



Published in final edited form as:

J Registry Manag. 2014 ; 41(2): 65–97.

Coding Completeness and Quality of Relative Survival-Related Variables in the National Program of Cancer Registries Cancer Surveillance System 1995–2008

R. Wilson, MPH^a, M.E. O’Neil, MPH^a, E. Ntekop, MD MPH^b, K. Zhang, PhD^b, and Y. Ren, PhD^b

^aDivision of Cancer Prevention and Control, National Center for Chronic Disease Prevention and Health Promotion, Coordinating Center for Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA 30333, U.S.A.

^bICF International, Fairfax, VA 22031, U.S.A.

Abstract

Background—Calculating accurate estimates of cancer survival are important for various analyses of cancer patient care and prognosis. Current U.S. survival rates are estimated based on data from the National Cancer Institute’s (NCI’s) Surveillance, Epidemiology, and End Results (SEER) program, covering approximately 28% of the U.S. population. The National Program of Cancer Registries (NPCR) covers about 96% of the U.S. population. Using a population-based database with greater U.S. population coverage to calculate survival rates at the national, state, and regional levels can further enhance the effective monitoring of cancer patient care and prognosis in the U.S. The first step is to establish the coding completeness and coding quality of the NPCR data needed for calculating survival rates and conducting related validation analyses.

Methods—Using data from the NPCR-Cancer Surveillance System (CSS) from 1995 through 2008, we assessed coding completeness and quality on 26 data elements that are needed to calculate cancer relative survival estimates and conduct related analyses. Data elements evaluated consisted of demographic, follow-up, prognostic, and cancer identification variables. Analyses showing trends of these variables by diagnostic year, state of residence at diagnosis, and cancer site were performed.

Results—Mean overall percent coding completeness by each NPCR central cancer registry averaged across all data elements and diagnosis years ranged from 92.3% to 100%. Results showing the mean percent coding completeness for the relative survival-related variables in NPCR data are presented. All data elements but one have a mean coding completeness greater than 90% as was the mean completeness by data item group type. Statistically significant differences in coding completeness were found in the ICD revision number, cause of death, vital status, and date of last contact variables when comparing diagnosis years. The majority of data items had a coding quality greater than 90%, with exceptions found in cause of death, follow-up source, and the Surveillance, Epidemiology, and End Results (SEER) Summary Stage 1977, and SEER Summary Stage 2000.

*Correspondence to: Reda Wilson, 4770 Buford Highway MS F-76, Atlanta, GA 30341, U.S.A. dfo8@cdc.gov.

Conclusion—Percent coding completeness and quality are very high for variables in the NPCR-CSS that are covariates to calculating relative survival. NPCR provides the opportunity to calculate relative survival that may be more generalizable to the U.S. population.

Keywords

Data quality; relative survival rates; cancer; National Program of Cancer Registries

Background

Estimation of cancer survival is an important part of assessing the overall strength of cancer care and success of prevention programs. Relative survival is a measure that can be used to describe the survival of a cohort of cancer patients by removing the effect of competing death events of a comparable general population. The measure is the ratio of observed survival among cancer patients divided by the expected survival of the general population that is comparable to the cancer patients with respect to covariates including age, sex, and year of diagnosis.^{1,2} Population-based cancer relative survival rates are important for medical and public health efforts, including measuring the survivorship of cancer patients after diagnosis and monitoring the impact of intervention and early detection programs.³

Current U.S. cancer survival rates are estimated based on data from the National Cancer Institute's (NCI's) Surveillance, Epidemiology, and End Results (SEER) program, which covers approximately 28% of the U.S. population.⁴ The National Program of Cancer Registries (NPCR), established by Congress in 1992 and administered by the Centers for Disease Control and Prevention (CDC), is conducted in 45 states, the District of Columbia, Puerto Rico, and the Pacific Island Jurisdictions, and covers approximately 96% of the U.S. population. Data submitted annually to the NPCR-Cancer Surveillance System (NPCR-CSS) may also be used to calculate survival rates and provide greater coverage at national, regional, and state levels, so clinicians, public health practitioners, and researchers can effectively monitor cancer patient care and prognosis in the U.S.

