2014

Cancer incidence in HIV-infected versus uninfected veterans: Comparison of cancer registry and ICD-9 code diagnoses

Lesley S. Park
Yale University

Janet P. Tate
Yale University

Maria Rodriguez-Barradas
Michael E. DeBakey Veterans Affairs (VA), Medical Cente

David Rimland
Emory University

Matthew B. Goetz
University of California, Los Angeles

See next page for additional authors

Recommended Citation

Authors

Lesley S. Park, Janet P. Tate, Maria Rodriguez-Barradas, David Rimland, Matthew B. Goetz, Cynthia L. Gibert, Sheldon T. Brown, Michael J. Kelley, Amy C. Justice, and Robert Dubrow

Cancer Incidence in HIV-Infected Versus Uninfected Veterans: Comparison of Cancer Registry and ICD-9 Code Diagnoses

Lesley S Park[1,2,*], Janet P Tate[3,4], Maria C Rodriguez-Barradas[5,6], David Rimland[7,8], Matthew Bidwell Goetz[9,10], Cynthia Gibert[11,12], Sheldon T Brown[13,14], Michael J Kelley[15,16,17], Amy C Justice[3,4] and Robert Dubrow[1,2]

[1]Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, USA
[2]Department of Epidemiology and Public Health, Yale School of Medicine, New Haven, CT, USA
[3]Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA
[4]Department of General Internal Medicine, Veterans Affairs Healthcare System, West Haven, CT, USA
[5]Infectious Diseases Section, Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX, USA
[6]Department of Medicine, Baylor College of Medicine, Houston, TX, USA
[7]Medical Specialty Care Service Line, Atlanta Veterans Affairs Medical Center, Atlanta, GA, USA
[8] Department of Medicine, Emory University School of Medicine, Atlanta, GA, USA
[9]Department of Medicine, Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, CA, USA
[10]Department of Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA
[11]Section of Infectious Diseases, Washington DC Veterans Affairs Medical Center, Washington, DC, USA
[12]Department of Medicine, George Washington University School of Medicine and Health Sciences, Washington, DC, USA
[13]Department of Medicine, James J. Peters Veterans Affairs Medical Center, Bronx, NY, USA
[14]Department of Medicine, Icahn School of Medicine, Mt. Sinai, New York, NY, USA
[15]Office of Patient Care Services, Department of Veterans Affairs, Washington, DC, USA
[16]Hematology-Oncology Service, Durham Veterans Affairs Medical Center, Durham, NC, USA
[17]Division of Medical Oncology, Duke University Medical Center, Durham, NC, USA

Abstract

Background: Given the growing interest in the cancer burden in persons living with HIV/AIDS, we examined the validity of data sources for cancer diagnoses (cancer registry versus International Classification of Diseases, Ninth Revision [ICD-9 codes]) and compared the association between HIV status and cancer risk using each data source in the Veterans Aging Cohort Study (VACS), a prospective cohort of HIV-infected and uninfected veterans from 1996 to 2008.

Methods: We reviewed charts to confirm potential incident cancers at four VACS sites. In the entire cohort, we calculated cancer-type-specific age-, sex-, race/ethnicity-, and calendar-period-standardized incidence rates and incidence rate ratios (IRR) (HIV-infected versus uninfected). We calculated standardized incidence ratios (SIR) to compare VACS and Surveillance, Epidemiology, and End Results rates.

Results: Compared to chart review, both Veterans Affairs Central Cancer Registry (VACCR) and ICD-9 diagnoses had approximately 90% sensitivity; however, VACCR had higher positive predictive value (96% versus 63%). There were 6,010 VACCR and 13,386 ICD-9 incident cancers among 116,072 veterans. Although ICD-9 rates tended to be double VACCR rates, most IRRs were in the same direction and of similar magnitude, regardless of data source. Using either source, all cancers combined, most viral-infection-related cancers, lung cancer, melanoma, and leukemia had significantly elevated IRRs. Using ICD-9, eight additional IRRs were significantly elevated, most likely due to false positive diagnoses. Most ICD-9 SIRs were significantly elevated and all were higher than the corresponding VACCR SIR.

Conclusions: ICD-9 may be used with caution for estimating IRRs, but should be avoided when estimating incidence or SIRs. Elevated cancer risk based on VACCR diagnoses among HIV-infected veterans was consistent with other studies.

Keywords: Neoplasms; Registries; International Classification of Diseases; HIV Infections

Abbreviations: ICD-9: International Classification of Diseases Ninth Revision; IRR: Incidence Rate Ratio; SIR: Standardized Incidence Ratio; VACCR: Veterans Affairs Central Cancer Registry; VACS: Veterans Aging Cohort Study

Introduction

While the incidence of AIDS-defining cancers (ADC) has decreased among persons living with HIV/AIDS (PLWHA) in the antiretroviral therapy era, the number of non-AIDS-defining cancer (NADC) diagnoses is increasing, mainly due to the growth and aging of the PLWHA population [1]. Furthermore, NADC risk among PLWHA compared to uninfected persons is elevated [2-4]. Given the growing interest in the cancer burden in PLWHA, we sought to examine how different sources of cancer diagnosis data influence epidemiologic analyses.

Studies of cancer incidence among PLWHA have used International

Classification of Diseases, Ninth Revision (ICD-9) codes[5]in claims databases to identify cancer diagnoses [6-8]. Claims data are primarily used for administrative purposes, such that their accuracy for epidemiologic investigations is variable. Investigators have assessed the validity of various algorithms usingICD-9 diagnosis and/or procedure codes for ascertaining incident cancer cases, and have observed varying

sensitivity and positive predictive value (PPV) [9-14], with the expected tradeoff between the two [9,11,14]. Reasons for false positive incident cancer diagnoses include ascertainment of prevalent cancers, incorrect coding of cancer site, benign or in situ cancers incorrectly classified as malignant, and rule out or provisional diagnoses [9,11,14]. Although the false positive rate may be reduced using procedure codes as evidence of cancer treatment [9,11,14], systematic differences between patients who do and do not receive treatment may introduce bias. Such bias would be of particular concern for comparisons of cancer incidence in HIV-infected and uninfected persons, since HIV-infected cancer patients are less likely than uninfected patients to be treated [15,16].

