The Impact of Underreported Veterans Affairs Data on National Cancer Statistics: Analysis Using Population-Based SEER Registries

Nadia Howlader, Lynn A. Ries, David G. Stinchcomb, Brenda K. Edwards

Reduced cancer reporting by the US Department of Veterans Affairs (VA) hospitals in 2007 (for patients diagnosed through 2005) impacted the most recent US cancer surveillance data. To quantify the impact of the reduced VA reporting on cancer incidence and trends produced by the Surveillance, Epidemiology, and End Results Program, we estimated numbers of missing VA patients in 2005 by sex, age, race, selected cancer sites, and registry and calculated adjustment factors to correct for the 2005 incidence rates and trends. Based on our adjustment factors, we estimated that as a result of the underreporting, the overall cancer burden was underestimated by 1.6% for males and 0.05% for females. For males, the percentage of patients missing ranged from 2.5% for liver cancer to 0.4% for melanoma of the skin. For age-adjusted male overall cancer incidence rates, the adjustment factors were 1.015, 1.012, and 1.035 for all races, white males, and black males, respectively. Modest changes in long-term incidence trends were observed, particularly in black males.

J Natl Cancer Inst 2009;101:533–536

A recent policy change by the US Department of Veterans Affairs (VA) regarding the sharing of VA cancer data has resulted in incomplete reporting of VA hospital patients to some central cancer registries. For example, the state of California reported a lower than expected number of cancer patients for 2005, due to the underreporting of VA hospital patients (Robbins AS, Bates JA, Cress RD, Nadia Howlader, MS, Lynn A. Ries, MS, David G. Stinchcomb, MS, MA, Brenda K. Edwards, PhD., unpublished data, 2007). The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (NCI) generates annual cancer statistics using data from population-based cancer registries covering approximately 26% of the US population. Underreporting of VA hospital patients potentially could distort population cancer incidence statistics (1,2) that were published by the SEER Program in April 2008 for patients diagnosed through 2005 (3).

This report estimates the impact of these missing VA hospital patients in 2005 on SEER incidence rates and trends. We used patient counts from VA hospitals in SEER regions to establish baseline VA hospital patient counts in years with no known underreporting and to estimate missing VA hospital patients for 2005. We then adjusted final patient counts for selected primary cancer sites based on the estimated missing patients. Finally, we compared long-term trends summarized by annual percent change (APC) (4) with and without the VA adjustment.

The de-identified data reported to SEER do not indicate whether a patient comes from a VA hospital. To estimate the underreporting, most SEER registries provided a special tabulation of VA hospital patient counts for 2000–2005 that were stratified by year of diagnosis, sex, primary cancer site, and 5-year age group, with no patient identifiers. VA patients may receive some treatment at non-VA facilities and would be reported by those facilities, but data provided by the SEER registries were for patients who were treated only at VA facilities. We excluded from our analysis four SEER registries that did not experience interruptions in VA reporting to state registries for cancers diagnosed through 2005. We also excluded the Louisiana Tumor Registry from our analysis. The Louisiana Tumor Registry submitted VA hospital patient counts, but population displacement following Hurricanes Katrina and Rita confounded the effects of the underreporting of 2005 VA hospital patients.

For each registry, we estimated the proportion of VA hospital patients by year of diagnosis (2000–2005) and sex. We used an average based on 2000–2003 proportions as a baseline for comparison with 2005. Differences between the 2005 proportions and the 2000–2003 average proportions were used to estimate the number of missing VA hospital patient counts in 2005 by registry, sex, age, and other covariates of interest, such as race and cancer site. We added the estimated numbers of missing patients for each age stratum to the total reported patients in 2005 to adjust for VA underreporting. Age-adjusted cancer incidence rates were then calculated with and without adjustment for VA underreporting. The ratio between these age-adjusted cancer incidence rates yielded a factor that was then used to adjust the 2005 overall incidence rates. We estimated VA adjustments based on nondelay adjusted counts. In addition, we estimated non–VA-adjusted rates using nondelay adjusted counts. Additional information on this calculation is available in the 2008 release of the SEER Cancer Statistics Review (5).