NPCR-CSS collects data on the occurrence of cancer including the type, extent, and anatomic location of the cancer and the type of initial treatment by providing funding and technical assistance to the central cancer registries (CCRs) within the program.⁵ Population-based CCRs are data systems that collect, manage, and analyze data about cancer cases. In each state, medical facilities (including hospitals, physicians' offices, therapeutic radiation facilities, free-standing surgical centers, and pathology laboratories) are required to report demographic and clinically-related data to their central cancer registry. Each year, CDC supports efforts to link registry data with the National Death Index (NDI), Indian Health Service data, and state vital records, receives data from NPCR registries, and assesses the completeness and accuracy of the data.⁵ The annual data submissions from CCRs add a new year of data, update data from prior diagnosis years, and include the variables needed to calculate survival rates (e.g., age, sex, year of diagnosis, date of last contact) as well as variables that are important surrogates of the quality of the follow-up information obtained (e.g., type of reporting source, follow-up source, ICD revision number, cause of death) or can be used to stratify analyses (e.g., stage, county or state, race, ethnicity).

Before using NPCR-CSS data to assess cancer survival, it is necessary to understand the coding completeness and coding quality of the data elements used in survival analyses across all the participating cancer registries. NPCR rigorously evaluates the completeness of case ascertainment and data quality for each annual data submission.³ Other studies have evaluated completeness, accuracy, and data quality of some, but not all, of the NPCR-CSS data required for conducting and validating survival analyses.^{6–10} For example, German and colleagues looked at the quality of breast and prostate cancer variables⁶; Hall and co-workers compared SEER and NPCR incidence rates of cutaneous melanoma⁷; McDavid's team assessed breast, prostate, and colon variables⁸; and Singh and co-workers investigated the quality of the census tract 2000 variable.⁹

Using 1998–2001 data from 34 CCRs, Thoburn and colleagues compared NPCR-CSS incidence data to medical record data for 13 data elements among 4 primary cancer sites (lungs/bronchus, colorectal, prostate, and female breast) to assess the case completeness and data accuracy.¹⁰ The elements investigated included date of birth, race, sex, state of residence at time of diagnosis, diagnosis date, primary site, histology, behavior, grade and SEER summary stage. The authors found data accuracy in 95% of the cases and case completeness in 96% of the cases; individual site-specific data element accuracy ranged from 81.2% to 100%, with a median accuracy of 98.1%.¹⁰

The purpose of the current study is to build upon these previous studies by evaluating the coding completeness and quality of 26 data elements, for all primary cancer sites, that are covariates to survival calculations and those used to assess the calculations across the 46 funded CCRs.

Materials and Methods

Using NPCR-CSS data from 46 CCRs inclusive of 1995 through 2008 diagnosis years, 26 data elements used in survival calculations and validation were examined for coding completeness and quality. Coding completeness was assessed by calculating the proportion of non-missing values by data element and by central registry; the numerator was the number of non-missing values and the denominator was the total number of values (Table 1). Coding quality was calculated through the proportion of known values; the numerator was the number of known values (excluding unknown or blank values) and the denominator was the total number of values (Table 1).

The data elements consisted of survival analysis variables, demographic variables, cancer identification variables, follow-up and death variables, and cancer stage and prognostic variables. They specifically included:

Survival analysis: date of birth, date of diagnosis, date of last contact/death, and sex;

Demographics: age at diagnosis, U.S. state of residence, county of residence at diagnosis, race, ethnicity (Hispanic), Indian Health Service (IHS) linkage (used to better classify cases of American Indian/Alaska Native heritage), North American Association of Central Cancer Registries' (NAACCR) Hispanic Identification Algorithm (NHIA) derived Hispanic origin

and NAACCR Asian/Pacific Islander Identification Algorithm (or NAPIIA, which is used to better classify cases of Asian Pacific Islander origin);

Cancer identification: behavior (benign, in situ, or malignant), diagnostic confirmation, histology, SEER primary site group, sequence number central (number of primary cancers), and type of reporting source (primary source from which original cancer incidence report received);

