improve patient quality of care at the institutional level.[8,9] Cancer surveillance programs collaborate to standardize definitions of relevant cancer data items and closely monitor estimates of cancer trends and outcomes calculated using different data sources.[9] Each cancer surveillance program works with oncology data specialists or certified tumor registrars who are educated, trained, and

certified in abstracting cancer data following established definitions and rules.[9,10] While these processes, among many others, have demonstrated consistency over time, they are also dynamic and undergo periodic revisions to incorporate advances in cancer care and ensure the availability of contemporary cancer data.[9,10] Thus, an assessment of existing quality processes and procedures is fundamentally important to ensuring the best possible data are being used to inform cancer practices and policies. While the Bray and Parkin quality control criteria were written primarily with population-based registries in mind, we propose their use for large hospital-based registries, such as the National Cancer Database (NCDB). The principal aim of this study is to assess the quality of cancer data collected by the NCDB using the Bray and Parkin framework.

Methods

The NCDB is a hospital-based cancer registry and contains approximately 40 million records, collecting data on patients with cancer since 1989.[11,12] The NCDB is jointly maintained by the American College of Surgeons Commission on Cancer (CoC) and the American Cancer Society as an obligatory component of CoC hospital accreditation.[12,13] To earn voluntary CoC accreditation, a hospital must meet quality of patient care and data quality standards.[13] Hospitals are evaluated on a triennial basis through a site visit process to maintain levels of excellence in the delivery of comprehensive patient-centered care.[13] Overall, approximately 1500 CoC-accredited hospitals submit data to the NCDB each year.[11] The NCDB collects data from patients in all phases of first course treatment in cancer care and cancer surveillance and includes the addition of roughly 1.5 million patients newly diagnosed cancers annually.[11,12,14]

Quality Procedures and Processes

Both population-based and hospital-based cancer registries adhere to uniform procedures during the record abstraction and coding process to ensure accuracy but serve different purposes. The reporting of cancer cases to the population-based central cancer registry (CCR) is mandated by legislation in all US states and territories.[15,16] The cases identified by these CCRs are then reported to national cancer registries.[15,16,19] The reporting of cancer cases within a hospital is mandated by the hospital-based NCDB.[13] Reportable cancer diagnoses will originate from single and multi-institution cancer registries.[17] The fundamental purpose of the NCDB is to capture data designed to improve patient outcomes.[17] Evidence-based quality measures representing clinical best practice are reported from the NCDB through interactive benchmarking reports.[13] This includes the Rapid Cancer Reporting System (RCRS), a web-based tool designed to facilitate real-time reporting of cancer cases.[13]

While registrars who submit data to the NCDB are involved in all aspects of both the population-based registries and hospital-based registries, not all quality procedures performed by registrars pertain to the NCDB (Table 1). Quality procedures identified by Bray and Parkin that are only relevant to population-based cancer registries include assessment of age-specific curves, incidence rates of childhood cancers, mortality incidence ratio stability, number and average sources per case, and death certificate methods.[16] Death

certificate-only analyses are performed routinely across all registries.[16] Death certificate analysis as a quality indicator does not directly affect the NCDB. Other quality procedures are conducted after data submission and as part of data aggregation, quality assessment, and reporting.

The NCDB is part of a multi-agency, national cancer registry community in the USA that works collaboratively to ensure consistent, high quality cancer data can be applied across diverse utilities (Figure). This surveillance community comprises the central cancer registries, including the Centers for Disease Control and Prevention (CDC) National Program of Cancer Registries (NPCR) and the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (NCI); the National Cancer Registrars Association (NCRA); and the CoC.[18] The North American Association of Central Cancer Registries (NAACCR) is also part of this community, and it serves a vital role as a consensus organization.[16] NAACCR facilitates standardization of data definitions, abstraction and coding rules, quality procedures, and registrar certification, which in turn ensures uniform registry processes and establishes data quality standards.[16] Instructions to support standardized data definitions, abstraction, and coding rules, as well as quality procedures, are detailed in key manuals and documents. [16]

Completeness

Completeness, defined as a measure of representation, is the extent to which all the incident cancer cases occurring in the population are included in the registry. Case-finding procedures are considered critical to both cancer registry coverage and for survival accuracy. Completeness includes 9 quality procedures (Table 1).[3,4]

Due to the legislative mandate to report cancer cases to population-based cancer registries in the US, population-based cancer registries are regarded as the gold standard for data completeness.[16] We evaluated data completeness within the NCDB by comparing the number of cases of incident cancer cases from participating central registries included in the United States Cancer Statistics (USCS), the official federal cancer statistics.[19] These statistics include cancer registry data from CDC's National Program of Cancer Registries (NPCR) and the NCI SEER Program.[19] The USCS internal quality control file includes cases from all 50 states and the District of Columbia, providing information on demographic and tumor characteristics.[19] Cases diagnosed at a Veterans Affairs hospital were excluded from the NCDB analysis. Cases were further limited to malignant disease except for benign and borderline brain and other nervous system cancers, as well as the inclusion of female in situ breast cancers. Only male and female cancers with a diagnosis of cancer within the USA during 2016 to 2020 were included. The percentage of cancer cases captured within the NCDB from 2016 to 2020 were compared against prior reports which included diagnostic years 2012–2014.[14] Comparisons were made by primary disease site using the SEER definitions of the World Health Organization (WHO) International Classification of Diseases for Oncology, Third edition (ICD-O-3) site recodes.[20] Additional stratification included sex, diagnosis year, patient age, race/ethnicity, and the state of diagnosis corresponding to the patient's residence.

Outcomes for other measures of completeness that affect all registries (Table 1) have been previously reported.[21] Incidence case ascertainment for the NCDB is continuously verified with CoC special studies which are required for accreditation and specifically capture additional data on previously submitted cancer diagnoses. This provides an extra level of detail and audit of abstraction accuracy. Independent studies utilizing data from the NCDB have demonstrated case ascertainment compared to trials and claims data.[22–24] This type of auditing may be extended to assess registry completeness.