We found that the underreporting of VA hospital patients led to the overall cancer burden in selected SEER registries to be underestimated by 1.6% for males (Table 1) and 0.05% for females (data not shown). Because female patients are a small proportion of the total VA hospital

Affiliation of authors: Cancer Statistic Branch, Surveillance Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD.

Correspondence to: Nadia Howlader, MS, Cancer Surveillance Branch, Division of Cancer Control and Population Sciences, National Cancer Institute, 6116 Executive Blvd Suite 504, Bethesda, MD 20892-8315 (e-mail: howladern@mail.nih.gov).

See "Funding" and "Notes" following "References."

DOI: 10.1093/jnci/djn517

CONTEXT AND CAVEATS

Prior knowledge

A recent policy change by the US. Department of Veterans Affairs (VA) has resulted in the underreporting of incident cancer cases to some cancer registries.

Study design

Estimated numbers of missing VA cancer patients in 2005 were used to correct cancer incidence rates and trends in the Surveillance, Epidemiology, and End Results Program.

Contributions

Due to the underreporting, overall cancer burden was underestimated by 1.6% for males and 0.05% for females. Changes in long-term incidence were modest and occurred mainly among black males.

Implications

The reduced reporting of VA patients to central cancer registries led to an underestimation of cancer incidence that modestly changed long-term trends.

Limitations

The variance in the estimated adjustment factors was not determined.

From the Editors

caseload, we focused on male VA hospital patients and examined overall cancer (by race) and the 10 leading primary cancer sites. In 2000–2003, male VA hospital patients were 3.7% of the overall cancer patients (3.4% white and 8.4% black); by contrast, in 2005 VA hospital patients were 2.1% (1.9% white and 4.5% black). Therefore, we estimated that the percentage of missing patients was 1.6% overall (1.5% white and 3.9% black). Among the top 10 cancers, the percentage estimated to be missing ranged from 2.5% for liver cancer to 0.4% for melanoma of the skin. Adjustment factors for overall cancer to correct for the underreporting of 2005 age-adjusted incidence rates for SEER-17, excluding the Louisiana Tumor Registry regions, were 1.015, 1.012, and 1.035 for all races, white males, and black males, respectively (Table 1).

We also derived adjustment factors for SEER-9 regions to investigate the impact of VA underreporting on long-term trends of delay-adjusted incidence rates. Trends for overall cancers and the

Table 1. Percentage of cancer patients from VA facilities by year of diagnosis, estimated percentage of missing cancer patients for 2005, and VA adjustment factors for 2005 age-adjusted incidence rates. All sites combined (all races, whites, blacks) and top 10 common cancer sites (all races) among males

Rank based on counts	Cancer site	Percentage of cancer patients from VA facilities based on selected SEER registries*							Estimated missing patients (%) 2005†	VA adjustment factor for 2005 age-adjusted incidence rate‡
		2000	2001	2002	2003	2004	2005	Mean (2000–2003)		
	All sites combined, all races	3.5	3.7	3.8	3.8	3.1	2.1	3.7	1.6	1.015
	All sites combined, whites	3.3	3.4	3.5	3.5	3.2	1.9	3.4	1.5	1.012
	All sites combined, blacks	8.5	8.3	8.5	8.3	8.0	4.5	8.4	3.9	1.035
1	Prostate	3.3	3.6	3.9	3.7	3.7	2.3	3.7	1.4	1.012
2	Lung and bronchus	4.9	5.2	4.8	5.0	4.3	2.6	5.0	2.4	1.023
3	Colon and rectum	3.0	3.3	3.3	3.2	2.9	1.7	3.2	1.5	1.016
4	Urinary bladder	3.1	2.8	3.4	3.3	3.2	2.1	3.2	1.0	1.011
5	Head and neck	4.6	5.1	5.3	4.4	4.3	2.5	4.9	2.4	1.027
6	Melanoma of the skin	2.1	2.4	2.6	3.1	3.3	2.1	2.6	0.4	1.005
7	Kidney and renal pelvis	3.9	3.8	3.8	3.9	3.4	1.8	3.8	2.0	1.020
8	Leukemia	2.6	2.5	3.1	3.1	2.1	1.3	2.8	1.5	1.015
9	Non–Hodgkin lymphoma	2.3	2.4	2.8	2.0	2.1	1.1	2.4	1.3	1.013
10	Liver	4.6	4.3	3.9	4.4	4.0	1.8	4.3	2.5	1.024

* Surveillance, Epidemiology, and End Results (SEER) registries include Atlanta, Connecticut, Iowa, New Mexico, San Francisco, Utah, San Jose-Monterey, Los Angeles, rural Georgia, greater California, Kentucky, and New Jersey. VA = US. Department of Veterans Affairs.