Follow-up, recurrence, and death: cause of death code, follow-up source central (2006–2008 diagnosis years for which the variable was captured by CCRs), follow-up source (1995–2005 diagnosis years for which the variable was captured by CCRs) (source from which follow-up information received), International Classification of Diseases (ICD) revision number (cause of death coding system version; in the analyses this variable was divided into separate groups as versions 7–9 and version 10 or combined into 1 variable, versions 7–10), and vital status;

Cancer stage and prognostic: SEER summary stage 1977 (1995–2000 diagnosis years for which the variable was in use), SEER summary stage 2000 (2001–2003 diagnosis years for which the variable was in use), and collaborative stage (CS) derived SEER summary stage 2000 (2004–2008 diagnosis years for which the variable was in use).

The mean percent coding completeness for each of the 46 central cancer registries was calculated and averaged over the 26 data elements combined and all diagnosis years (1995–2008) combined; the mean percent coding completeness was also calculated for each of the 26 data elements averaged over all the diagnosis years combined and all central cancer registries combined. Mean percent coding quality was assessed by each of the 26 data elements, taking the average of all diagnosis years and all central cancer registries combined. General linear modeling was performed to assess statistical differences for each data element by diagnosis year (year was coded as a categorical variable; the latest year available, 2008, was used as the referent year) and NPCR central cancer registry (assessed individually; the referent variable was a state that has maintained high coding quality and stability over time) ($\alpha=0.05$). Coding completeness of each data element was modeled as the outcome variable with diagnosis year and NPCR central cancer registry as the independent variables, respectively, in a least squares linear model. All analyses were conducted using SAS (version 9.2, SAS Institute, Inc., Cary, NC).

Results

Coding Completeness

The mean overall coding completeness by each NPCR central cancer registry averaged across all 26 data elements and all diagnosis years ranged from 92.3% to 100% (Figure 1). Twenty-one of the central cancer registries (46%) had a mean overall coding completeness greater than or equal to 99% (Figure 1). All, but one, of the 26 data elements' mean coding completeness by combined diagnosis year was greater than 90%; the completeness for the elements ranged from 91.9% to 100% (Table 2). Follow-up source (1995–2005) had a mean completeness of 42.2% (Table 2). Similarly, the data elements' mean coding completeness

by all central cancer registries combined was high; the majority ranged from 96.8% to 100% (Table 2). Only two of the elements were lower than that range: mean coding completeness by central cancer registry for Indian Health Service linkage was 88.9% and follow-up source (1995–2005) was 64.9% (Table 2). When we examined the data elements grouped by variable type, the mean coding completeness was above 90% for each group; 99.3% over the 4 survival analysis variables combined, 98.6% over the 8 demographic variables, 100% over the 6 cancer identification data elements combined, 92.1% for the 5 follow-up/death data elements, and 100% for the 3 cancer stage/prognostic (not shown in tables).

In assessing the statistical difference in mean coding completeness of each data element by diagnosis year through the general linear models procedure, we found that date of birth, SEER summary stage 2000, CS derived SEER summary stage 2000, SEER summary stage 1977, cause of death, and follow-up source (2006–2008) showed no statistically significant differences by diagnosis year (Table 3). For ICD revision number, there was statistically significant difference comparing 2008 to 1996, but there were no statistically significant differences for the other diagnosis years. The mean coding completeness rates for the vital status data element are statistically significantly different comparing 2008 to the thirteen years, 1995 through 2007 diagnosis years. Date of last contact's coding completeness rates for 1995 and 2000–2003 were significantly different compared to 2008.

We also found that the percent coding completeness for the follow-up source variable for 20 NPCR central cancer registries were less than and statistically significantly different from the remaining NPCR central cancer registries (Table 3). For variables ICD revision number, date of birth, cause of death, and SEER summary stage 2000, only one central cancer registry has a less than statistically significant difference in mean coding completeness percent for each of these data elements. CS derived SEER summary stage 2000, IHS linkage, SEER summary stage 1977, vital status, follow-up source central, and date of last contact/death data elements all had two central cancer registries with a less than mean coding completeness percent statistically significantly different from all other NPCR central cancer registries.