Comparability

Comparability is ensured by using standardized international guidelines on coding and classification procedures for cancer data abstraction.[3,4] Cancers reported to the NCDB are identified by the WHO ICD-O-3 topography, morphology, behavior, and grade codes.[25] The ICD-O-3 and topography and histology codes are categorized into cancer types.[10,26–28] Coding rules are maintained in registry manuals so that data items are abstracted and submitted to the registry with universal rules and codes.[10,26–28] Staging standards are defined by the American Joint Committee on Cancer.[29] The rules for coding include timing relative to initiation of treatment. Clinical staging is inclusive of the extent of cancer information before initiation of definitive treatment or within 4 months after the date of diagnosis, whichever is shorter.[29–30] Pathological staging includes any information obtained about the extent of cancer through completion of definitive surgery or within 4 months after the date of diagnosis, whichever is longer.[29–30] Secondary diagnosis codes are captured by the cancer registry as International Classification of Diseases, 10th Revision codes.[30] The CoC also requires registries to submit up to 10 comorbid conditions to the NCDB; these conditions influence the health status of the patient and treatment complications.[30]

SEER maintains an interactive drug database that facilitates the proper coding of treatment fields.[31] The rules for diagnostic confirmation require the reportability of both clinically diagnosed and microscopically confirmed tumors.[30] Clinically diagnosed tumors are based only on diagnostic imaging, laboratory tests, or other clinical examinations whereas microscopically confirmed tumors include all tumors with positive histopathology.[16,30] Cancer registries reference both “ambiguous terms at diagnosis” to determine case reportability and “ambiguous terms describing tumor spread” for staging purposes.[30] For reportability, the NCDB follows rules for class of case to describe the patient’s relationship to the facility. Rules exist for the reporting of multiple primaries to the NCDB.[32] These solid tumor rules are aimed at promoting consistent and standardized coding by cancer registrars and are intended to guide registrars through the process of determining the correct number of primaries.[32]

Timeliness

No international guidelines for cancer registry data submission timeliness exist, although the cancer surveillance community has specific timeliness standards for their respective registries.[16] Timeliness of NCDB data submission was assessed with compliance to CoC Standard 6.4 (Table 1).[13] There are 3 components to this standard. The first criterion assesses compliance with monthly data submissions of all new and updated cancer cases.[13] The second ensures all analytic cases are submitted to the NCDB’s annual call for data. [13]

The third requires hospitals to review at least twice each calendar year the quality measures performance rates, which are affected by timeliness of data submission.[13]

Validity

Validity is defined by Bray and Parkin as the proportion of cases in a dataset with a given characteristic that actually has that characteristic.[3,4] Data validity is maintained through procedures specific to quality control that are integral to the registry and tied to Standards 3.2, 4.3, 5.1, and 6.1 for CoC accreditation (Table 1).[13]

Accreditation for anatomic pathology by a qualifying organization is a component of Standard 3.2, designed to further structure quality assurance protocols.[13] Histological verification is also assessed in compliance with CoC Standard 3.2 and ensures each hospital provides diagnostic imaging services, radiation oncology services, and systemic therapy services on site with accreditation by a qualifying organization for anatomic pathology.[13]

Compliance with CoC Standard 4.3 is assessed for internal consistency, which ensures that all case abstraction is performed by tumor registrars who hold current certification by NCRA.[10,13] This ensures that registrars utilize, maintain, and continue their formal education through NCRA and thus continue working towards correctly interpreting and coding cancer diagnoses.[10,13] Educational assessment may additionally include participation in reliability studies designed to measure abstractor and coder compliance with existing coding rules and standards.[16] Reproducibility is a goal to assess the reliability study measures to help identify ambiguity or inadequacy of existing data definitions and rules as well as education needs.[16]

Standard 5.1 requires College of American Pathologists[33] synoptic reporting and for each hospital to perform an annual internal audit, confirming at least 90% of all cancer pathology reports are in synoptic format.[13] The synoptic format must be structured and include all core elements reported in a “diagnostic parameter pair” format.[13] Each diagnostic parameter pair must be listed together in synoptic format in 1 location in the pathology report.[13]

The database validity criteria for re-abstracting, recoding, and reliability procedures identified by Bray and Parkin are measured in compliance with CoC Standard 6.1. This requires the review of a minimum of 10% of cases each year and for CoC hospitals to establish a cancer registry quality control plan.[13] The re-abstracting and recoding auditing approaches involve data captured by the registry compared with data collected by a designated auditor.[16]

Data edits are electronic logical rules that evaluate internal consistency of values or data items.[16] For instance, a biological woman diagnosed with prostate cancer will fail edits. Edits are currently maintained by NAACCR based on edits originally developed by SEER.[34] NAACCR Edits’ Metafile comprises validation checks applied to cancer data.[34] CDC develops and maintains software (EditWriter and GenEDITS Plus) for registries to obtain edit reports on their cases using the standards maintained by NAACCR.[34,38] NCDB assigns scores that are applied to the call for data and RCRS reporting requirements, causing a case

to be rejected or accepted into either dataset.[35] An edit score of 200 will cause a record to be rejected from the NCDB.[35]

All data were analyzed using SAS version 9.4 (SAS Institute, Cary NC)[39] or SEER Surveillance Research Program, National Cancer Institute SEER*Stat software version 8.4.2.[40]

Results

The exclusion and inclusion criteria resulted in 9,269,442 cases from USCS and 6,828,507 cases from the NCDB. Compared to USCS, the official cancer statistics[36], the NCDB demonstrated 73.7% completeness of cancer cases diagnosed in the USA during 2016 to 2020 (Table 2). Among the top 10 major cancer sites, breast cancer, males and females, had the highest coverage at 81.9%, and the lowest was found for melanoma of the skin, males and females, at 52.0% (Table 2). Age group comparisons showed the lowest coverage (61.1%) for the 85 years and older age group, with the highest coverage for those aged 20-74 years (73.1%-80.4%). Race and ethnicity comparisons showed coverage for White patients to be 68.4%, Black patients 73.7%, American Indian/Alaskan Native patients 41.0%, Asian/Pacific Islander patients 70.7%, and Hispanic patients to be 56.4% (Table 3). Lastly, by state, Arkansas demonstrated the lowest coverage (24.0%), and North Dakota demonstrated the highest coverage (98.9%) (Table 4).