Other studies of cancer incidence among PLWHA have used cancer registries to identify cancer diagnoses [1,17,18]. Registries aim to collect standardized information on all cancer diagnoses that occur within a defined population, such as residents of a defined geographic area [19,20]. They require substantial resources, but meet stringent quality standards [21].

To our knowledge, the impact on the results of epidemiologic analyses of using registry versus claims data for cancer diagnoses has not been studied. To assess this impact on studies of cancer incidence in HIV-infected versus uninfected persons, we compared findings using registry versus claims data to identify incident cancer diagnoses in the Veterans Aging Cohort Study (VACS), which includes HIV-infected veterans and demographically-matched uninfected veterans in the US. Previously, ICD-9 code diagnoses were used to study cancer incidence in this cohort [7,22]; subsequently VACS was linked to the Veterans Affairs Central Cancer Registry (VACCR), which seeks to identify all cancer cases diagnosed or treated at Veterans Health Administration (VHA) medical centers [23]. We hypothesized that VACCR diagnoses would have greater validity than ICD-9 code diagnoses and that the two data **sources** would lead to different study conclusions.

Materials and Methods

VACS consists of two cohorts. The larger VACS Virtual Cohort (VACS-VC) is based on national VHA databases (e.g., demographic, vital status, inpatient and outpatient encounters with associated diagnosis and procedure codes, pharmacy, laboratory results) with no patient-contact [24]. VACS-VC enrolls HIV-infected veterans when they begin HIV care in the VHA system and two matched uninfected patients also in VHA care. The VACS Eight Site Study (VACS-8) [25] recruits HIV-infected patients at Infectious Disease clinics(a subset of VACS-VC) and frequency-matched uninfected patients from the General Internal Medicine clinics at eight VHA facilities. VACS-8 subjects complete surveys and provide consent to access medical charts. In both cohorts, subjects are matched on age, race/ethnicity, gender, and clinical site.VHA Connecticut Healthcare System and Yale University Institutional Review Boards have approved both cohorts and have granted VACS-VC a waiver of informed consent.

VACCR, which adheres to North American Association of Central Cancer Registries standards, maintains a database of cancer cases diagnosed or treated at the VHA, aggregated from local VHA medical center registries [23]. We linked VACS and VACCR to ascertain incident cancer cases. We then mapped ICD for Oncology, third edition (ICD-O-3) [26] topography and morphology codes from VACCR records and ICD-9 cancer diagnosis codes from VHA inpatient and outpatient encounter files to specific cancer types, consistent with Surveillance, Epidemiology, and End Results (SEER) recoding algorithms [27,28]. We required ICD-9 code diagnoses to be associated with at least one inpatient or two outpatient encounters on separate dates for the same cancer type, similar to methods used in prior VACS analyses [7,22,29]. The diagnosis date was the date provided by VACCR for VACCR diagnoses or the earliest ICD-9 code date for ICD-9 code diagnoses.

We ran parallel analyses to compare the findings using VACCR versus ICD-9 codes as the cancer diagnosis source. First, we assessed the validity of each data source for cancer-type-specific diagnoses using medical chart review as the gold standard; next, we compared HIV-status-specific, type-specific incidence rates and the association of HIV infection with cancer risk; and finally, we calculated HIV-status-specific, type-specific standardized incidence ratios (SIRs) by comparing observed and expected numbers of cases based on SEER rates. Tables for IRR and SIR analysis results include ADC, cancer types established to be associated with HIV, and select common NADC. The appendix includes complete results for all specific cancer types. Statistical analyses were performed using SAS version 9.2 [30].We defined statistical significance as p<0.05 (two-sided).

Validity of VACCR and ICD-9 code diagnoses by chart review

We identified all potential incident cancer cases (excluding non-melanoma skin cancers) from 1999-2008 for VACS-8 subjects at four VACS-8 sites with available electronic medical charts (Los Angeles, Pittsburgh, Bronx, Houston). Potential cases included all diagnoses identified from four sources: ICD-9 codes, VACCR, local VHA hospital cancer registries, or electronic medical record text searches for malignancy-related terms ("malignant", "cancer", "lymphoma", "carcinoma", etc.). We then adjudicated each potential case from the four sources by reviewing pathology, cytology, operative, radiology, and progress reports, as well as other records from the electronic medical chart. We compared sensitivity, specificity, PPV, and negative predictive value between VACCR and ICD-9 code diagnoses for all cancers combined and for specific cancer types, overall and by HIV status.

Impact of VACCR and ICD-9 code diagnoses on cancer incidence rates and rate ratios

Using VACS-VC data from 1996-2008, we compared the impact of the two data sources on HIV-status-specific cancer incidence rates and on rate ratios (HIV-infected vs. uninfected). We conducted each analysis for all cancers combined and for specific cancer types. We calculated observation time from 180 days post-baseline (to remove prevalent cases) to the date of cancer diagnosis, death, last healthcare utilization date, or December 31, 2008. In the all-cancer analysis, we used the first incident primary cancer diagnosis per subject as the endpoint. For specific cancer types, the endpoint was the first type-specific diagnosis.