Coding Quality

All of the survival analysis variables (date of birth, date of diagnosis, date last contact/death, and sex) achieved 100% coding quality (Figure 2). The majority of other variables also had a mean percent coding quality greater than 90%. The exceptions to this high percent were: cause of death (78% ICD versions 7, 8, or 9 and 81% ICD version 10), follow-up source (1995–2005) (33%), SEER Summary Stage 1977 (85%), and SEER Summary Stage 2000 (87%).

Discussion

The results show a high level of coding completeness and quality for the 4 survival analysis variables (date of diagnosis, date of birth, date of last contact or death, and sex) across central cancer registry sites and diagnosis years. The majority of variables used to assess quality of the follow-up information and to stratify analyses also had high averaged means of coding completeness and quality by central registry sites and diagnosis years. These

findings may be indicative of the training and support CDC NPCR provides in monitoring and improving coding completeness.

The mean coding completeness percent for the relative survival data elements examined is relatively high, compared with the previous studies,^{6–10} with an increase in the 2006 diagnosis year, and is similar among NPCR CCRs. This increase may result from the increase in NPCR CCRs conducting linkage processes with the National Death Index (NDI), including additional data editing and record updating, as well as the availability of training sessions and other resources. Demographic (some of which are evaluated annually to determine compliance with NPCR Program Standards), cancer identification, cancer stage/prognostic, and the majority of the follow-up/death data elements have high mean coding completeness percentages.

Even though the results are very promising, additional work may be needed for some of the data elements (e.g., date of last contact/death, follow-up source, IHS linkage). Some NPCR central cancer registries have concerns releasing the full date of last contact/death due to confidentiality while other central cancer registries may not be updating this data element following death certificate clearance procedures and/or NDI linkages. Additional discussions to assure confidentiality or resources to facilitate automatic record updates may be needed to improve the completeness of date of last contact /death. Our analysis also showed how competing priorities can affect coding completeness, as exhibited with the date of last contact/death variable. Starting in 2001, NPCR established a linkage agreement with NDI, which facilitated improved linkages and date of last contact/death information. However, in 2004, when Collaborative Stage activities became a priority for CCRs, the linkages were not completed as frequently and the completeness of the variable was affected.

The data element follow-up source (1995–2005 diagnosis years) has not been required by CDC for the NPCR registries, so the low level of completeness is not surprising. However, this data element is important; it makes it possible to identify records with information resulting from NDI linkages and, when necessary, release of that information can be recorded and reported back to NDI. The data item can also serve as a surrogate for the quality of the follow-up information.

As shown in the Results, the cause of death variable has a low percent of coding quality, 78% and 81% for the different ICD versions. The cause of death is dependent upon information recorded on death certificates, available through vital statistics linkages, and data quality issues have been identified in other evaluations. For this reason, researchers generally rely on relative survival rates for cancer rather than cause-specific survival rates.

Not all NPCR CCRs link with the IHS Administrative database on an annual basis, but all do link every 5 years. If a record is not sent for IHS Administrative database linkage, this data element is not coded. This most likely explains the lower percent completeness for the IHS linkage variable. Additional analyses may be needed, limiting the analyses to only those CCRs that conduct the IHS linkage annually or to the years where all CCRs conduct the

linkage. The coding quality evaluation, however, showed that for records that were sent for linkage, the percent with a coded known value is very high.

Our results show that the NPCR-CSS data can be a complete source of information for researchers interested in using population-based cancer data to study cancer relative survival in the U.S. Another strength of the NPCR data is the potential to calculate relative survival by race and ethnicity, which may assist researchers and comprehensive cancer control coalitions in making decisions about the type of cancer care and cancer programs they provide to the various ethnicities in the U.S., thereby having the potential to reduce disparities in cancer incidence and survival.

