CASE COMPLETENESS AND DATA QUALITY IN THE NATIONAL PROGRAM OF CANCER REGISTRIES (NPCR)

KK Thoburn[1], RR German[2], MB Lewis[2], PJ Nichols[2], and J Jackson-Thompson[3]

[1]New York State Cancer Registry, [2]Centers for Disease Control and Prevention, [3]University of Missouri-Columbia

ABSTRACT

CASE COMPLETENESS AND DATA QUALITY IN THE NATIONAL PROGRAM OF CANCER REGISTRIES (NPCR). KK Thoburn[1], RR German[2], MB Lewis[2], PJ Nichols[2] and J Jackson-Thompson[3], [1]New York State Cancer Registry, [2]Centers for Disease Control and Prevention, [3]University of Missouri-Columbia

To address the issues of case completeness and data quality (CCDQ) within the NPCR, the NPCR conducts Technical Assistance and Audits (NPCR-TAA) of NPCR-funded central cancer registries (CCRs) following guidelines established by CDC and NAACCR. Case completeness is assessed by independently casefinding cancer cases in sample hospitals with differing caseloads. The level of data quality is assessed by reabstracting a sample of cancer cases from the same hospitals and comparing the reabstracted values for each sample case with the values existing in the central registry.

In addition to the NPCR-TAA, the NPCR conducts a web-based Annual Program Evaluation Instrument (NPCR-APEI). The NPCR-APEI is used to evaluate various attributes of CCRs, monitor their progress towards program standards, goals and objectives and respond to data inquiries.

To be able to draw accurate conclusions regarding cancer incidence, the user of the data must be aware of any case completeness or data quality issues. A large number of studies have been published utilizing incidence data from NPCR registries. Because both CCRs and researchers will benefit from the perspective obtained by a summary analysis of NPCR-TAA results across multiple states along with information gleaned from the NPCR-APEI, a national-level analysis of 1998-2001 NPCR-TAA audit results linked with NPCR-APEI information was undertaken. This presentation will convey the results of these analyses and characterize the outcome of NPCR support for a statewide, population-based CCR.

INTRODUCTION

NPCR Cancer Surveillance System (NPCR-CSS)
- Established in 2000
- Measures incidence rates by various characteristics
- Data estimates CCR progress meeting NPCR program standards

NPCR Technical Assistance and Audit Program (NPCR-TAA)
- ORC Macro International Inc. conducts NPCR–TAA
- CCDQ audits performed at the level of hospital reporting to CCR
- Uses sample of cases from selected hospitals
- Technical assistance provided to CCRs
- Four most common cancers reabstracted: lung and bronchus, colorectal, prostate and female breast (represent over 50% of cases reported to NPCR-CSS)

NPCR Annual Program Evaluation Instrument (NPCR-APEI)
- Assessment of NPCR program objectives, registry operations and data use
- Web-based instrument completed annually by all NPCR-funded CCRs
- Serves as one required progress report
- Aggregate results presented at NAACCR with future dissemination planned at state/territory level

NAACCR Certification
- All NPCR registries are members of NAACCR
- Registries certified for high quality data based on standardized evaluation of CCR data for completeness, accuracy and timeliness
- CCRs meeting NAACCR standards for data quality are recognized annually through certification at the Gold or Silver level

Present Study
- Usefulness of incidence data from NPCR-CSS depends upon the completeness of case reporting and the accuracy of information contained in the reports of submitted cases
- Data from the NPCR-CSS are frequently utilized to measure incidence and survival according to various characteristics such as race, gender, age, subsite, and stage at diagnosis
- To be able to draw accurate conclusions the user of the data must be aware of any CCDQ issues
- Present study consists of a unique, national-level covariate analysis of NPCR-TAA results with information from the NPCR-APEI, as well as information regarding CCR achievement of USCS publication standards and NAACCR certification

STUDY OBJECTIVES

1. **Summarize NPCR-TAA results for lung and bronchus, colorectal, prostate and female breast cancer for the diagnosis years 1998-2001 across all states audited for those diagnosis years.**

2. **Analyze reabstraction results for specific data elements critical to the surveillance of lung and bronchus, colorectal, prostate and female breast cancer and identify primary site-specific data quality issues.**