For timeliness, CoC Standard 6.4 was assessed with the requirement for timely data submission with compliance at 92.7% (Table 5).[13]

Validity was assessed with compliance for Standard 6.1, at 94.2% (Table 5). Compliance with histological verification standards were high, at 99.1% and CoC Standard 5.1 with pathological synoptic reporting at 93.6% (Table 5). Compliance with CoC Standard 4.3 was at 99.1% (Table 5). Edits checks at time of data submission are part of the NCDB validity criteria and covered in the Bray and Parkin criteria.[3] During the 2023 annual call for data, which began in March 2023, the NCDB processed 12,151,768 records consisting of 2021 diagnoses and follow-up resubmissions from prior years. Of the total, 71,854 cases failed the NCDB edits score representing <1%.

Discussion

The present study characterizes the NCDB data quality in all 4 domains defined by Bray and Parkin, including high rates of completeness, comparability, timeliness, and validity.[3,4] Coordination with the cancer surveillance community as demonstrated in the figure lead to standardization of abstraction practice with universal coding definitions. The CoC accreditation standards layered an additional component to quality assurance with regards to histologic verification, registry staff credentials, synoptic reports, and including submission timeliness. Altogether nearly all framework that applies to the hospital-based NCDB, identified by Bray and Parkin criteria, is maintained with results indicative of consistency and stability over time.

The CoC Standards for data quality that we examined are associated with high compliance and are a necessary component to maintain accreditation by the CoC. Cancer hospitals of the CoC are diverse by region, patient case mix, and volume and yet still display unified adherence to compliance with metrics designed to promote high quality of data. Many of the countries that previously reported on national registry data quality have universal healthcare coverage with a single or 2-tiered national provider.[5,6] Norway has an 11-digit personal identification assigned to all newborns and people residing in the country.[5] In contrast, the USA has a complex system of insurance options and eligibility criteria that patients navigate on their own or through their employer. In the USA there is no national patient identifier, and the gathering of cancer data could be further complicated by the variability in electronic health record systems, which may not be interoperable. Hospitals are required to follow standard processes and procedures to abstract and report data to NCDB, including treatment information, and are therefore a valuable resource for evaluating cancer treatment patterns. While central registries capture treatment information, this varies by state and therefore is not routinely available in public-facing NPCR and SEER data.

There are limitations to note. First, the NCDB does not capture data beyond those hospitals accredited by the CoC. There are approximately 6,000 hospitals in the US[37] with variable definitions and practices. Through this study, we determined that NCDB captures 73.7% of cancer patients in the USA when compared with national data. A second limitation is that the NCDB does not collect direct patient identifiers, including name and Hispanic ethnicity which are under-ascertained in the NCDB. Name is necessary to run the NAACCR algorithm used by population-based registries to identify Hispanic identity. Finally, NCDB is not designed to assess changes in clinical practices or quality of care in real time, although with the launch of RCRS more timely evaluation of sudden changes in cancer care and outcomes, such as those that occurred during the first months of the COVID-19 pandemic, is increasingly feasible. Mandatory concurrent data abstraction rules are in place and required of accredited hospitals of the CoC. Data submission rules are presently in place that require all new and updated cancer cases to be submitted monthly.[13] Additional progress with timeliness is expected as the Standards for concurrent abstraction are adjusted to include the diagnostic and first treatment phase of care. We have planned future studies to evaluate the completeness, comparability, validity, and timeliness of RCRS data and the feasibility of using real-time data in research.

Advances in cancer control are information dependent. As new data sources and analytic platforms become available, it is imperative that data quality be considered alongside data availability to ensure information validity and reliability. The data quality standards described in this paper and adhered to by NCDB facilitate reporting to hospital administration personnel for decision making, researchers and epidemiologists, quality analysts, and to governments that mandate reporting of cancer. Registry data must be comprehensive, granular, and valid. High quality data allows the use of NCDB, during the CoC accreditation process, to include reports on quality-of-care measures and patient outcomes assessments. The NCDB provides a comprehensive view of cancer care in the USA within CoC-accredited hospitals.

Disclaimer

The findings and conclusions in this article are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