† Difference between average proportion based on 2000–2003 and 2005.

‡ Adjustment factors are generated for SEER-17 (excluding Louisiana) registry.

Table 2. Comparison of long-term delay-adjusted incidence trends with and without VA adjustment for SEER-9 areas among males, 1975–2005*

		Joinpoint analyses (1975–2005)†									
		Trend 1		Trend 2		Trend 3		Trend 4		Trend 5	
Cancer site	Race	Years	APC‡	Years	APC‡	Years	APC‡	Years	APC‡	Years	APC‡
All sites	All	1975–1989	1.3§	1989–1992	5.2	1992–1995	−4.8§	1995–2001	0.3	2001–2005	–1.8§
All sites (VA adjusted)	All		‖		‖		‖		‖	2001–2005	–1.7§
All sites	White	1975–1989	1.4§	1989–1992	5.0§	1992–1995	–5.0§	1995–2001	0.6§	2001–2005	–1.6§
All sites (VA adjusted)	White		‖		‖		‖		‖	2001–2005	–1.5§
All sites	Black	1975–1981	2.8§	1981–1989	0.7	1989–1992	6.8§	1992–2005	–1.7§		
All sites (VA adjusted)	Black		‖		‖	1989–1992	7.6§	1992–1995	–2.9	1995–2005	–1.3§
Oral cavity and pharynx	All	1975–2005	–1.2§								
Oral cavity and pharynx (VA adjusted)	All		‖								
Colon and rectum	All	1975–1985	1.1§	1985–1991	–1.2§	1991–1995	–3.1§	1995–1998	1.9	1998–2005	–2.8§
Colon and rectum (VA adjusted)	All		‖		‖		‖		‖		‖
Liver and IBD	All	1975–2005	3.6§								
Liver and IBD (VA adjusted)	All		‖								
Lung and bronchus	All	1975–1982	1.5§	1982–1991	–0.5	1991–2005	–1.8§				
Lung and bronchus (VA adjusted)	All		‖		‖		–1.7§				
Melanoma of the skin	All	1975–1985	5.4§	1985–2000	3.4§	2000–2003	–0.2	2003–2005	7.7§		
Melanoma of the skin (VA adjusted)	All		‖		‖		‖	2003–2005	7.9§		
Prostate	All	1975–1988	2.6§	1988–1992	16.5§	1992–1995	–11.5§	1995–2001	2.1§	2001–2005	–4.4§
Prostate (VA adjusted)	All		‖		‖		‖		‖	2001–2005	–4.3§
Urinary bladder	All	1975–1986	0.9§	1986–2005	0.0						
Urinary bladder (VA adjusted)	All		‖		‖						
Kidney and renal pelvis	All	1975–2005	1.8§								
Kidney and renal pelvis (VA adjusted)	All		‖								
Non–Hodgkin lymphoma	All	1975–1991	4.2§	1991–2005	0.4§						
Non–Hodgkin lymphoma (VA adjusted)	All		‖		‖						
Leukemia	All	1975–2005	0.1								
Leukemia (VA adjusted)	All		‖								

* VA adjustment was generated for SEER-9 regions. Source: SEER-9 areas covering about 10% of the US population (Connecticut, Hawaii, Iowa, Utah, and New Mexico and the metropolitan areas of San Francisco, Detroit, Atlanta, and Seattle–Puget Sound). VA = US. Department of Veterans Affairs; SEER = Surveillance, Epidemiology, and End Results; APC = annual percent change; NOS = not otherwise specified. IBD = intrahepatic bile duct.

† Joinpoint analyses with up to four joinpoints are based on rates per 100 000 persons and were age adjusted to the 2000 US standard population (19 age groups—Census p25–1130). Joinpoint Regression Program, v 3.3.1, April 2008, National Cancer Institute.

‡ APC is based on delay-adjusted rates that were age adjusted to the 2000 US standard population (19 age groups—Census p25–1130).