3. **Conduct a covariate analysis on CCR audit case completeness and error rates utilizing information from CCR responses to the NPCR-APEI.**

4. **Characterize the outcome of NPCR support for a statewide, population-based CCR.**

METHODS

Table 1. States Included for Analysis

Name of State	# of States	Diagnosis Year of NPCR Audit	Year of NPCR-APEI[1]
Colorado, Massachusetts, New Hampshire, North Dakota, Ohio, West Virginia	6	1998	2000
Idaho, Illinois, Maryland, Minnesota, Missouri, Nebraska, North Carolina, Tennessee, Vermont, Washington	10	1999	2001
Arizona, District of Columbia, Michigan, Montana, Nevada, New York, Oregon, Pennsylvania, Virginia	9	2000	2003[2]
Arkansas, Florida, Georgia, Kansas, Maine, Oklahoma, Rhode Island, Texas, Wisconsin	9	2001	2003
TOTAL	**34**		

[1]The APEI information utilized described the diagnosis year being audited. As the majority of cancer cases are reported within 24 months of diagnosis and many of the APEI questions inquire about the past 12 or 24 months, information from the APEI that was administered 2 years following that of the diagnosis year audited was utilized.
[2]Since APEI information from 2002 is not available, information from the 2003 APEI was used to impute values.

- NPCR-TAA audit results for diagnosis years of 1998 through 2001 for 34 NPCR-supported CCRs were included in the analysis (see Table 1)
- Data for each state and audit diagnosis year were aggregated for a descriptive analysis of average case completeness rates and cancer site-specific error rates by covariates obtained from the NPCR-APEI, NPCR staff and the NAACCR website (see Table 2)
- ORC Macro provided a SAS data file of NPCR-TAA data for each state and audit diagnosis year for analysis
- CDC staff obtained select, covariate data from the NPCR-APEI, NPCR staff and the NAACCR website; these data were linked by state and audit diagnosis year to the ORC Macro SAS data file
- Average case completeness and site-specific error rates were calculated; error rates were calculated by primary site for each of the 13 data elements (see Figure 3) examined in the audit program

  Case completeness rates (%) = 100 - ((number of missed cases / total number of cases identified) x 100)
  Error rates (%) = (number of discrepancies / total number of data elements reabstracted) x 100

- When possible, continuous covariates were dichotomized at the median
- Average case completeness and site-specific error rates were examined across the covariates, and small-sample Student's t tests were used to determine statistically significant differences ($\alpha = 0.05$)

RESULTS: Case Completeness

- Across all 34 CCRs that were audited, the total number of lung and bronchus, colorectal, prostate and female breast cancer cases were found upon casefinding audit in sample hospitals was 41,521
- 1,503 (3.6%) of these cases were identified as originally 'missed' by the CCRs, resulting in an **overall case completeness rate of 96.4%**
- Female breast and prostate cancer cases were more likely to be missed than lung and bronchus and colorectal cases; 491 female breast and 430 prostate cancer cases were missed, which accounted for 61.3 % of all missed cases (see Figure 1)

Figure 1. Distribution of Missed Cases by Primary Site
N = 1,503

- The 1,503 missed cases were predominantly identified in one casefinding source (see Table 3)

Distribution of Missed Cases by Number of Casefinding Sources:

1 Source	1,367	(91.0%)
2 Sources	133	(8.8%)
3 or More Sources	3	(0.2%)

Table 3. Number of Missed Cases by Number of Missed Casefinding Sources

Primary Site	Lung and Bronchus	Colon and Rectum	Prostate	Female Breast
Number (%) of Cases Missed in 1 Source	274 (91.7)	251 (88.7)	403 (93.7)	439 (89.4)
Number (%) of Cases Missed in 2 Sources	24 (8.0)	32 (11.3)	26 (6.1)	51 (10.4)
Number (%) of Cases Missed in 3 or More Sources	1 (0.3)	0 (0.0)	1 (0.2)	1 (0.2)

- Missed cases were predominantly identified in the Medical Records Disease Index (MRDI) and pathology reports

Distribution of Missed Cases by Type of Casefinding Source:

MRDI	906	(60.3%)
Pathology Reports	665	(44.2%)
Radiation Therapy (RT) Log	39	(2.6%)
Cytology Reports	17	(1.1%)
By Autopsy	12	(0.8%)