References

1. Pollock BH. Quality assurance for interventions in clinical trials. Multicenter data monitoring, data management, and analysis. Cancer. 1994;74(9 Suppl):2647–2652. [PubMed: 7954280]
2. Menard T, Barmaz Y, Koneswarakantha B, Bowling R, Popko L. Enabling data-driven clinical quality assurance: predicting adverse event reporting in clinical trials using machine learning. Drug Saf. 2019;42(9):1045–1053. [PubMed: 31123940]
3. Bray F and Parkin DM, Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. Eur J Cancer. 2009;45(5):747–755. [PubMed: 19117750]
4. Parkin DM and Bray F. Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. Eur J Cancer. 2009;45(5):756–764. [PubMed: 19128954]
5. Larsen IK, Smastuen M, Johannesen TB, et al. Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. Eur J Cancer. 2009;45(7):1218–1231. [PubMed: 19091545]
6. Dimitrova N, Parkin DM. Data quality at the Bulgarian National Cancer Registry: an overview of comparability, completeness, validity and timeliness. Cancer Epidemiol. 2015;39(3):405–413. [PubMed: 25914137]
7. Wanner M, Matthes KL, Korol S, Dehler S, Rohrmann S. Indicators of data quality at the Cancer Registry Zurich and Zug in Switzerland. Biomed Res Int. Published 2018 Jun 13.
8. SEER Cancer Registration and Training Modules Accessed November 13, 2023. https://training.seer.cancer.gov/registration/types/hospital.html .
9. Cancer Surveillance Programs. Accessed July 17, 2023. https://www.cancer.org/cancer/managing-cancer/making-treatment-decisions/clinical-trials/cancer-surveillance-programs-and-registries-in-the-united-states.html#:~:text=The%20National%20Cancer%20Institute's%20(NCI,Survival.
10. National Cancer Registrars Association (NCRA) Cancer Registry Manual Principles and Practices Fourth Edition. The CTR Certification. Accessed August 22, 2023. https://www.ncra-usa.org/CTR .
11. National Cancer Database. Accessed November 14, 2023. https://www.facs.org/quality-programs/cancer-programs/national-cancer-database/about/.
12. Lerro CC, Robbins AS, Phillips JL, Stewart AK. Comparison of cases captured in the National Cancer Data Base with those in population-based Central Cancer Registries. Ann Surg Oncol. 2013;20:1759–65. [PubMed: 23475400]
13. Optimal Resources for Cancer Care, 2020 Standards. Accessed July 17, 2023. https://acs.wufoo.com/forms/download-optimal-resources-for-cancer-care/ .
14. Mallin K, Browner A, Palis B, et al. Incident cases captured in the National Cancer Database Compared with those in U.S. Population Based Central Cancer Registries in 2012–2014. Ann Surg Oncol. 2019;26(6):1604–1612. [PubMed: 30737668]
15. Duggan MA, Anderson WF, Altekruse S, Penberthy L, Sherman ME. The Surveillance, Epidemiology, and End Results (SEER) program and pathology: toward strengthening the critical relationship. Am J Surg Pathol. 2016;40(12):e94–e102. [PubMed: 27740970]
16. North American Association of Central Cancer Registries (NAACCR) Standards for Cancer Registries Volume III: Standards for Completeness, Quality, Analysis, Management, Security, and Confidentiality of Data. Accessed July 17, 2023. https://www.naaccr.org/wp-content/uploads/2016/11/Standards-for-Completeness-Quality-Analysis-Management-Security-and-Confidentiality-of-Data-August-2008PDF.pdf .
17. Bilimoria KY, Stewart AK, Winchester DP, Ko CY. The National Cancer Data Base: a powerful initiative to improve cancer care in the United States. Ann Surg Oncol. 2008;15(3):683–90. [PubMed: 18183467]

18. SEER Cancer Registration and Training Modules. Accessed November 14, 2023. https://training.seer.cancer.gov/operations/standards/setters/ .
19. U.S. Cancer Statistics Working Group. U.S. Cancer Statistics Data Visualizations Tool, based on 2022 submission data (1999–2020): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; Accessed November 14, 2023. https://www.cdc.gov/cancer/dataviz .
20. SEER Site Recodes https://seer.cancer.gov/siterecode/ Accessed November 14th, 2023.
21. Islami F, Ward EM, Sung H, et al. Annual Report to the Nation on the Status of Cancer, Part 1: National Cancer Statistics. J Natl Cancer Inst. 2021;113(12):1648–1669. [PubMed: 34240195]
22. Shi Q, You YN, Nelson H, et al. Cancer registries: a novel alternative to long-term clinical trial follow-up based on results of a comparative study. Clinical Trials. 2010;7(6):686–695. Doi:10.1177/1740774510380953. [PubMed: 20729254]
23. Mallin K, Palis BE, Watroba N, et al. Completeness of American Cancer Registry Treatment data: implications for quality of care research. J Am Coll Surg. 2013;216(3):428–37. [PubMed: 23357724]
24. Lin CC, Virgo KS, Robbins AS, Jemal A, Ward EM. Comparison of comorbid medical conditions in the National Cancer Database and the SEER-Medicare Database. Ann Surg Oncol. 2016;23(13):4139–4148. [PubMed: 27535406]
25. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin LH, et al. International Classification Of Diseases For Oncology, 3rd ed. World Health Organization; 2000.
26. Surveillance and Epidemiology End Result Program Coding and Staging Manual 2023. Accessed July 17, 2023. https://seer.cancer.gov/tools/codingmanuals/ .
27. North American Association of Central Cancer Registries (NAACCR) SSDI/Grade Manual. Accessed July 17, 2023. https://www.naaccr.org/data-standards-data-dictionary/ .
28. North American Association of Central Cancer Registries (NAACCR) Data Standards and Data Dictionary. Accessed July 17, 2023. https://www.naaccr.org/data-standards-data-dictionary/ .
29. Amin MB, Edge S, Greene F, Byrd DR, Brookland RK, Washington MK, et al. (Eds.) AJCC Cancer Staging Manual (8th edition). Springer International Publishing: American Joint Commission on Cancer; 2017.
30. Standards for Oncology Registry Entry (STORE 2022). Accessed July 17, 2023. https://www.facs.org/media/weqje4pk/store-2022-12102021-final.pdf .
31. Surveillance and Epidemiology End Result (SEER) *Rx Drug Database. Accessed July 17, 2023. https://seer.cancer.gov/seertools/seerrx/ .
32. Surveillance and Epidemiology End Result (SEER) Solid Tumor Rules. Accessed July 17, 2023. https://seer.cancer.gov/tools/codingmanuals/historical.html#solid .
33. CAP Protocol Templates Accessed November 14, 2023. https://www.cap.org/protocols-and-guidelines/cancer-reporting-tools/cancer-protocol-templates .
34. North American Association of Central Cancer Registries (NAACCR) Standards for Cancer Registries, Standard Data Edits, Volume IV. Accessed July 17, 2023. https://www.naaccr.org/standard-data-edits/ .
35. American College of Surgeons v22b and v23b NCDB/RCRS Edits and RCRS Data Submission Requirements. 2023. Accessed July 17, 2023. https://www.facs.org/media/qw5bqvew/v22b-and-v23b-ncdb-rcrs-edits-and-rcrs-data-submission-requirements-07112023.pdf .
36. About US Cancer Statistics. November 14, 2023. https://www.cdc.gov/cancer/uscs/about/index.htm Accessed.
37. Fast Facts on US Hospitals. 2023. Accessed September 21, 2023. https://www.aha.org/statistics/fast-facts-us-hospitals .
38. Centers for Disease Control and Prevention National Program of Cancer Registries Edits. 2023. Accessed December 14, 2023. https://www.cdc.gov/cancer/npcr/tools/edits/index.htm.
39. SAS Academic Programs. 2023. Accessed December 14, 2023. https://sas.com.
40. Surveillance, Epidemiology, and End Results Program SEER*Stat Software. 2023. SEER*Stat Software (cancer.gov). Accessed December 14, 2023.

