Nicotine & Tobacco Research, Volume 13, Number 12 (December 2011) *1233–1239*

Original Investigation

Validating Smoking Data From the Veteran's Affairs Health Factors Dataset, an Electronic Data Source

Kathleen A. McGinnis, M.S.,[1] *Cynthia A. Brandt, M.D., M.P.H.,*[2,3] *Melissa Skanderson, M.S.W.,*[1] *Amy C. Justice, M.D., Ph.D.,*[2,4] *Shahida Shahrir, M.P.H.,*[5] *Adeel A. Butt, M.D., M.S.,*[6] *Sheldon T. Brown, M.D.,*[7] *Matthew S. Freiberg, M.D., M.Sc.,*[6] *Cynthia L. Gibert, M.D.,*[8] *Matthew Bidwell Goetz, M.D.,*[9,10] *Joon Woo Kim, M.D.,*[7] *Margaret A. Pisani, M.D., M.P.H.,*[4] *David Rimland, M.D.,*[11] *Maria C. Rodriguez-Barradas, M.D.,*[12] *Jason J. Sico, M.D.,*[2,13] *Hilary A. Tindle, M.D., M.P.H.,*[6] *& Kristina Crothers, M.D.*[5]

[1] *Center for Health Equity Research and Promotion, VA Pittsburgh Healthcare System, Pittsburgh, PA*
[2] *VA Connecticut Healthcare System, West Haven, CT*
[3] *Department of Emergency Medicine, Yale University School of Medicine, New Haven, CT*
[4] *Department of Medicine, Yale University, New Haven, CT*
[5] *Department of Medicine, Harborview Medical Center, University of Washington, Seattle, WA*
[6] *Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA*
[7] *James J. Peterson VAMC, Bronx, NY*
[8] *VA Medical Center, Washington DC*
[9] *VA Greater Los Angeles Healthcare System, Los Angeles, CA*
[10] *David Geffen School of Medicine at UCLA, Los Angeles, CA*
[11] *VA Medical Center and Emory University School of Medicine, Atlanta, GA*
[12] *Infectious Diseases Section, Michael E. DeBakey VAMC and Department of Medicine, Baylor College of Medicine, Houston, TX*
[13] *Department of Neurology, Yale University School of Medicine, New Haven, CT*

Corresponding Author: Kathleen A. McGinnis, M.S., *Center for Health Equity Research and Promotion, VA Pittsburgh Healthcare System, 7180 Highland Drive (151C-H), Pittsburgh, PA 15206-4900, USA. Telephone: 412-365-4157; Fax: 412-365-4386; E-mail: kathleen.mcginnis3@va.gov*

Received April 29, 2011; accepted August 10, 2011

Abstract

Introduction: We assessed smoking data from the Veterans Health Administration (VHA) electronic medical record (EMR) Health Factors dataset.

Methods: To assess the validity of the EMR Health Factors smoking data, we first created an algorithm to convert text entries into a 3-category smoking variable (never, former, and current). We compared this EMR smoking variable to 2 different sources of patient self-reported smoking survey data: (a) 6,816 HIV-infected and -uninfected participants in the 8-site Veterans Aging Cohort Study (VACS-8) and (b) a subset of 13,689 participants from the national VACS Virtual Cohort (VACS-VC), who also completed the 1999 Large Health Study (LHS) survey. Sensitivity, specificity, and kappa statistics were used to evaluate agreement of EMR Health Factors smoking data with self-report smoking data.

Results: For the EMR Health Factors and VACS-8 comparison of current, former, and never smoking categories, the kappa statistic was .66. For EMR Health Factors and VACS-VC/LHS comparison of smoking, the kappa statistic was .61.

Conclusions: Based on kappa statistics, agreement between the EMR Health Factors and survey sources is substantial. Identification of current smokers nationally within the VHA can be used in future studies to track smoking status over time, to evaluate smoking interventions, and to adjust for smoking status in research. Our methodology may provide insights for other organizations seeking to use EMR data for accurate determination of smoking status.