We calculated age-, sex-, race/ethnicity-, and calendar-period-standardized incidence rates (IR) using the person-year distribution of the entire cohort as the standard weights [31]. We used exact person-time data calculation methods where time-dependent variables (i.e., age, calendar period) are classified at each day of observation [32]. The categories used for standardization were the same as those presented in Table 1.To quantify the association of HIV with cancer risk, we calculated incidence rate ratios (IRR) (HIV-infected versus uninfected) and 95% confidence intervals (CI) [31]. For the remainder of the text, IR and IRR refer to calculations using standardized rates. To examine the effects of missed cases due to healthcare outside the VHA system, we used the one inpatient/two outpatient encounter algorithm to count additional ICD-9 code cancer diagnoses from Medicare and Medicaid encounters to recalculate IRs and IRRs.

Impact of VACCR and ICD-9 code diagnoses on SIRs

We calculated SIRs with Wald confidence limits [33] to compare cancer incidence in VACS with the SEER program cancer registries, which are representative of the US population [19]. For each cancer type, we multiplied age-, sex-, race/ethnicity-, and calendar-period-specific SEER incidence rates [34] obtained using SEER*Stat version 8.0.4 [35] by the stratified person-time in VACS-VC to calculate the expected number of cases. We selected the first type-specific diagnosis for type-specific SEER incidence rates and the first cancer diagnosis of any type for the all-cancer SEER incidence rates.

Results

There were 124,936 VACS-VC subjects enrolled during 1996-2008, of whom 116,072 contributed observation time to this investigation

(i.e., had observation time after 180 days post-baseline). Subjects who contributed observation time were predominantly male (98%); about half were non-Hispanic black and about two-fifths were non-Hispanic white; approximately 70% were 35-54 years of age at baseline; and more than half entered the cohort in 1996-1999 (Table 1). The median observation time was 6.0 years for HIV-infected subjects and 7.7 years for uninfected subjects. The medical chart review from the four VACS-8 sites included 3,222 subjects, whose baseline characteristics were similar to VACS-VC subjects, except that about 90%of VACS-8 subjects were enrolled during 2000-2005.

Validity of VACCR and ICD-9 code diagnoses by chart review

Medical chart review confirmed a total of 252 incident cancers in the four VACS-8 sites (Table 2). VACCR identified 229 potential cancers, and ICD-9 codes identified 369 potential cancers. Using chart review as the gold standard, overall VACCR type-specific sensitivity was slightly lower than ICD-9 code sensitivity(87% versus92%), but overall VACCR type-specific PPV was much higher(96% versus 63%), with 136 false positive ICD-9 code diagnoses, but only 10 false positive VACCR diagnoses. When we excluded 28 "cancers of unspecified site," theICD-9 code PPV remained low (68%).

VACCR diagnosis sensitivity and PPV did not meaningfully differ by HIV status (Table 3). ICD-9 code diagnosis sensitivity also did not differ by HIV status, but PPV was lower for HIV-infected subjects (59% versus 69%). VACCR sensitivity was below 90% for four of the nine common cancer types shown in Table 3 (anal, liver, lung, Kaposi sarcoma), whereas ICD-9 code sensitivity was below 90% for only two of these cancer types (anal, leukemia). However, VACCR PPV was at least 90% for all of the common cancer types except oral cavity and pharynx, whereas ICD-9 code PPV was below 90% for all but non-Hodgkin lymphoma. In all cases, PPV was lower using ICD-9 code than VACCR; the lowest ICD-9 code PPVs were for oral cavity and pharynx (23%) and colorectal (34%) cancers.

We compared the cancer diagnosis dates determined by the VACCR and ICD-9 codes (date of first code for the cancer type of interest) with the diagnosis date determined by the gold standard chart review. The VACCR diagnosis date was identical to the gold standard diagnosis date for 192 (88%) of the 219 cancer diagnoses identified by VACCR and confirmed by chart review. Of the 27 VACCR diagnosis dates that were incorrect, 24 (89%) were within one year of the gold standard date (median: 15 days, interquartile range: 4-54 days). The ICD-9 code diagnosis date was identical for 166 (71%) of the 233 cancer diagnoses identified by ICD-9 code and confirmed by chart review. Similar to VACCR, of the 67 ICD-9 diagnosis dates that were incorrect, 60 (90%)

Covariate		HIV status			
		HIV+ (N=38,123)		HIV- (N=77,949)	
		N	(%)	N	(%)
Sex	Female	905	(2)	1,831	(2)
	Male	37,218	(98)	76,118	(98)
Race/ethnicity	Non-Hispanic white	14,792	(39)	31,149	(40)
	Non-Hispanic black	18,593	(49)	36,738	(47)
	Hispanic	2,741	(7)	6,055	(8)
	Other/unknown	1,997	(5)	4,007	(5)
Age (years)	<30	1,599	(4)	2,915	(4)
	30-34	2,789	(7)	5,415	(7)
	35-39	5,099	(13)	10,094	(13)
	40-44	7,743	(20)	15,578	(20)
	45-49	8,310	(22)	17,133	(22)
	50-54	5,700	(15)	12,023	(15)
	55-59	3,182	(8)	6,759	(9)
	60-64	1,789	(5)	3,871	(5)
	≥65	1,912	(5)	4,161	(5)
Calendar-period	1996-1999	21,493	(56)	45,393	(58)
	2000-2002	6,698	(18)	13,750	(18)
	2003-2005	6,034	(16)	11,720	(15)
	2006-2008	3,898	(10)	7,086	(9)

Table 1: Baseline characteristics of VACS Virtual Cohort subjects who contributed observation time.