More work is needed to improve coding completeness for cancer case follow-up, the information with the lowest mean coding completeness percentages in this database. A limitation of this study is that we did not assess data accuracy for these relative survival variables. Evaluating the data accuracy requires an audit of the source documents and the assigned codes. Other projects are conducting this evaluation and include some, but not all, of the data elements assessed in this project.

Survival analysis estimates are critical for many prevention, control, and treatment activities, including evaluation of the impact of screening and comprehensive cancer control programs and assessing the progress in cancer treatments. Because NPCR provides data for approximately 96% of the U.S. population, it has the potential to provide near-national estimates as well as regional and state-based measures that have not before been available to researchers, clinicians, and public health decision makers. Our analyses demonstrate the high coding completeness and quality of the NPCR-CSS variables that are needed to calculate relative survival estimates and variables used to validate and stratify the estimates.

Acknowledgments

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

References

1. Andersson TM, Dickman PW, Eloranta S, Lambert PC. Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models. *BMC Med Res Methodol*. 2011; 11:96. [PubMed: 21696598]
2. Dickman PW, Adami HO. Interpreting trends in cancer patient survival. *J Intern Med*. 2006 Aug; 260(2):103–117. [PubMed: 16882274]
3. U.S. Cancer Statistics Working Group. Atlanta: U.S. Department of Health and Human Services, for Disease Control and Prevention and National Cancer Institute; 2013. United States Cancer Statistics 1999–2009 Incidence and Mortality Web-based Report. www.cdc.gov/uscs. [Accessed 12/5/2013]
4. Howlander, N.; Noone, AM.; Krapcho, M., et al. [Accessed 12/5/2013] SEER Cancer Statistics Review, 1975–2010, National Cancer Institute. Based on November 2012 SEER data submission, posted to the SEER web site. 2013 Apr. http://seer.cancer.gov/csr/1975_2010/
5. Centers for Disease Control and Prevention. [Accessed 12/5/2013] About the National Program of Cancer Registries. <http://www.cdc.gov/cancer/npcr/about.htm>.
6. German RR, Wike JM, Bauer KR, et al. Quality of cancer registry data: findings from CDC-NPCR's Breast and Prostate Cancer Data Quality and Patterns of Care Study. *J Registry Management*. 2011 Summer;38(2):75–86.

7. Hall HI, Jamison P, Fulton JP, Clutter G, Roffers S, Parrish P. Reporting cutaneous melanoma to cancer registries in the United States. *J Amer Acad Dermatol*. 2003 Oct; 49(4):624–630. [PubMed: 14512907]
8. McDavid K, Schymura MJ, Armstrong L, et al. Rationale and design of the National Program of Cancer Registries' Breast, Colon, and Prostate Cancer Patterns of Care Study. *Cancer Causes Contr*. 2004 Dec; 15(10):1057–1066.
9. Singh S, Ajani UA, Wilson RJ, Thomas CC. Evaluating quality of census tract data in the NPCR dataset. *J Registry Management*. 2009 Winter;36(4):143–146.
10. Thoburn KK, German RR, Lewis M, et al. Case completeness and data accuracy in the Centers for Disease Control and Prevention's National Program of Cancer Registries. *Cancer*. 2007 Apr 15; 109(8):1607–1616. [PubMed: 17343277]