Introduction

Smoking is a leading cause of mortality (Crothers et al., 2009; Ezzati & Lopez, 2003; Mokdad, Marks, Stroup, & Gerberding, 2004) and is a major risk factor for comorbidities, such as bacterial pneumonia, pulmonary disease, cardiovascular disease, and cancer (Crothers et al., 2006, 2009; Diaz et al., 2000; Kirk et al., 2007; Sudano et al., 2010; Thompson & St-Hilaire, 2010). It is important to be able to adjust analyses for smoking status, particularly when comparing outcomes between populations with varied smoking rates. Accurate determination of smoking status can also be used to efficiently and inexpensively track and monitor smoking over time in order to evaluate smoking interventions.

doi: 10.1093/ntr/ntr206
Advance Access published on September 12, 2011

The Veterans Health Administration (VHA) benefits from one of the most highly developed health information systems in the world (Corrigan, Eden, & Smith, 2002; McQueen, Mittman, & Demakis, 2004). However, research studies using electronic medical record (EMR) data have been limited in the past by the absence of valid and complete smoking information (Bedimo, McGinnis, Dullap, Rodriguez-Barradas, & Justice, 2009; Fultz, McGinnis, Skanderson, Ragni, & Justice, 2004; McAfee, Grossman, Dacey, & McClure, 2002; McGinnis et al., 2006). Although International Classification of Diseases, 9th Revision (ICD-9) smoking diagnosis codes in VHA electronic databases exist nationally, they are susceptible to underreporting (Thompson & St-Hilaire, 2010). Recently, smoking data from the VHA EMR Health Factors dataset have become available to VHA researchers.

Prior to utilizing these EMR Health Factors smoking data, we sought to first perform a validation of the Health Factors smoking data by comparing it with smoking data from other available sources for the same individuals. With validated smoking data, many future research studies will be substantially improved as they can adjust for smoking as a critical risk factor for many outcomes. Not only is this useful for investigating smoking behaviors within large VHA populations but as more health care systems convert to the EMR, this methodology for determining smoking status can serve as an example of how to validate EMR smoking data, in general.

Methods

Two sources of self-reported smoking data were available for validating the EMR Health Factors smoking data: (a) the Veterans Aging Cohort Study (VACS-8) dataset that contains data recently collected from 8 sites and (b) a subset of the VACS Virtual Cohort (VACS-VC) participants who also completed the 1999 Veterans Longitudinal Health Survey (VACS-VC/LHS) containing data from all 128 VHA sites. Both cohorts were necessary to analyze because the former comparison allowed us to examine more recent data, while the latter provided a larger sample from all sites.

Data Sources

EMR Health Factors Data

The smoking data recently became available through the VHA Corporate Data Warehouse (CDW) through which the VACS study obtained Health Factors data. The CDW is a national repository that incorporates data from the clinical and administrative systems into one standard database structure. The objective of the CDW as stated on their Veterans Affairs (VA) intranet website is "to provide data and tools to support management decision making, performance measurement, and research objectives (VA Information Resource Center [VIReC], 2010)." As of September 2009, Health Factors data from all four Regional Data Warehouses have become available for any records that exist in the Veterans Health Information Systems and Technology Architecture (VistA), the VHA EMR, with a visit date later than October 1, 1999. The CDW cautions that Health Factors data have not been standardized across the VA, so site exploration and comparisons are important (VIReC, 2010).

EMR Health Factors data are collected nationally using the clinical reminder process and are stored in the Health Factors tables within the VHA EMR databases. Clinical reminders are automated requirements that providers must complete on their patients on a regular basis. Providers are automatically prompted by the computer to ask patients about their tobacco use. The exact content of these prompts, their frequency, and possible response entries can vary by site and over time. EMR Health Factors smoking information has been collected since 1999, and our database contains information from October 1999 to May 2009. Health Factors smoking data consist of text values representing an answer to a clinical reminder or question a health care provider has asked a patient.

A team of clinicians and researchers, including a pulmonologist, reviewed text entries and created an algorithm to assign each text value entry to a category: never, former, or current by assigning variables based on face validity. For example, "CURRENT," "TOBACCO COUNSELING," "TOBACCO CESSATION MEDS OFFERED," and/or "CURRENT SMOKER," identified current smokers; "QUIT," "FORMER," and/or "PREVIOUS" and "FORMER SMOKER," identified former smokers; and "NONE," "NEVER SMOKED," "LIFETIME NONSMOKER," and/or "NEVER-SMOKER," identified never-smokers. Clinical reminders on tobacco use can incorporate cigarette smoking as well as other tobacco products like smokeless tobacco, pipes, or cigars; we did not include entries that specified use of smokeless tobacco ("CURRENT SMOKELESS TOBACCO USER" and/or "SMOKELESS TOBACCO USER") in generation of smoking status. The mapping strategies used are available on the VACS website: www.vacohort.org.