		Chart review			
		Incident cancer	No cancer	Total	
Source of diagnosis	**VACCR**				
	Incident cancer	219	10	229	Sensitivity=219/252=87%
	No cancer	33†	2,982	3,015	Specificity=2982/2992=99.67%
	Total	252	2,992	3,244	PPV =219/229=96% NPV =2982/3015=99%
	ICD-9				
	Incident cancer	233	136	369	Sensitivity=233/252=92%
	No cancer	19†	2,926	2,945	Specificity =2926/3062=96%
	Total	252	3,062	3,314	PPV =233/369=63% NPV =2926/2945=99%

PPV=Positive Predictive Value, NPV=Negative Predictive Value
*For subjects with >1 potential cancer, each cancer was counted separately. Subjects with no potential cancers were counted once. ICD-9 codes identified more subjects with >1 potential cancer, resulting in a higher total.
†False negative diagnoses include those with no cancer diagnosis according to VACCR or ICD-9 code, respectively, as well as those with incorrect coding of the cancer type (two VACCR diagnoses and nine ICD-9 code diagnoses).

Table 2: Validity of Veterans Affairs Central Cancer Registry (VACCR) versus ICD-9 code cancer-type-specific diagnoses, with medical chart review as the gold standard*.

	VACCR		ICD-9	
	Sensitivity	**PPV**	**Sensitivity**	**PPV**
HIV+	124/144=86%	124/131=95%	131/144=91%	131/221=59%
HIV-	95/108=88%	95/98=97%	102/108=94%	102/148=69%
Oral cavity and pharynx	3/3=100%	3/6=50%	3/3=100%	3/13=23%
Colorectal	10/11=91%	10/11=91%	10/11=91%	10/29=34%
Anal	12/16=75%	12/13=92%	13/16=81%	13/20=65%
Liver	19/22=86%	19/19=100%	20/22=91%	20/24=83%
Lung	38/48=79%	38/38=100%	45/48=94%	45/56=80%
Prostate	51/56=91%	51/52=98%	52/56=93%	52/62=84%
Non-Hodgkin lymphoma	20/21=95%	20/20=100%	21/21=100%	21/23=91%
Leukemia	9/9=100%	9/10=90%	8/9=89%	8/10=80%
Kaposi sarcoma	10/12=83%	10/10=100%	12/12=100%	12/15=80%
Other specified cancers†	45/51=88%	45/46=98%	47/51=92%	47/89=53%
Cancers of unspecified site	2/3=67%	2/4=50%	2/3=67%	2/28=7%

*Cancer types with at least 10 VACCR or ICD-9 code diagnoses
†Cancer-type specific sensitivity and PPV for the following cancer types combined: esophagus, stomach, small intestine, biliary tract, pancreas, peritoneum and retroperitoneum, larynx, pleura, melanoma, breast, uterus or corpus, penis, testicular, bladder, kidney, brain and nervous system, thyroid, Hodgkin lymphoma, multiple myeloma

Table 3: Sensitivity and positive predictive value (PPV) of Veterans Affairs Central Cancer Registry (VACCR) and ICD-9 code diagnoses by HIV status and for common cancer types* (chart review as gold standard).

Cancer type	HIV status	VACCR				VHA ICD-9			
		# cases	**IR**	**IRR**	**95% CI**	**# cases**	**IR**	**IRR**	**95% CI**
All cancers*	+	2,421	1,103.0	1.93	(1.82, 2.05)	4,282	2,119.6	1.95	(1.86, 2.05)
	-	3,194	571.2			5,525	1,087.0		
Oral cavity & pharynx	+	120	51.3	1.30	(1.03, 1.64)	286	123.6	1.65	(1.40, 1.94)
	-	227	39.4			427	75.0		
Colorectal	+	124	54.8	1.22	(0.97, 1.52)	376	162.5	1.99	(1.70, 2.32)
	-	264	45.1			471	81.8		
Anal	+	149	62.0	28.14	(16.06, 49.32)	281	118.1	20.78	(14.24, 30.31)
	-	13	2.2			33	5.7		
Liver	+	173	75.9	3.27	(2.50, 4.29)	272	118.9	2.70	(2.21, 3.31)
	-	139	23.2			259	44.0		
Lung	+	445	195.8	1.87	(1.63, 2.14)	647	284.9	1.84	(1.64, 2.06)
	-	614	104.9			894	155.0		
Melanoma	+	46	19.5	2.21	(1.40, 3.49)	110	47.3	1.66	(1.27, 2.17)
	-	51	8.8			162	28.5		
Cervical	+	0	0.0	-	(-)	10	4.0	5.57	(1.44, 21.54)
	-	0	0.0			4	0.7		
Prostate	+	396	181.1	0.96	(0.85, 1.07)	588	276.4	0.98	(0.89, 1.08)
	-	1,085	188.9			1,555	281.9		
Penis	+	11	4.8	4.59	(1.38, 15.27)	24	10.0	5.11	(2.19, 11.94)
	-	6	1.0			11	2.0		
Hodgkin lymphoma	+	75	32.3	9.08	(5.09, 16.19)	136	58.4	6.63	(4.50, 9.78)
	-	20	3.6			49	8.8		
Non-Hodgkin lymphoma	+	315	133.8	6.93	(5.35, 8.96)	627	271.1	6.68	(5.57, 8.00)
	-	113	19.3			229	40.6		
Leukemia	+	51	22.2	2.02	(1.33, 3.08)	90	39.9	1.45	(1.10, 1.92)
	-	63	11.0			156	27.6		
Kaposi sarcoma	+	227	94.4	519.61	(229.02, 1,178.9)	528	223.3	180.31	(115.18, 282.26)
	-	1	0.2			7	1.2		

*Includes cancers of unspecified type
Table 4: Age-, sex-, race/ethnicity-, and calendar-period-standardized incidence rates (IR, per 100,000 person-years), incidence rate ratios (IRR) and 95% confidence intervals (95% CI) for Veterans Affairs Central Cancer Registry (VACCR) diagnoses and Veterans Health Administration (VHA) ICD-9 code diagnoses for selected cancer types.

were within one year of the gold standard date (median: 28 days, interquartile range: 6-116 days).