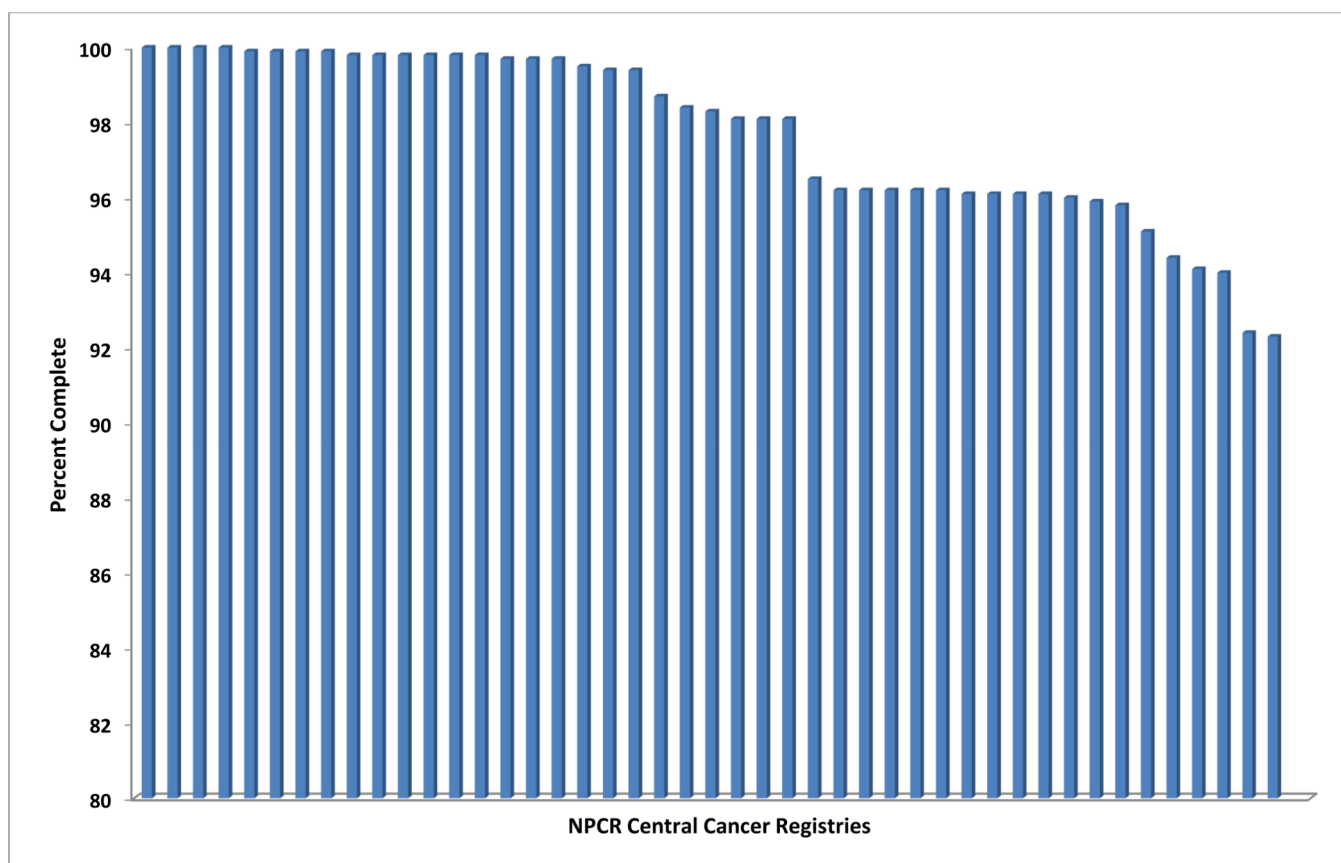
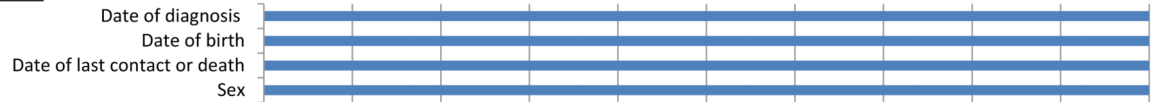
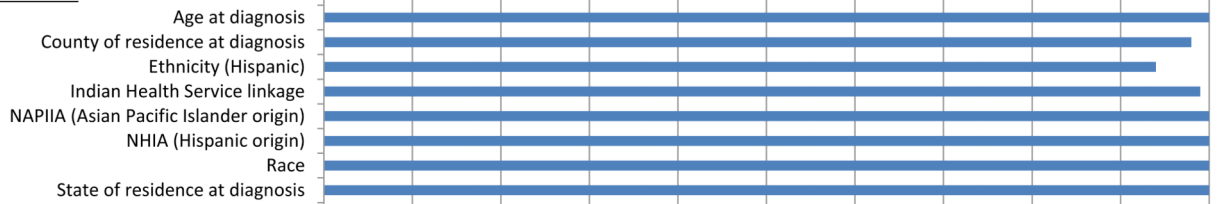
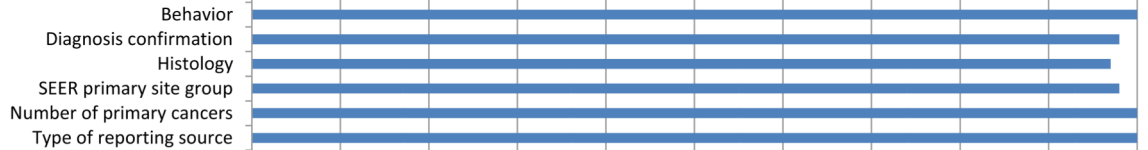
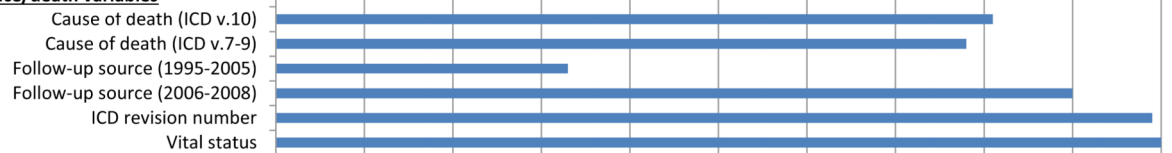