Veterans Aging Cohort 8-Site Study (VACS-8)

VACS-8 has been described in detail in previous publications (Justice, Dombrowski, et al., 2006; Justice, Erdos, et al., 2006; Justice et al., 2001). Briefly, VACS-8 is an ongoing prospective cohort study being conducted at eight VA medical facilities in the United States (Atlanta, GA; Baltimore, MD; Bronx, NY; Houston, TX; Los Angeles, CA; New York City, NY; Pittsburgh, PA; and Washington, DC). Enrollment in VACS-8 began in June 2002 and reached its initial target of 3,000 HIV-infected individuals and 3,000 HIV-uninfected controls in August 2004, though recruitment is ongoing. HIV-infected individuals are recruited from the Infectious Diseases clinics at the participating sites. HIV-uninfected controls are recruited from the General Internal Medicine clinics at the same sites and are targeted to match the demographics of the Infectious Diseases clinics by 5-year age blocks, race/ethnicity, and gender. At baseline enrollment, the individuals complete a comprehensive self-administered paper survey that includes information on tobacco use. Smoking from the VACS-8 baseline survey was coded as current if a person reported that they "now smoke cigarettes (i.e., within the past week)" or quit smoking within the last 4 weeks. A person was considered to be a former smoker if s/he reported ever smoking cigarettes for "as long as a year" and quitting more than four weeks ago. A person was considered to be a never-smoker if s/he does not currently smoke and did not ever smoke for as long as a year.

VACS Virtual Cohort (VACS-VC) Subgroup Who Completed the 1999 Large Health Study (VACS-VC/LHS)

The VACS Virtual Cohort (VACS-VC) is a national cohort of 40,594 HIV-infected and 81,188 age, race/ethnicity, gender, and clinical site–matched HIV-uninfected patients who were identified

from VHA electronic data in the fiscal years 1997–2008 using a modified existing algorithm (Fultz et al., 2006). This cohort consists of data from the Immunology Case Registry; the Pharmacy Benefits Management database, the VHA's centralized database of outpatient prescriptions; and the Decision Support System, a national database of VHA clinical and financial data, including laboratory data. The VACS-VC is updated annually. In addition, EMR Health Factors smoking data are available for VACS-VC patients.

A subgroup of the VC also completed the 1999 Large Health Survey (LHS) of veteran enrollees, which was designed to assess the health status of veterans in the VHA. A self-completed paper survey was administered to 887,775 veterans between June 1999 and January 2000 (Iqbal et al., 2008). It contains measures of health, health behaviors, and sociodemographic and economic status. Of the 40,594 HIV-infected and 81,188 HIV-uninfected VC subjects identified from 1997 to 2008, 13,250 participants also completed the 1999 LHS survey.

Smoking from the LHS survey was coded as current if the respondent reported that they "now smoke cigarettes everyday" or "some days." A person was considered to be a former smoker if they are not a current smoker and reported smoking "at least 100 cigarettes" in their entire life and "last smoked cigarettes regularly, that is, daily" one month ago or longer. A person was considered to be a never-smoker if s/he did not smoke at least 100 cigarettes over his/her lifetime and does not currently smoke.

We examined differences in performance of the EMR Health Factors data by varying several factors. We restricted the time interval allowed between EMR data and VACS-8 survey data to 1 year. We also examined results using the most recent EMR Health Factors smoking response as compared with the most frequently recorded EMR Health Factors response.

Analyses

To compare the EMR Health Factors smoking data with self-completed survey smoking data, we examined summaries of percent missing, percent correctly identified into the three groups (assuming survey data are the gold standard), sensitivity, specificity, and agreement using kappa statistics. Analyses were run overall and for each site.

The kappa statistic measure of agreement ranges from 0 to 1, with 0 representing agreement when it is what would be expected from chance alone and 1 representing perfect agreement. Landis and Koch (1977) suggest interpreting intermediate values as follows: below .00—poor, .00 to .20—slight; .21 to .40—fair; .41 to .60—moderate; .61 to .80—substantial; .81 to 1.00—almost perfect. We compared agreement using the three-level smoking variable and two dichotomous variables: ever/never smoking and current/not current smoking. For the three-level smoking variable, we ran the kappa statistics in two ways. First, agreement is either *perfect* (1) or *not perfect* (0). The second way includes weighting that acknowledges a difference between being two categories apart (never smoking and currently smoking) versus being only one category apart (never smoking and former smoking or current smoking and former smoking). If agreement is different by only one category, weighting would be .5 rather than 0.

Results

Comparison of EMR Health Factors With VACS-8 Survey Data

VACS-8 participants are predominantly black males, with a mean age of 49.8 years. Because of the study design, half are HIV infected (Table 1). Of the 6,819 VACS-8 participants, only 1% were missing VACS-8 survey smoking data and 5% were missing EMR Health Factors smoking data. Sites 1, 5, and 7 had the most missing EMR Health Factors data with 15%, 7%, and 6% missing, respectively. Site 1 had started capturing the data at a later date than the other sites.