Impact of VACCR and ICD-9 code diagnoses on cancer incidence rates and rate ratios

In the VACS-VC, there were 6,010 VACCR incident cancer diagnoses (5,615 first primary cancers of any type) and 13,386 ICD-9 code incident cancer diagnoses (9,807 first primary cancers of any type). All-cancer IRs per 100,000 person-years were about twice as high using ICD-9 code (IR_{HIV+}=2,120, IR_{HIV-}=1,087) compared to VACCR (IR_{HIV+}=1,103, IR_{HIV-}=571) diagnoses, regardless of HIV status (Table 4). For a few cancer types (i.e., trachea and mediastinum, bone and joint, soft tissue, eye, and brain and nervous system cancers), number of diagnoses and consequent IRs were on the order of five- to fifty-fold higher using ICD-9 code compared to VACCR (Appendix Table A1). The most extreme instance was cervical cancer, with no VACCR cases, but 14 ICD-9 code cases (Table 4). VACCR indicated that four of these cases were intraepithelial (i.e., in situ) cancers. Furthermore, a search for procedure codes (e.g., hysterectomy) and pharmacy records (e.g., cisplatin) that would suggest treatment of invasive cervical cancer provided insufficient evidence to definitely support an incident invasive cervical cancer diagnosis among any of the 14 ICD-9 code cases.

The all-cancer IRRs (HIV-infected vs. uninfected) were 1.93 (95% CI: 1.82, 2.05) for VACCR and 1.95, (95% CI: 1.86, 2.05) for ICD-9 code (Table 4). Most type-specific IRRs were in the same direction and of similar magnitude, regardless of data source. Using either VACCR or ICD-9 code diagnoses, most viral-infection-related cancers (oral cavity and pharynx, anal, liver, and penis cancers; Hodgkin and non-Hodgkin lymphoma; and Kaposi sarcoma), lung cancer, melanoma, and leukemia had significantly elevated IRRs. Using ICD-9 code diagnoses, colorectal, biliary tract, nasal cavity and middle ear, bone and joint, soft tissue, cervical, brain and nervous system cancers, and multiple myeloma were also significantly elevated (Table 4; Appendix Table A1). Among ADC, the magnitude of significantly elevated IRRs ranged from about 7 for non-Hodgkin lymphoma to 180 (ICD-9) to 520 (VACCR) for Kaposi sarcoma. Significantly elevated NADC IRRs ranged from 1.30 (oral cavity and pharynx cancer) to almost 30 (anal cancer) using VACCR diagnoses and from 1.45 (leukemia) to about 21 (anal cancer) using ICD-9 code diagnoses.

In our secondary analysis of outside utilization with Medicare/Medicaid, we found that 48% of HIV-infected subjects and 36% of uninfected subjects accessed Medicare/Medicaid services between 1997 and 2008. In the all-cancer analysis, the addition of Medicare/Medicaid diagnoses increased ICD-9 code first primary cancer diagnoses of any type by 19% for HIV-infected subjects, but only by 11% for uninfected subjects (Appendix Table A1). Therefore, adding Medicare/Medicaid ICD-9 code diagnoses to VHA ICD-9 code diagnoses resulted in higher IRRs (for all cancers combined and for specific cancer types). However, most results were qualitatively similar to results using VHA ICD-9 code diagnoses alone.

Impact of VACCR and ICD-9 code diagnoses on SIRs

The greater number of observed ICD-9 code diagnoses compared to VACCR diagnoses resulted in more censoring based on cancer

Cancer type	HIV status	VACCR				VHA ICD-9			
		Observed # cases	Expected # cases	SIR	(95% CI)	Observed # cases	Expected # cases	SIR	(95% CI)
All cancers†	+	2,421	1,407	1.72	(1.65, 1.79)	4,282	1,277	3.35	(3.25, 3.45)
	-	3,194	3,669	0.87	(0.84, 0.90)	5,525	3,401	1.62	(1.58, 1.67)
Oral cavity & pharynx	+	120	61	1.96	(1.64, 2.35)	286	61	4.71	(4.19, 5.29)
	-	227	154	1.48	(1.30, 1.68)	427	153	2.79	(2.53, 3.06)
Colorectal	+	124	164	0.76	(0.64, 0.90)	376	162	2.31	(2.09, 2.56)
	-	264	423	0.62	(0.55, 0.70)	471	420	1.12	(1.02, 1.23)
Anal	+	149	6	23.05	(19.63, 27.07)	281	6	43.56	(38.75, 48.96)
	-	13	16	0.82	(0.48, 1.42)	33	16	2.09	(1.49, 2.94)
Liver	+	173	46	3.80	(3.28, 4.41)	272	45	5.98	(5.31, 6.74)
	-	139	118	1.18	(1.00, 1.39)	259	118	2.20	(1.95, 2.48)
Lung	+	445	216	2.06	(1.88, 2.26)	647	214	3.02	(2.79, 3.26)
	-	614	564	1.09	(1.01, 1.18)	894	561	1.59	(1.49, 1.70)
Melanoma	+	46	44	1.04	(0.78, 1.39)	110	44	2.50	(2.07, 3.01)
	-	51	110	0.46	(0.35, 0.61)	162	109	1.48	(1.27, 1.73)
Cervical	+	0	1	0.00	(-)	10	1	13.41	(7.21, 24.92)
	-	0	2	0.00	(-)	4	2	2.32	(0.87, 6.18)
Prostate	+	396	571	0.69	(0.63, 0.77)	588	558	1.05	(0.97, 1.14)
	-	1,085	1,496	0.73	(0.68, 0.77)	1555	1456	1.07	(1.02, 1.12)
Penis	+	11	2	5.91	(3.27, 10.67)	24	2	12.90	(8.64, 19.24)
	-	6	5	1.24	(0.56, 2.75)	11	5	2.27	(1.26, 4.10)
Hodgkin lymphoma	+	75	9	8.24	(6.57, 10.33)	136	9	14.97	(12.65, 17.71)
	-	20	21	0.93	(0.60, 1.44)	49	21	2.29	(1.73, 3.02)
Non-Hodgkin lymphoma	+	315	67	4.69	(4.20, 5.23)	627	67	9.42	(8.71, 10.19)
	-	113	169	0.67	(0.55, 0.80)	229	169	1.36	(1.19, 1.54)
Leukemia	+	51	35	1.44	(1.10, 1.90)	90	35	2.55	(2.07, 3.14)
	-	63	90	0.70	(0.54, 0.89)	156	90	1.73	(1.48, 2.02)
Kaposi sarcoma	+	227	8	28.07	(24.65, 31.97)	528	8	65.92	(60.53, 71.79)
	-	1	18	0.05	(0.01, 0.39)	7	18	0.38	(0.18, 0.81)