- Number and distribution of missed sources by type of casefinding source varied by primary site (see Figure 2)

Figure 2. Distribution of Missed Casefinding Sources by Type of Casefinding Source and Primary Site

Lung & Bronchus N = 324
Colorectal N = 315
Prostate N = 458
Female Breast N = 544

Legend: MRDI, Pathology, RT Log, Cytology, Autopsy, Other

RESULTS: Covariate Analysis

Number of Years from CCR's NPCR Reference Year
- Significantly higher average case completeness found for registries 7 or more years from their reference year (p = 0.04)
- Unexpectedly, significantly higher error rate found for registries 7 or more years from their reference year for the site of female breast (p = 0.05); similar (not significant) pattern observed for other primary sites (see Table 4)

FTE and CTR-to-Caseload Ratios for CCR
- Higher average case completeness and lower average error rates found for CCRs staffed with more FTEs
- Significantly lower average error rate found for the colorectal and prostate primary sites (p = 0.03 and 0.04, respectively) for CCRs staffed with more CTRs; similar (not significant) pattern observed for other primary sites
- Findings emphasize the need for adequate and well-trained staff in CCRs

Supplementary Reporting Sources
- Significantly higher average case completeness rate found for registries that had pathology laboratories and/or radiation therapy centers reporting (p < 0.01)
- Lower average error rate was observed across all primary sites for CCRs that identified reporting by supplemental reporting sources
- Findings suggest that CCRs can attain higher case completeness and better data quality by obtaining reports from supplemental reporting sources such as pathology labs and radiation therapy centers

Met USCS Publication Standards/Achieved NAACCR Certification
- Lower average error rate found across all sites for CCRs that had achieved these two data quality milestones
- Findings demonstrate a clear benefit to CCR data quality as a result of attaining national data quality standards

Table 2. Listing of Variables and Their Sources

Variable[1]	Data Source of Variable
Primary variables of interest	
Case completeness rate[2] (percent, plus numerators and denominators)	Audit Report
Error rate (percent, plus numerators and denominators)	Audit Report
Covariates	
Anatomic site (lung/bronchus, colorectal, female breast, prostate)	Audit Report
Data elements examined in audit program	Audit Report
Source of missed cases (1 or 2 sources vs. 3 or more sources)	Audit Report
Years from NPCR reference year (6 years or less vs. 7 or more years)	Evaluation Instrument
Type of current funding (enhancement vs. planning)	Evaluation Instrument
Proportion of FTE positions to central registry caseload	Evaluation Instrument
Proportion of CTRs to central registry caseload	Evaluation Instrument
Supplementary reporting sources report cases (yes [path labs and radiation therapy centers] / no)	Evaluation Instrument
Casefinding audits at reporting facilities (yes/no)	Evaluation Instrument
Reabstracting audits at reporting facilities (yes/no)	Evaluation Instrument
Annual report issued (yes [hardcopy, electronic, web] / no)	Evaluation Instrument
Met NPCR data standards (yes/no – 1999-2001 only)	NPCR Staff
Certified by NAACCR (yes/no)	NAACCR Web Site

[1]Variables are for each state and audit diagnosis year.
[2]Not examined by primary site due to unavailability of denominator data by primary site.

RESULTS: Data Quality

- For each of the 1,641 records sampled, 13 data elements were reabstracted (total of 130,130 data elements examined)
- 6,531 data elements were found to have discrepancies upon reabstraction resulting in an **overall data accuracy rate of 95.0%**
- Percentage of error-free data elements varied by primary site:
  - Prostate 97.1%
  - Colorectal 95.2%
  - Female Breast 94.3%
  - Lung and Bronchus 93.6%
- The data discrepancy rates for each of the 13 audited data elements varied widely both between data elements as well as across primary sites (see Figure 3)

Overall Discrepancy Rates:

Gender	0.2%
State of Residence	0.3%
Behavior	0.7%
Primary Site	0.7%
Date of Birth	1.4%
Laterality	1.7%
Race	2.0%
Sequence Number	2.6%
Histology	8.9%
Subsite	10.0%
Grade	10.6%
SEER Summary Stage	11.2%
Date of Diagnosis	14.9%