Synopsis

The National Cancer Database conforms to the standardized framework developed by Bray and Parkin in 2009 for registry and data quality. The data are demonstrated to be of high quality for clinical and policy decision making at the national level.

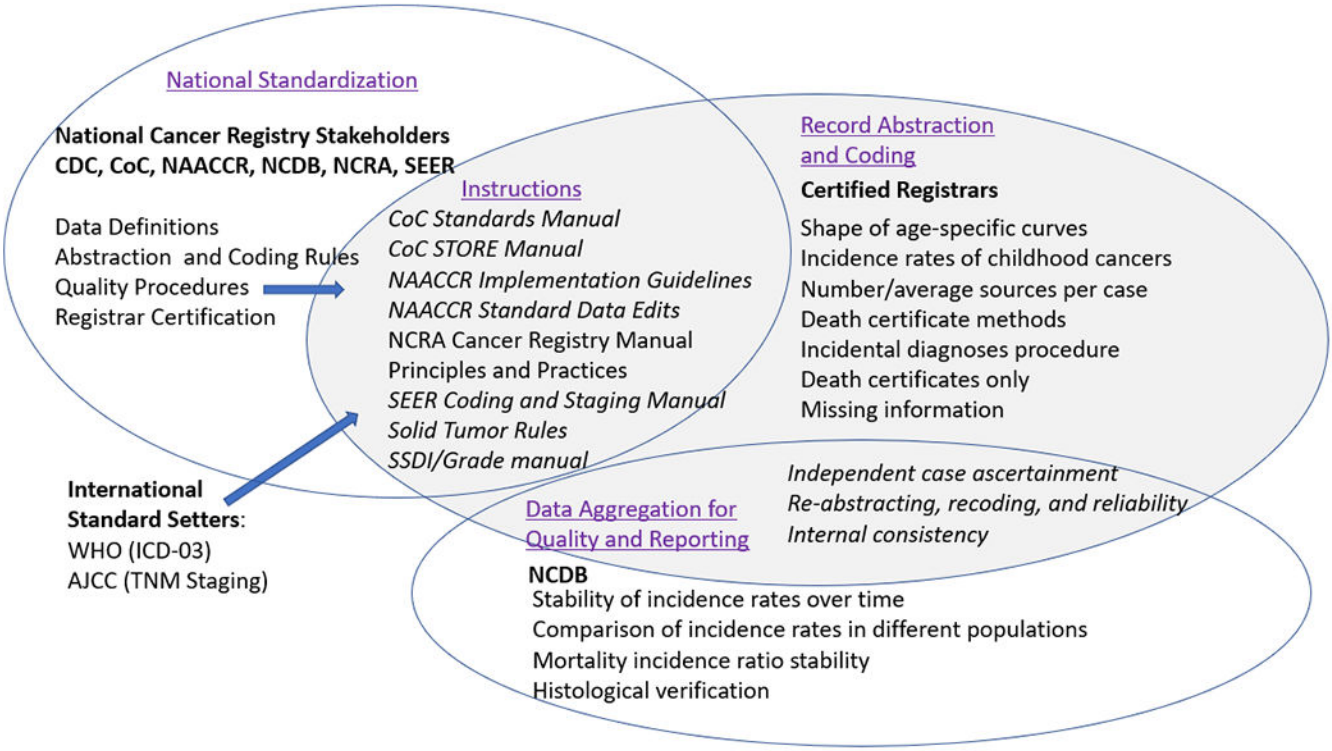

Figure 1. National Cancer Registry Quality Processes*

*The quality of cancer data in the United States is supported by a large, multi-agency, national cancer registry stakeholders community in the United States that works collaboratively to ensure consistent, high quality cancer data that can be applied across diverse utilities. These national cancer registry stakeholders standardize cancer data definitions, abstraction and coding rules, registry-based quality procedures as well as registrar education, training and certification. These national standards are monitored at the hospital level through compliance with quality procedures during the record abstraction and coding process as well at the national level during the process of data aggregation for quality and reporting.

AJCC American Joint Committee on Cancer
CDC Centers for Disease Control and Prevention
CoC Commission on Cancer
NAACCR North American Association of Central Registries, Inc.
NCDB National Cancer Data Base
NCRA National Cancer Registrars Association
SEER Surveillance, Epidemiology, and End Results Program
STORE Standards for Oncology Registry Entry
SSDI Site-Specific Data Item
WHO World Health Organization

Table 1.

Assessment of NCDB registry and data quality according to Bray and Parkin criteria[1]