*Expected numbers of cases are rounded to the nearest whole number. SIR calculations used the exact expected number of cases (unrounded).
†Includes cancers of unspecified type
Table 5: Observed and expected numbers of cancer cases, standardized incidence ratios (SIR), and 95% confidence intervals (95% CI) for Veterans Affairs Central Cancer Registry (VACCR) diagnoses and Veterans Health Administration (VHA) ICD-9 code diagnoses for selected cancer types*.

diagnosis date for ICD-9 code diagnoses, yielding less observation time and consequently a lower number of expected cancer cases based on SEER rates (HIV+: 1,277 versus 1,407 cases; HIV-: 3,401 versus 3,669 cases) (Table 5). The all-cancer SIR for HIV-infected subjects was almost twice as high for ICD-9 code diagnoses (SIR=3.35; 95% CI: 3.25, 3.45) than for VACCR diagnoses (SIR=1.72; 95% CI: 1.65, 1.79). For uninfected subjects, VACCR identified fewer cancers than expected (SIR=0.87, 95% CI: 0.84, 0.90), but ICD-9 code diagnoses identified more cancers than expected (SIR=1.62, 95% CI: 1.58, 1.67).

Using VACCR diagnoses, with the exception of melanoma, cancer types with significantly elevated IRRs (most viral-infection-related cancers, lung cancer, and leukemia) in HIV-infected subjects also had significantly elevated SIRs (Table 5). Biliary tract, nasal cavity and middle ear, larynx, and pleura cancers were also significantly elevated (Appendix Table A2). For uninfected subjects, SIRs were significantly elevated for oral cavity and pharynx, esophagus, larynx, lung, and pleura cancers, but the majority of SIRs were less than or around 1.0.

Using ICD-9 code diagnoses, most SIRs for both HIV-infected and uninfected subjects were significantly elevated (Table 5; Appendix Table A2). Furthermore, all of the ICD-9 SIRs were greater than the corresponding VACCR SIRs. The strikingly higher number of ICD-9 code than VACCR observed cases for trachea and mediastinum, bone and joint, soft tissue, brain and nervous system, and cervical cancers resulted in markedly higher SIRs by ICD-9 code than by VACCR.

Discussion

This investigation not only reviewed the validity of cancer registry and ICD-9 code diagnoses compared to medical chart review, but also examined the impact of these data sources in a study of the association between HIV status and cancer risk. As expected, we found VACCR diagnoses to be more valid than ICD-9 code diagnoses. With chart review as gold standard, we found that both VACCR and ICD-9 code diagnoses exhibited high sensitivity, missing only about 10% of cancer cases. However, the validity of the two data sources diverged when we evaluated PPV, which was appreciably lower using ICD-9 codes. Furthermore, the VACCR cancer diagnosis date was more accurate than the ICD-9 code diagnosis date, with 88% of VACCR diagnosis dates, but only 71% of ICD-9 diagnosis dates, in exact agreement with the gold standard. Based on these results, we can adopt VACCR as our preferred data source for cancer epidemiologic analyses, and can predict that the low PPV (and resultant high proportion of false positive cases) for ICD-9 code diagnoses will lead to artifactually inflated estimates of IRs and SIRs.

Our analyses using VACCR diagnoses were broadly consistent with previous findings [2,3], providing face validity for our adoption of VACCR diagnoses in future analyses. First, the incidence of the two main ADC, Kaposi sarcoma and non-Hodgkin lymphoma, was appreciably elevated among PLWHA as shown by both IRRs and SIRs. Second, our all-cancer IRR, as well as our SIR for PLWHA, were both almost two, similar to the result for all-NADC combined in a meta-analysis of the incidence of NADC in PLWHA compared to the general population [3]. Third, the incidence of oral cavity and pharynx, anal, liver, lung, and penis cancers, Hodgkin lymphoma, and leukemia, NADCs well-established to be elevated among PLWHA [2,3], was significantly elevated among PLWHA according to both IRRs and SIRs.