Figure 3. Discrepancy Rates by Primary Site for Audited Data Elements

Legend: Lung & Bronchus, Colorectal, Prostate, Female Breast

Table 4. Results from Covariate Analysis of Case Completeness & Error Rates Utilizing APEI and NAACCR Certification Information

Item	N[1]	Case Completeness Rate		Site-Specific Error Rate							
				Lung & Bronchus		Colorectal		Prostate		Female Breast	
		Average	P-value[2]	Average	P-value	Average	P-value	Average	P-value	Average	P-value
Overall	34	96.5		6.3		4.8		2.9		5.8	
Years from reference year for NPCR											
7 years or more	17	97.5	0.04	6.9	0.21	5.4	0.18	3.3	0.18	6.6	0.05
Less than or equal to 6 years	17	95.5		5.7		4.3		2.5		4.9	
Type of current funding from NPCR											
Enhancement	32	96.7	0.15	6.3	0.70	4.7	0.25	3.0	0.42	5.8	0.87
Planning	2	93.6		7.0		6.8		2.0		5.5	
FTEs-to-caseload ratio											
Greater than 6.62 FTEs per 10,000 cases	11	98.2	0.09	6.1	0.74	4.4	0.40	2.9	0.93	5.2	0.37
Less than or equal to 6.62 FTEs per 10,000 cases	13	97.1		6.5		5.2		3.0		6.2	
CTRs-to-caseload ratio											
Greater than 2.18 CTRs per 10,000 cases	15	97.9	0.23	5.6	0.13	3.9	0.03	2.4	0.04	5.1	0.15
Less than or equal to 2.18 CTRs per 10,000 cases	13	97.1		7.2		5.9		3.8		6.6	
Supplementary reporting sources report cases											
Yes, path laboratories or radiation therapy centers	29	97.1	<0.01	6.1	0.35	4.7	0.45	2.9	0.52	5.6	0.32
No	5	93.0		7.3		5.6		3.4		6.8	
Case finding audits at reporting facilities											
Yes	20	96.6	0.87	6.4	0.87	4.7	0.68	3.2	0.15	6.0	0.44
No	14	96.4		6.2		5.0		2.5		5.4	
Re-abstracting audits at reporting facilities											
Yes	20	96.9	0.36	6.1	0.58	4.5	0.32	2.9	0.79	5.8	0.90
No	14	96.0		6.6		5.3		3.0		5.7	
Annual report issued											
Yes, hardcopy, electronic, and/or web	14	96.9	0.39	5.8	0.41	4.0	0.12	2.6	0.30	5.4	0.53
No	20	96.1		6.6		5.4		3.2		6.0	
Met data standards for publication in NPCR's USCS (only 1999, 2000, and 2001)											
Yes	26	97.2	0.22	6.1	0.22	4.6	0.02	2.9	0.27	5.6	0.11
No	3	92.4		8.3		8.1		4.0		8.2	
Certified by NAACCR											
Yes, either gold or silver	30	96.9	0.34	6.1	0.21	4.5	0.04	2.8	0.33	5.5	0.06
No	4	94.0		7.3		6.3		3.7		7.9	

[1]N equals less than 34 when 1 or more registries did not respond to the APEI question being examined.
[2]P-value from Student's t-test.

CONCLUSIONS

- NPCR's standard for CCR case completeness is 95.0% of reportable cancer cases reported within 24 months of the end of the diagnosis year. The overall estimated case completeness rate for the 34 CCRs audited from 1998-2001 was 96.4%, exceeding the national standard.
- The overall data accuracy rate for the 34 CCRs audited from 1998-2001 was 95.0%. Therefore, users of NPCR-CSS incidence data may have confidence that the data accurately represents what occurred at the level of patient diagnosis and treatment. In addition, the identified site-specific data quality issues will facilitate the appropriate interpretation of findings from studies using NPCR-CSS incidence data.
- The findings from the covariate analysis illustrate and emphasize the importance and positive effect on CCR case completeness and data quality of CCRs having adequate, well-trained staff, procuring supplemental reporting sources and attaining compliance with national data standards.
- As NPCR funding assists states in the development and enhancement of effective registry operations, especially in areas such as staffing, training and monitoring and improving the completeness and quality of registry data, this study suggests a positive outcome of NPCR support for a statewide, population-based CCR.