We compared the Health Factors data with VACS-8 data in three different ways: (a) limiting Health Factors data to a 1-year timeframe around the VACS-8 baseline survey, (b) using the most recent entry in the Health Factors dataset, and (c) using the most frequent entry in the Health Factors dataset. Limiting Health Factors data to a 1-year timeframe around the VACS-8 baseline survey resulted in kappa = .69; however, there was a substantial amount of missing data for some sites (overall 40%). When we compared agreement between using the most frequent versus the most recent EMR Health Factors entry, only 5% of data are missing, and we found that agreement was better using the most frequent smoking entry (kappa = .66 vs. .57).

Overall, the EMR Health Factors smoking data and VACS-8 survey data had substantial agreement. Kappa statistics ranged from .56 to .74 for the eight sites (Table 2). Table 3 shows that of those who were never-smokers according to the VACS-8 survey,

Table 1. Demographics for the 8-Site Veterans Aging Cohort Study (VACS-8) and National VACS Virtual Cohort Subset Who Completed the 1999 Large Health Study Survey (VACS-VC/LHS)

	VACS-8	VACS-VC/LHS
N	6,819	13,689
Mean age (*SD*)	49.8 (9.4)	48.1 (10.4)
Race/ethnicity (%)		
White, non-Hispanic	22	45
Black, non-Hispanic	65	41
Hispanic	10	10
Other/unknown	4	4
Male (%)	95	98
HIV infected (%)	50	27
Smoking from VACS survey (%)		
Never	25	28
Former	26	14
Current	49	58
Smoking from Health Factors dataset (%)		
Never	22	29
Former	32	20
Current	46	51

Table 2. Smoking in the 8-Site Veterans Aging Cohort Study (VACS-8) by Site Based on VACS-8 Survey and Most Common Electronic Medical Record Health Factors Information

		VACS-8				Health factors				VACS-8 and Health factors	
		Smoking %				Smoking %					
Site	*N*	% Missing	Current	Former	Never	% Missing	Current	Former	Never	*n*	Kappa
1	1,053	0.3	39	23	38	15.2	42	11	46	890	.69
2	721	1.1	48	30	22	1.1	53	28	19	705	.69
3	919	0.4	53	23	24	0.0	53	20	28	915	.64
4	812	0.1	43	30	28	0.5	50	27	24	807	.73
5	1,090	1.7	49	28	23	7.4	63	2	35	991	.56
6	899	0.1	61	24	15	1.6	77	9	14	884	.62
7	1,126	2.8	54	22	25	5.9	69	4	28	1,029	.61
8	199	0.0	51	30	19	4.0	50	33	17	191	.74
All	6,819	1.0	49	26	25	5.0	58	14	28	6,412	.66

84% were never-smokers based on EMR Health Factors data. Of those who were current smokers on VACS-8 survey, 95% were current smokers based on Health Factors data. Of former smokers based on VACS-8 survey, 43% were former smokers based on Health Factors data. The overall kappa statistic is .66, representing substantial agreement, and the weighted kappa statistic is even higher at .74 (Table 3). When categories are collapsed into ever/never, the kappa statistic is .72 (sensitivity = 91%; specificity = 84%), and for current/not current, the kappa statistic is .75 (sensitivity = 95%; specificity = 79%).

Comparison of EMR Health Factors With National VACS-VC/LHS

Individuals in the national VACS-VC/LHS are demographically similar to subjects in VACS-8, although slightly younger and more likely to be White. Twenty-seven percent are HIV infected due to the design of the study (Table 1). Of the 13,689 individuals in the national VACS-VC/LHS subset, 0.2% of participants were missing LHS survey smoking data and 17% were missing EMR Health Factors smoking data.

We found that agreement between EMR Health Factors smoking data and LHS survey was substantial. Of those who never smoked according to the LHS survey, 82% were never-smokers based on Health Factors data. Of those who were current smokers based on LHS survey, 88% were current smokers based on Health Factors data. Of former smokers based on LHS survey, 48% were former smokers based on Health Factors data (Table 4). Kappa statistics ranged from .2 to .9 for the 128 sites, and 121 of the 128 sites (95%) had kappa statistics of .4 or higher, which represents moderate agreement or better. The overall kappa statistic was .61, representing substantial agreement. The weighted kappa statistic was even higher at .69. When categories are collapsed into ever/never, the kappa statistic is .63 (sensitivity = 87%; specificity = 82%); and for current/not current, the kappa statistic is .72 (sensitivity = 88%; specificity = 84%).