Consistent with our prediction, the all-cancer IR was about two-fold higher using ICD-9 code than VACCR diagnoses. However, for some cancer types (e.g., bone and joint and brain and nervous

system), the number of diagnoses was many-fold higher using ICD-9 codes, resulting in IRs (and SIRs) that were biased upward many-fold. Interestingly, in spite of the inflated IRs using ICD-9 code diagnoses, many type-specific IRRs were similar across the two data sources, apparently due to similar type-specific multiplicative inflation factors (ICD-9 IRs versus VACCR IRs) across HIV status. Consistent with this observation, the 2009 VACS study that used ICD-9 code cancer diagnoses reported IRs that we now know to be inflated. However, IRRs for anal, liver, lung, and prostate cancers, Hodgkin lymphoma, non-Hodgkin lymphoma, and Kaposi sarcoma were of similar magnitude and significance to the VACCR IRRs reported here [7]; only the cervical cancer IRR was notably erroneous. Furthermore, in the current study, IRRs for other cancer types differed across the two data sources. For example, for colorectal cancer, which is generally found not to have elevated incidence in PLWHA, the VACCR IRR was 1.22 (95% CI: 0.97, 1.52), whereas the ICD-9 code IRR was 1.99 (95% CI: 1.70, 2.32) (Table 4). Finally, for other cancer types, such as biliary tract and nasal cavity and middle ear, the magnitude of the IRR was similar across the two data sources, but the IRR based on ICD-9 code diagnoses was statistically significant, whereas the IRR based on VACCR diagnoses was not.

The effect of low PPV and large number of false positive ICD-9 code diagnoses was more evident in our SIR analysis, with ICD-9 code diagnoses resulting in inflated SIRs, significantly greater than 1.0, for most cancer types in both HIV-infected and uninfected subjects. The all-cancer SIR of 3.35 for HIV-infected subjects was unrealistically high [3]. Although we would expect cancer incidence in uninfected veterans to be higher than in the general US population due to higher prevalence of certain cancer risk factors, such as smoking [36] and hepatitis C virus infection [37], the all-cancer SIR of 1.62 was also unrealistically high. Furthermore, cancer types that appear to be unrelated to HIV-infection, including colorectal, pancreas and bladder cancers [3], exhibited significantly elevated SIRs among PLWHA.

The inflated ICD-9 code SIRs, which resulted from comparing cancer incidence rates measured by ICD-9 code in the VACS to cancer incidence rates measured by cancer registry in the general US population, highlight the danger of comparing disease risk in two populations for which the method of case ascertainment differed. On the other hand, the VACCR SIRs, which we established to be more valid, compared cancer incidence rates measured by cancer registry in both populations.

This study had appreciable strengths. Few studies have compared both registry and claims data to a medical chart review; and to our knowledge, no study has examined data source impact on epidemiologic analyses. VACS has the strengths of a very large sample size in the VACS-VC, providing generally high statistical power for epidemiologic analyses, combined with the smaller VACS-8 sample, providing in-depth medical record data for validation analyses. Thus, this study was able to review the fine detail of cancer data accuracy with chart review, while also calculating IRs, IRRs, and SIRs for many specific cancer types in the larger sample.

Our study also had limitations. Since our cohort was 98% male, we had scant statistical power to assess female cancer types and were unable to generalize our results to women. In addition, we did not review the medical chart for each of the 3,222 subjects from the four VACS-8 sites. However, we believe that the text search for malignancy-related terms should have rendered all of the subjects with potential incident cancer diagnoses.

As demonstrated in this study, analyses that use VACCR data will underestimate cancer incidence. First, although the VACCR diagnoses exhibited a relatively high sensitivity, VACCR did miss 13% of cases. Second, the all-cancer SIR for uninfected subjects was significantly below 1.0, suggesting that VACCR is missing cancer cases since we would expect cancer incidence for uninfected veterans to be greater than in the general population. Finally, we found that adding Medicare/Medicaid ICD-9 code diagnoses to VHA ICD-9 code diagnoses resulted in 19% more diagnoses among HIV-infected veterans and 11% more diagnoses among uninfected veterans. Although some of these diagnoses were certainly false positives, many would be true positives not captured by VACCR.

The underestimation of cancer incidence in the VACS-VC using VACCR diagnoses would result in SIRs that are biased downward. Given the differential outside utilization with Medicare/Medicaid apparently resulting in a greater proportion of outside cancer diagnoses among HIV-infected veterans, we can project that IRRs would be biased downward as well. Unfortunately, we were unable to directly show how outside utilization may affect IRRs based on VACCR diagnoses.

In conclusion, ICD-9 codes are a convenient source of data because of low cost and availability, especially since linking cancer registry data to HIV cohorts is not always feasible. However, the low PPV of ICD-9 code diagnoses resulted in overestimation of incident cancers cases, IRs, and SIRs, such that ICD-9 code diagnoses should not be used to estimate these measures. Although VACCR and ICD-9 code IRRs tended to be similar, there were meaningful exceptions, indicating that ICD-9 code IRR estimates should be interpreted with caution. Elevated cancer risk based on VACCR diagnoses among HIV-infected veterans was consistent with other studies. Cancer registry diagnoses should be used for epidemiologic analyses whenever feasible.

Acknowledgements

References

1. Shiels MS, Pfeiffer RM, Gail MH, Hall HI, Li J, et al. (2011) Cancer burden in the HIV-infected population in the United States. J Natl Cancer Inst 103: 753-762.

2. Dubrow R, Silverberg MJ, Park LS, Crothers K, Justice AC (2012) HIV infection, aging, and immune function: implications for cancer risk and prevention. Curr Opin Oncol 24: 506-516.

3. Shiels MS, Cole SR, Kirk GD, Poole C (2009) A meta-analysis of the incidence of non-AIDS cancers in HIV-infected individuals. J Acquir Immune Defic Syndr 52: 611-622.

4. Grulich AE, van Leeuwen MT, Falster MO, Vajdic CM (2007) Incidence of cancers in people with HIV/AIDS compared with immunosuppressed transplant recipients: a meta-analysis. Lancet 370: 59-67.

5. National Center for Health Statistics (2012) ICD-9: International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). Centers for Disease Control and Prevention.