Figure 1. Mean Percent Coding Completeness Averaged Over All Data Elements Combined (n=26) and Over All Diagnosis Years Combined (1995–2008) by National Program of Cancer Registries (NPCR) Central Cancer Registry (n=46).

Survival analysis variables**Demographic variables****Cancer identification variables****Follow-up/recurrence/death variables****Cancer stage/prognostic variables**

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Coding Quality Percent

Figure 2.

Mean Percent Coding Quality of Known Value for Relative Survival Data Elements
Averaged Over All Diagnosis Years (1995–2008) and Over All NPCR Central Cancer
Registries (n=46) Data

Table 1

Coding Completeness and Coding Quality: Example Using the County of Residence Variable

County of Residence at Diagnosis – Codes		
	Coding Completeness	Coding Quality
Numerator	000, 001–840, 999	001–840
Denominator	000, 001–840, 999, invalid/blank	000, 001–840, 999, invalid/blank

Table 2

Mean Percent Coding Completeness for Relative Survival Data Elements Averaged Over All Diagnosis Years (1995–2008) and Over All National Program of Cancer Registries (NPCR) Central Cancer Registries (n=46), NPCR Data

Data Elements	Mean Coding Completeness	
	By Diagnosis Years Combined (1995–2008) (%) [range]	By All NPCR Central Cancer Registries (n=46) (%) [range]
Survival Analysis Variables		
Date of birth	99.9 [99.8–100]	99.9 [98.9–100]
Date of diagnosis	100 [–]	100 [–]
Date of last contact or death	96.2 [95.3–98.2]	97.1 [47.4–100]
Sex	100 [–]	100 [–]
Demographic variables		
Age at diagnosis	100 [–]	100 [–]
County of residence at diagnosis	100 [–]	100 [–]
Ethnicity (Hispanic)	100 [–]	100 [–]
Indian Health Service linkage	91.9 [72.6–98.0]	88.9 [59.8–100]
NHIA (Hispanic origin)	99.3 [99.0–99.6]	100 [–]
NAPIIA (Asian Pacific Islander origin)		
Race	100 [–]	100 [–]
State of residence at diagnosis	100 [–]	100 [–]
Cancer identification variables		
Behavior	100 [–]	100 [–]
Diagnostic confirmation	100 [–]	100 [–]
Histology	100 [–]	100 [–]
SEER primary site group	100 [–]	100 [–]
Number of primary cancers	100 [–]	100 [–]
Type of reporting source	100 [–]	100 [–]
Follow-up/recurrence/death variables		
Cause of death (ICD v.7–10)	99.7 [99.2–99.9]	99.6 [91.4–100]
Follow-up source (1995–2005)	42.2 [38.5–43.6]	64.9 [0–100]
Follow-up source (2006–2008)	98.0 [97.5–98.7]	96.8 [9.6–100]
ICD revision number	99.8 [99.2–100]	99.9 [94.8–100]
Vital status	99.5 [94.9–100]	99.8 [98.9–100]
Cancer stage/prognostic variables		
SEER Summary Stage 1977 (1995–2000)	100 [–]	100 [–]
SEER Summary Stage 2000 (2001–2003)	100 [–]	100 [99.2–100]
SEER Summary Stage 2000 (CS derived) (2004–2009)	100 [–]	100 [99.9–100]