Bray and Parkin Criteria		Application to NCDB[2]	Data Quality Mechanism	Registry and Data Quality Category
Completeness				
Historic data methods	Stability of incidence rates over time	Yes	Annual NCDB Warehouse quality assurance check; Annual Benchmarking Report trends	Quality Procedure
	Comparison of incidence rates in different populations	Yes	NCDB coverage to USCS	Quality Procedure
	Shape of age-specific curves	No	NAACCR Standards for Cancer Registries Volume III (age specific/adjusted incidence rates)	Quality Procedure
	Incidence rates of childhood cancers	No	NAACCR Standards for Cancer Registries Volume III (Percent Incidence Ratio)	Quality Procedure
Mortality incidence ratio stability		No	NAACCR Standards for Cancer Registries Volume III	Quality Procedure
Number/average sources per case		No	NAACCR Standards for Cancer Registries Volume III (sources for reporting)	Quality Procedure
Histological verification		Yes	NAACCR Standards for Cancer Registries Volume III	Quality Procedure
Independent case ascertainment		Yes	CoC Special Studies NAACCR Standards for Cancer Registries Volume III (NAACCR abstraction and recoding reliability studies and audits) NCRA Cancer Registry Manual Principles and Practices, 4th Edition	Quality Procedure
Death certificate methods		No	NAACCR Standards for Cancer Registries Volume III (Percent DCO) NAACCR Standards for Cancer Registries Volume III (Death Clearance Follow-Back)	Quality Procedure
Comparability				
Identification	Topography	Yes	WHO ICD-O-3 (C00.0-80.9)	Standardized Data Definition
	Histology	Yes	WHO ICD-O-3 (8000-9993)	
	Behavior	Yes	WHO ICD-O-3 (0-3)	
	Grade	Yes	NAACCR SSDI/grade manual WHO ICD-O-3	
Stage		Yes	AJCC Staging standards	Standardized Data Definition
Secondary diagnosis		Yes	WHO ICD-10	Standardized Data Definition
Standard coding schema		Yes	CoC STORE manual NAACCR SSDI/Grade manual NAACCR Data Standards and Data Dictionary SEER Coding and Staging Manual 2023 SEER Drug Database	Abstraction and Coding Rules
Definition of incidence (case and date)		Yes	NAACCR Standards for Cancer Registries Volume III (diagnostic confirmation, class of case, type of submission, ambiguous terminology) CoC Store manual	Abstraction and Coding Rules
Primary cancer (new case) rules		Yes	Solid Tumor Rules (collaborative product of CDC, NAACCR, SEER, and Central Registries)	Abstraction and Coding Rules

Bray and Parkin Criteria	Application to NCDB[2]	Data Quality Mechanism	Registry and Data Quality Category
Timeliness			
Abstraction and submission timeliness	Yes	NAACCR Standards for Cancer Registries Volume III CoC Standards Manual, 6.4	Abstraction and Coding Rules
Validity			
Re-abstracting, recoding, and reliability	Yes	CoC Standards Manual 6.1 (review of 10% analytic caseload annually) NAACCR Standards for Cancer Registries Volume III (QA process controls, special assessments, re-abstraction audits, recoding audits, reliability studies)	Quality Procedure
Histological verification	Yes	CoC Standards Manual, 3.2 and 5.1 (accreditation for anatomic pathology, internal audit of 90% of pathology reports annually)	Quality Procedure
Death certificate only	No	NAACCR Standards for Cancer Registries Volume III (DCO validity) SEER Coding and Staging Manual 2023	Quality Procedure
Reviews missing information	Yes	NAACCR Standards for Cancer Registries Volume III (edits, process controls for unknown values) Requirements NAACCR Standard Edits for Cancer Registry Volume IV	Quality Procedure
Reviews internal consistency	Yes	CoC Standards Manual 4.3 (cancer registry staff credentials) NAACCR Standards for Cancer Registries Volume III (QA standards, QA staffing guidelines, QA procedures, staff credentials) NCRA Cancer Registry Manual Principles and Practices, 4th Edition V22B and V23B NCDB/RCRS Edits and Submission	Quality Procedure

[1]AJCC American Joint Committee on Cancer

CDC Centers for Disease Control and Prevention

CoC Commission on Cancer

NAACCR North American Association of Central Registries, Inc.

NCDB National Cancer Database

NCRA National Cancer Registrars Association

SEER Surveillance, Epidemiology, and End Results Program

STORE Standards for Oncology Registry Entry

SSDI Site-Specific Data Item

USCS United States Cancer Statistics

WHO World Health Organization

[2]Procedures followed by all registrars for purposes of reporting to population-based registries that may not directly impact reporting to the NCDB

Table 2.

Comparison of incidence for completeness by disease sites in 2016-2020[1]

	USCS Count	NCDB Count	Coverage %	USCS Count, Male	NCDB Count, Male	Coverage %	USCS Count, Female	NCDB Count, Female	Coverage %
Total [2]	9,269,442	6,828,507	73.7	4,522,387	3,142,113	69.5	4,747,055	3,686,394	77.7
Oral Cavity and Pharynx	239,509	188,806	78.8	171,188	134,296	78.4	68,321	54,510	79.8
Lip	9,231	5,192	56.2	6,529	3,724	57.0	2,702	1,468	54.3
Tongue	77,495	62,151	80.2	55,989	44,717	79.9	21,506	17,434	81.1
Salivary Gland	24,196	18,421	76.1	14,148	10,414	73.6	10,048	8,007	79.7
Floor of Mouth	9,601	8,144	84.8	6,473	5,454	84.3	3,128	2,690	86.0
Gum and Other Mouth	31,711	26,133	82.4	17,751	14,617	82.3	13,960	11,516	82.5
Nasopharynx	9,606	7,278	75.8	6,808	5,123	75.2	2,798	2,155	77.0
Tonsil	46,527	37,638	80.9	38,815	31,401	80.9	7,712	6,237	80.9
Oropharynx	15,298	12,401	81.1	12,230	9,902	81.0	3,068	2,499	81.5
Hypopharynx	11,255	9,217	81.9	8,949	7,286	81.4	2,306	1,931	83.7
Other Oral Cavity and Pharynx	4,589	2,231	48.6	3,496	1,658	47.4	1,093	573	52.4
Digestive System	1,549,130	1,169,589	75.5	867,417	651,185	75.1	681,713	518,404	76.0
Esophagus	92,634	71,329	77.0	73,239	56,134	76.6	19,395	15,195	78.3
Stomach	122,455	92,974	75.9	75,013	57,554	76.7	47,442	35,420	74.7
Small Intestine	49,807	39,377	79.1	26,716	21,047	78.8	23,091	18,330	79.4
Colon and Rectum	711,415	527,686	74.2	375,758	277,230	73.8	335,657	250,456	74.6
Colon excluding Rectum	502,914	366,984	73.0	252,651	182,490	72.2	250,263	184,494	73.7
Rectum and Rectosigmoid Junction	208,501	160,702	77.1	123,107	94,740	77.0	85,394	65,962	77.2
Anus, Anal Canal and Anorectum	39,893	32,411	81.2	13,778	11,061	80.3	26,115	21,350	81.8
Liver and Intrahepatic Bile Duct	179,172	131,386	73.3	126,466	92,104	72.8	52,706	39,282	74.5
Gallbladder	21,348	16,380	76.7	7,009	5,305	75.7	14,339	11,075	77.2
Other Biliary	33,101	28,506	86.1	18,280	15,814	86.5	14,821	12,692	85.6
Pancreas	267,894	204,543	76.4	139,094	105,999	76.2	128,800	98,544	76.5
Retroperitoneum	7,771	6,859	88.3	3,968	3,448	86.9	3,803	3,411	89.7
Peritoneum, Omentum and Mesentery	9,430	8,398	89.1	866	669	77.3	8,564	7,729	90.2