§ APC is statistically significantly different from zero ($P < .05$, two-sided Student t test.).

‖ APCs are same as delay adjusted.

10 leading cancer sites with and without VA adjustment factors for the 2005 incidence data were compared (Table 2). Joinpoint statistical software (6,7) was used to fit trends over time and to evaluate when changes in trends occurred. Results show that adjusting for VA underreporting raised the APC slightly in the most recent reporting years for all sites combined and for cancers of the prostate, lung and bronchus, and melanoma of the skin. Although the VA-adjusted incidence trends showed very little change in the interpretation and conclusion of long-term trends, we detected a new change point for overall trends in black males in 1995, when trends were stable. We also observed a statistically significant decrease in trends for black males in the final segment (1995–2005), when incidence decreased 1.3% per year.

In summary, VA-adjusted incidence rates may be more accurate in capturing current SEER trends, even if the change in rates appears to be slight. A limitation of the adjustment factors is that they are not presented with the statistical uncertainty of the estimates. In reporting cancer trends, a change of as little as 1% per year demonstrates improvements or causes alerts in cancer control efforts. Such changes could easily be obscured by incomplete reporting of VA hospital data. Trends for black males in particular could be underestimated severely in the future. Our current analysis also shows that the number of patients missed in underreporting can be as large as those missed in the reporting delay that impacts national cancer incidence rates every year (7). Reporting delay is predicted and adjusted for in the presentation of the annual cancer statistics. More importantly, unlike reporting delays in which patients are captured with the passage of time, underreporting of the VA patients could result in patients being permanently missed. Thus, the absence of VA data could compromise the accuracy and completeness of estimates of the US national cancer burden. The VA and the cancer registry community therefore must continue their efforts to restore integration of cancer data on VA patients with data from state and regional population-based registries. This should not be an impossible task. The cancer registry community has a strong commitment to and experience in protecting patient confidentiality while advancing cancer control and surveillance research.

References

1. Savage L. Unreported VA data may affect SEER research, cancer surveillance, and statistics gathering [News]. *J Natl Cancer Inst.* 2007;99(23):1744–1745.
2. Furlow B. Accuracy of US cancer surveillance under threat. *Lancet Oncol.* 2007;8(9):762–763.
3. Ries LAG, Melbert D, Krapcho M, et al. *SEER Cancer Statistics Review, 1975–2005.* Bethesda, MD: National Cancer Institute; 2008. Available at: http://seer.cancer.gov/csr/1975_2005/, based on November 2007 SEER data submission, posted to the SEER web site. Accessed April 16, 2008.
4. Kim HJ, Fay MP, Feuer EJ, et al. Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med.* 2001;19(3):335–351.
5. Ries LAG, Melbert D, Krapcho M, et al. eds. SEER Cancer Statistics Review, 1975–2005, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2005/, based on November 2007 SEER data submission, posted to the SEER web site, 2008. Access Date: April 16, 2008.
6. Kim HJ, Fay MP, Feuer EJ, Midthune DN, Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med.* 2000;19:335–351.
7. Joinpoint Regression Program, Version 3.3 - April 2008; Statistical Research and Applications Branch, National Cancer Institute.
8. Clegg LX, Feuer E, Midthune D, Fay M, Hankey B. Impact of reporting delay and reporting error on cancer incidence rates and trends. *J Natl Cancer Inst.* 2002;94(20): 1537–1545.

Funding

Division of Cancer Control and Population Sciences, Surveillance Research Program, National Cancer Institute, National Institutes of Health to BK Edwards.

Notes

SEER is a set of geographically defined, population-based, central cancer registries in the United States, operated by local nonprofit organizations under contract to the NCI. Registry data are submitted electronically without personal identifiers to the NCI on a biannual basis, and the NCI makes the data available to the public for scientific research.

We would like thank Dr Eric J Feuer for discussion of the analysis, William Waldron of the Information Management System for programming assistance, and the SEER registries for participation in this project.

The authors take full responsibility for the study design, the data collection and analysis, the interpretation of the results, the preparation of the manuscript, and the decision to submit the manuscript for publication.

Manuscript received July 23, 2008; revised December 10, 2008; accepted December 30, 2008.