6. Guiguet M, Boué F, Cadranel J, Lang JM, Rosenthal E, et al. (2009) Effect of immunodeficiency, HIV viral load, and antiretroviral therapy on the risk of individual malignancies (FHDH-ANRS CO4): a prospective cohort study. Lancet Oncol 10: 1152-1159.

7. Bedimo RJ, McGinnis KA, Dunlap M, Rodriguez-Barradas MC, Justice AC (2009) Incidence of non-AIDS-defining malignancies in HIV-infected versus noninfected patients in the HAART era: impact of immunosuppression. J Acquir Immune Defic Syndr 52: 203-208.

8. Ruiz M, Johnson D, Reske T, Cefalu C, Estrada J (2013) Non-AIDS-defining cancers in New Orleans. J Int Assoc Provid AIDS Care 12: 173-177.

9. Baldi I, Vicari P, Di Cuonzo D, Zanetti R, Pagano E, et al. (2008) A high positive predictive value algorithm using hospital administrative data identified incident cancer cases. J Clin Epidemiol. 61: 373-379.

10. Cooper GS, Yuan Z, Stange KC, Dennis LK, Amini SB, et al. (1999) The sensitivity of Medicare claims data for case ascertainment of six common cancers. Med Care 37: 436-444.

11. Gold HT, Do HT (2007) Evaluation of three algorithms to identify incident breast cancer in Medicare claims data. Health Serv Res 42: 2056-2069.

12. Kashner TM (1998) Agreement between administrative files and written medical records: a case of the Department of Veterans Affairs. Med Care 36: 1324-1336.

13. McClish DK, Penberthy L, Whittemore M, Newschaffer C, Woolard D, et al. (1997) Ability of Medicare claims data and cancer registries to identify cancer cases and treatment. Am J Epidemiol 145: 227-233.

14. Penberthy L, McClish D, Manning C, Retchin S, Smith T (2005) The added value of claims for cancer surveillance: results of varying case definitions. Med Care 43: 705-712.

15. Suneja G, Shiels MS, Melville SK, Williams MA, Rengan R, et al. (2013) Disparities in the treatment and outcomes of lung cancer among HIV-infected individuals. AIDS 27: 459-468.

16. Suneja G (2013) Cancer treatment disparities in HIV-infected individuals in the United States. 14th International Conference on Malignancies in AIDS and Other Acquired Immunodeficiencies. National Institutes of Health Main Campus, Bethesda, MD.

17. Silverberg MJ, Chao C, Leyden WA, Xu L, Horberg MA, et al. (2011) HIV infection, immunodeficiency, viral replication, and the risk of cancer. Cancer Epidemiol Biomarkers Prev 20: 2551-2559.

18. Engels EA, Biggar RJ, Hall HI, Cross H, Crutchfield A, et al. (2008) Cancer risk in people infected with human immunodeficiency virus in the United States. Int J Cancer 123: 187-194.

19. Hankey BF, Ries LA, Edwards BK (1999) The surveillance, epidemiology, and end results program: a national resource. Cancer Epidemiol Biomarkers Prev 8: 1117-1121.

20. Hutton MD SL, Miller DS, Weir HK, McDavid K, Hall HI (2001) Progress toward nationwide cancer surveillance: an evaluation of the National Program of Cancer Registries, 1994–1999. J Registry Manag. 28: 113-120.

21. German RR, Wike JM, Bauer KR, Fleming ST, Trentham-Dietz A, et al. (2011) Quality of cancer registry data: findings from CDC-NPCR's Breast and Prostate Cancer Data Quality and Patterns of Care Study. J Registry Manag 38: 75-86.

22. McGinnis KA, Fultz SL, Skanderson M, Conigliaro J, Bryant K, et al. (2006) Hepatocellular carcinoma and non-Hodgkin's lymphoma: the roles of HIV, hepatitis C infection, and alcohol abuse. J Clin Oncol 24: 5005-5009.

23. Zullig LL, Jackson GL, Dorn RA, Provenzale DT, McNeil R, et al. (2012) Cancer incidence among patients of the U.S. Veterans Affairs Health Care System. Mil Med 177: 693-701.

24. Fultz SL, Skanderson M, Mole LA, Gandhi N, Bryant K, et al. (2006) Development and verification of a "virtual" cohort using the National VA Health Information System. Med Care 44: S25-30.

25. Justice AC, Dombrowski E, Conigliaro J, Fultz SL, Gibson D, et al. (2006) Veterans Aging Cohort Study (VACS): Overview and description. Med Care 44: S13-24.

26. Fritz A PC, Jack A, Shanmugarathnam K, Sobin L, Parkin DM, et al. (2000) International Classification of Diseases for Oncology (ICD-O). Geneva: World Health Organization.

27. Surveillance, Epidemiology, and End Results Program (2008) Site Recode ICD-O-3/WHO 2008 Definition. Surveillance Research Program, National Cancer Institute.

28. Surveillance, Epidemiology, and End Results Program (2012) SEER Cause of Death Recode 1969+. Surveillance Research Program, National Cancer Institute.

29. Justice AC, Lasky E, McGinnis KA, Skanderson M, Conigliaro J, et al. (2006) Medical disease and alcohol use among veterans with human immunodeficiency infection: A comparison of disease measurement strategies. Med Care 44: S52-60.

30. SAS Institute Inc (2013) SAS software [computer program]. Version 9.2, Cary, NC.

31. Boyle P, Parkin DM (1991) Chapter 11. Statistical methods of registries. In: Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG, (Eds.), Cancer registration: principles and methods. Statistical methods for registries. Lyon, France: International Agency for Research on Cancer, 126-58.

32. Wood J, Richardson D, Wing S (1997) A simple program to create exact person-time data in cohort analyses. Int J Epidemiol 26: 395-399.