	USCS Count	NCDB Count	Coverage %	USCS Count, Male	NCDB Count, Male	Coverage %	USCS Count, Female	NCDB Count, Female	Coverage %
Other Digestive Organs	14,210	9,740	68.5	7,230	4,820	66.7	6,980	4,920	70.5
Respiratory System	1,189,661	903,630	76.0	627,383	467,371	74.5	562,278	436,259	77.6
Nose, Nasal Cavity and Middle Ear	12,771	11,010	86.2	7,795	6,691	85.8	4,976	4,319	86.8
Larynx	61,328	47,936	78.2	48,699	37,680	77.4	12,629	10,256	81.2
Lung and Bronchus	1,111,987	841,895	75.7	568,510	421,113	74.1	543,477	420,782	77.4
Pleura	489	362	74.0	273	207	75.8	216	155	71.8
Trachea, Mediastinum and Other Respiratory Organs	3,086	2,427	78.6	2,106	1,680	79.8	980	747	76.2
Bones and Joints	17,176	14,054	81.8	9,671	7,986	82.6	7,505	6,068	80.9
Soft Tissue including Heart	60,381	50,436	83.5	33,745	27,929	82.8	26,636	22,507	84.5
Skin excluding Basal and Squamous	463,759	245,084	52.8	274,878	144,835	52.7	188,881	100,249	53.1
Melanoma of the Skin	430,808	224,051	52.0	254,565	132,045	51.9	176,243	92,006	52.2
Other Non-Epithelial Skin	32,951	21,033	63.8	20,313	12,790	63.0	12,638	8,243	65.2
Breast, In situ	NA						283,751	233,502	82.3
Breast, Malignant	1,294,951	1,060,064	81.9	11,236	9,735	86.6	1,283,715	1,050,329	81.8
Female Genital System	NA						514,641	432,279	83.9
Cervix Uteri	NA						64,810	52,943	81.7
Corpus and Uterus, NOS	NA						292,506	247,649	84.7
Ovary	NA						102,157	84,872	83.1
Vagina	NA						6,784	5,170	76.2
Vulva	NA						27,782	22,834	82.2
Other Female Genital Organs	NA						20,602	18,811	91.3
Male Genital System	NA			1,146,461	704,569	61.5	NA		
Prostate	NA			1,091,626	665,462	61.0	NA		
Testis	NA			45,227	32,402	71.6	NA		
Penis	NA			7,592	5,439	71.6	NA		
Other Male Genital Organs	NA			2,016	1,266	62.8	NA		
Urinary System	736,493	545,604	74.1	517,044	379,579	73.4	219,449	166,025	75.7

	USCS Count	NCDB Count	Coverage %	USCS Count, Male	NCDB Count, Male	Coverage %	USCS Count, Female	NCDB Count, Female	Coverage %
Urinary Bladder	381,247	266,866	70.0	290,764	202,408	69.6	90,483	64,458	71.2
Kidney and Renal Pelvis	337,171	264,252	78.4	214,718	167,923	78.2	122,453	96,329	78.7
Ureter	10,720	8,935	83.3	6,604	5,561	84.2	4,116	3,374	82.0
Other Urinary Organs	7,355	5,551	75.5	4,958	3,687	74.4	2,397	1,864	77.8
Eye and Orbit	15,541	11,901	76.6	8,360	6,336	75.8	7,181	5,565	77.5
Brain and Other Nervous System Benign	224,893	173,036	76.9	69,734	53,856	77.2	155,159	119,180	76.8
Brain, Benign	10,829	8,112	74.9	5,081	3,864	76.0	5,748	4,248	73.9
Cranial Nerves Other Nervous System, Benign	214,064	164,924	77.0	64,653	49,992	77.3	149,411	114,932	76.9
Brain and Other Nervous System Borderline	23,444	17,652	75.3	11,363	8,511	74.9	12,081	9,141	75.7
Brain, Borderline	10,831	7,515	69.4	5,851	4,088	69.9	4,980	3,427	68.8
Cranial Nerves Other Nervous System, Borderline	12,613	10,137	80.4	5,512	4,423	80.2	7,101	5,714	80.5
Brain and Other Nervous System Malignant	116,569	100,037	85.8	65,525	56,741	86.6	51,044	43,296	84.8
Brain, Malignant	110,062	95,140	86.4	62,282	54,256	87.1	47,780	40,884	85.6
Cranial Nerves Other Nervous System, Malignant	6,507	4,897	75.3	3,243	2,485	76.6	3,264	2,412	73.9
Endocrine System	243,327	196,182	80.6	68,677	55,919	81.4	174,650	140,263	80.3
Thyroid	228,738	184,589	80.7	61,039	49,845	81.7	167,699	134,744	80.3
Other Endocrine including Thymus	14,589	11,593	79.5	7,638	6,074	79.5	6,951	5,519	79.4
Lymphoma	404,391	285,779	70.7	223,341	156,882	70.2	181,050	128,897	71.2
Hodgkin Lymphoma	42,843	33,108	77.3	23,562	18,130	76.9	19,281	14,978	77.7
Non-Hodgkin Lymphoma	361,548	252,671	69.9	199,779	138,752	69.5	161,769	113,919	70.4
Myeloma	140,054	100,911	72.1	77,923	55,985	71.8	62,131	44,926	72.3
Leukemia	264,670	173,955	65.7	154,654	101,013	65.3	110,016	72,942	66.3
Lymphocytic Leukemia	127,298	76,132	59.8	77,380	46,338	59.9	49,918	29,794	59.7