Table 3

General linear models procedure (GLM) to assess percent mean coding completeness differences for relative survival data elements by diagnosis year and by NPCR central cancer registry using NPCR-CSS data (1995–2008)

Data Elements	Diagnosis Year		NPCR Central Cancer Registry (CCR) ***	
	Statistical Difference **	p-value *	Statistical Difference **	p-value *
Survival analysis variables				
Date of birth	NSD		1 NPCR CCR SD from all other NPCR CCRs	<0.0001
Date of diagnosis	NSD		NSD	
Date of last contact or death	2008 SD from 1995, 2000 – 2003 2008 NSD from 1996 – 1999, 2004 – 2007	0.02 (1995) 0.04 (2000) 0.02 (2001) 0.02 (2002) 0.03 (2003)	2 NPCR CCRs SD from all other NPCR CCRs	<0.01 <0.01
Sex	NSD		NSD	
Demographic variables				
Age at diagnosis	NSD		NSD	
County of residence at diagnosis	NSD		NSD	
Ethnicity (Hispanic)	NSD		NSD	
Indian Health Service linkage	2008 SD from 1995 – 2002 2008 NSD from 2003 – 2007	0.03 (1995) 0.04 (1996–2002)	2 NPCR CCRs SD from all other NPCR CCRs	<0.01 <0.01
NHIA (Hispanic origin)	2008 SD from 1995	0.01 (1995)	NSD	
NAPIIA (Asian Pacific Islander origin)				
Race	NSD		NSD	
State of residence at diagnosis	NSD		NSD	
Cancer identification variables				
Behavior	NSD		NSD	
Diagnosis confirmation	NSD		NSD	
Histology	NSD		NSD	
SEER primary site group	NSD		NSD	
Number of primary cancers	NSD		NSD	
Type of reporting source	NSD		NSD	
Follow-up/recurrence/death variables				
Cause of death (ICD v.7–10)	NSD		1 NPCR CCR SD from all other NPCR CCRs	<0.0001
Follow-up source (1995–2005)	2005 NSD from other diagnosis years		20 NPCR CCRs SD from all other NPCR CCRs	
Follow-up source (2006–2008)	NSD		2 NPCR CCRs SD from all other NPCR CCRs	<0.01 <0.02
ICD revision number	2008 SD from 1996 2008 NSD from other diagnosis years	0.04 (1996)	1 NPCR CCR SD from all other NPCR CCRs	<0.0001

Data Elements	Diagnosis Year		NPCR Central Cancer Registry (CCR) ***	
	Statistical Difference **	p-value *	Statistical Difference **	p-value *
Vital status	2008 SD from 1995–2007	0.02 (1995–2007)	2 NPCR CCRs SD from all other NPCR CCRs	<0.01 <0.01
Cancer stage/prognostic variables				
SEER Summary Stage 1977 (1995–2000)	NSD		2 NPCR CCRs SD from all other NPCR CCRs	<0.01 <0.01
SEER Summary Stage 2000 (2001–2003)	NSD		1 NPCR CCR SD from all other NPCR CCRs	<0.0001
SEER Summary Stage 2000 (CS derived) (2004–2009)	NSD		2 NPCR CCRs SD from all other NPCR CCRs	<0.01 <0.01

* p-value calculated at alpha = 0.05 level of significance

** **SD** – statistically significant difference; **NSD** – no statistically significant difference

*** Statistically significant NPCR Central Cancer Registries all had a mean completeness less than that of the referent NPCR Central Cancer Registry