	USCS Count	NCDB Count	Coverage %	USCS Count, Male	NCDB Count, Male	Coverage %	USCS Count, Female	NCDB Count, Female	Coverage %
Myeloid and Monocytic Leukemia	122,520	90,476	73.8	69,220	50,670	73.2	53,300	39,806	74.7
Other Leukemia	14,852	7,347	49.5	8,054	4,005	49.7	6,798	3,342	49.2
Mesothelioma	15,187	12,046	79.3	11,136	8,670	77.9	4,051	3,376	83.3
Kaposi Sarcoma	5,330	3,318	62.3	4,821	3,056	63.4	509	262	51.5

[1] https://seer.cancer.gov/siterecode/icdo3_dwhoheme/index.html

[2] Totals include all breast disease both male and female, miscellaneous primaries, and invalid primaries not defined in the SEER site recode ICD-O 3/WHO 2008 definitions not shown in the table

Table 3.

Comparison of incidence for completeness by patient demographics in 2016-2020

	USCS count	NCDB count	Case coverage %
Diagnosis Year			
2016	1,835,671	1,340,154	73.0
2017	1,868,195	1,371,180	73.4
2018	1,888,798	1,389,910	73.6
2019	1,931,814	1,430,765	74.1
2020	1,744,964	1,296,498	74.3
Age group (years)			
0-19	84,061	56,090	66.7
20-44	661,256	531,721	80.4
45-54	1,051,339	837,344	79.6
55-64	2,234,851	1,714,153	76.7
65-74	2,801,072	2,047,766	73.1
75-84	1,752,985	1,223,798	69.8
≥ 85	683,878	417,635	61.1
Pediatric, Young Adult age groups (years)			
0-14	56,416	35,642	63.2
15-29	143,796	113,376	78.8
30-39	286,235	229,475	80.2
Race/ethnicity [1]			
White	7,673,661	5,252,315	68.4
Black	1,036,310	763,280	73.7
American Indian/Alaskan Native	59,068	24,224	41.0
Asian/Pacific Islander	336,216	237,810	70.7
Hispanic[2]	786,254	443,101	56.4

[1] White, Black, American Indian/Alaskan Native, and Asian/Pacific Islander is shown regardless of Hispanic origin

[2] Due to Hispanic origin misclassification, data for North Dakota and Wisconsin may be underestimated for any Hispanic race groups and overestimated for any Non-Hispanic race groups

Table 4.

Comparison of incidence for completeness by patient state for all cancer sites in 2016-2020

	USCS count	NCDB count	Case coverage %
Alabama	142,136	92,044	64.8
Alaska	16,534	8,493	51.4
Arizona	178,632	43,284	24.2
Arkansas	92,417	22,163	24.0
California	925,531	545,472	58.9
Colorado	133,685	106,901	80.0
Connecticut	113,707	109,051	95.9
Delaware	31,314	29,273	93.5
District of Columbia	15,210	12,162	80.0
Florida	719,491	440,952	61.3
Georgia	288,885	235,055	81.4
Hawaii	40,440	31,987	79.1
Idaho	48,273	31,760	65.8
Illinois	373,086	320,209	85.8
Indiana*	184,281	166,420	90.3
Iowa	101,525	71,108	70.0
Kansas	82,434	53,838	65.3
Kentucky	147,448	125,915	85.4
Louisiana	140,097	101,811	72.7
Maine	48,473	40,083	82.7
Maryland	173,825	137,654	79.2
Massachusetts	204,835	160,318	78.3
Michigan	295,481	230,478	78.0
Minnesota	168,322	134,805	80.1
Mississippi	88,204	66,443	75.3
Missouri	182,992	153,295	83.8
Montana	33,977	27,115	79.8
Nebraska	54,526	45,279	83.0
Nevada *	73,340	27,130	37.0
New Hampshire	46,420	39,366	84.8
New Jersey	286,034	246,754	86.3
New Mexico	50,510	26,833	53.1
New York	617,261	441,331	71.5
North Carolina	314,527	257,235	81.8
North Dakota	20,603	20,376	98.9
Ohio	362,198	323,061	89.2

	USCS count	NCDB count	Case coverage %
Oklahoma	107,891	67,105	62.2
Oregon	117,334	88,899	75.8
Pennsylvania	422,345	356,727	84.5
Rhode Island	33,528	28,437	84.8
South Carolina	149,771	115,465	77.1
South Dakota	25,878	18,686	72.2
Tennessee	202,099	165,499	81.9
Texas	641,500	409,066	63.8
Utah	63,052	43,483	69.0
Vermont	20,649	211,524	87.6
Virginia	220,387	18,093	96.0
Washington	206,138	169,238	82.1
West Virginia	63,733	52,584	82.5
Wisconsin	183,331	151,654	82.7
Wyoming	15,152	6,593	43.5

* These states did not meet the requirements for USCS Publication criteria for diagnosis year 2020

Table 5.

Program Compliance with Commision on Cancer Data and Registry Quality Accreditaiton Standards; Based on Commission on Cancer Accreditation Site Visits 2022 (N=329 programs)

Validity and Timeliness Quality Standards	CoC Program Compliance, N (%)
Histological verification for validity	
Standard 3.2; Evaluation of Treatment Services[1]	326/329 (99.1%)
Standard 5.1; College of American Pathologists Synoptic Reporting	308/329 (93.6%)
Reviews internal consistency for validity	
Standard 4.3; Cancer Registry Staff Credentials	326/329 (99.1%)
Re-abstracting, recoding, and reliability for validity	
Standard 6.1; Cancer Registry Quality Control	310/329 (94.2%)
Abstraction and submission timeliness	
Standard 6.4; Rapid Cancer Reporting System: Data Submission[2]	283/305 (92.7%)

[1] Accreditation for anatomic pathology by a qualifying organization.

[2] Newly accredited hospitals are not rated on Standard 6.4 until their first reaccreditation visit resulting in the discrepant N