Conclusions

We compared the performance of EMR Health Factors smoking data with two different sources of self-completed survey data. We found that agreement was substantial between EMR Health Factors data and both the VACS-8 survey data and the national VACS-VC/LHS survey data. In both comparisons, the lowest agreement was for the former smoking group (43% of former smokers based on self-completed VACS-8 were former smokers based on EMR Health Factors data, and 48% of former smokers based on self-completed LHS were former smokers based on the

Table 3. Smoking From Electronic Medical Record Health Factors Data Compared With Self-report on 8-Site Veterans Aging Cohort Study (VACS-8) Survey as Gold Standard (*n* = 6,412)

	VACS-8 survey		
Health factors	Never	Former	Current
Never, *n* (%)	1,364 (84)	372 (24)	40 (1)
Former, *n* (%)	117 (7)	663 (43)	119 (4)
Current, *n* (%)	150 (9)	519 (33)	3,068 (95)
Kappa statistic = .66			
Weighted kappa statistic = .74			

Table 4. Smoking From Electronic Medical Record Health Factors Data Compared With Self-report on LHS survey for the Subset of National VACS Virtual Cohort Subset Who Completed the 1999 Large Health Study Survey (VACS-VC/LHS) as Gold Standard (*n* = 11,355)

	LHS survey		
Health factors	Never	Former	Current
Never, *n* (%)	2,136 (82)	973 (29)	179 (3)
Former, *n* (%)	246 (9)	1,612 (48)	459 (8)
Current, *n* (%)	213 (8)	760 (23)	4,777 (88)
Kappa statistic = .61			
Weighted kappa statistic = .68			

EMR Health Factors data). This is not surprising, given that former smokers are the group most likely to vacillate between smoking status groups. In addition, our survey definition of former smokers included individuals who could have quit as recently as four weeks ago; given the high rates of recidivism among recent quitters, this may also be a reason for the lower agreement in this group.

There are several strengths of this analysis. We used national data on a racially and ethnically diverse population over a long period of time and show substantial agreement between self-reported survey data and EMR data. For analyses using electronic VHA data, the EMR Health Factors smoking data will improve future research since smoking data are available for 83% of the veterans identified in the VACS-VC/LHS. For more recent years, an even higher percent of veterans have smoking data available. Furthermore, analyses using EMR Health Factors data can be performed longitudinally on a large number of individuals nationwide and do not require substantial data resources or any additional participant burden. Utilizing data from EMRs for analyses is much less costly than collecting data from surveys and biochemical testing.

There are some limitations to this analysis. The EMR Health Factors data are collected during face-to-face interactions and may be subject to underreporting. Our "gold standard" measurements for smoking are based on self-completed survey data rather than a biological assessment for smoking and may also be subject to underreporting. Another limitation is that some of the Health Factors prompts to providers are inquiring about tobacco use, which can include smokeless tobacco, pipes, and cigars and are not necessarily restricted to cigarette smoking. However, we were careful to not include Health Factors data that explicitly specified smokeless tobacco use to define smoking status.

Additionally, we acknowledge that our approach of using the most frequent Health Factors data entry can result in potential for misclassification, particularly of recent quitters as current smokers. For many health studies in which adjusting for smoking is important, this misclassification may be acceptable, especially as many of the benefits of smoking cessation take place in the long-term rather than in the short-term. We included a long timeframe to create the EMR Health Factors smoking variable based on most frequent observation, comparing it with cross-sectional report of smoking from surveys because limiting the Health Factors data to a shorter timeframe closer to the survey dates resulted in a large amount of missing Health Factors data. For this to be a meaningful validation, we chose to maximize available data. However, we do not necessarily recommend that all researchers use the most frequent observation as we did. For example, for studies that involve assessing quit rates over time, researchers may wish to use the most recent rather than the most frequent Health Factors entry, which we demonstrate performs as well or better than the most frequent Health Factors entry when available. The choice of most frequent versus most recent observation should be based on what is most relevant to a particular research question.

Despite these limitations, we found that agreement is nonetheless substantial between Health Factors data and self-reported smoking variables. Additionally, although there is variation in how the Health Factors smoking data are collected by sites, there is not substantial variation between agreement between sites. Despite the potential misclassification, we believe it is important to describe and assess the three smoking categories (current, former, and never) as many analyses will benefit from using all three categories. We additionally provide data on agreement between two categories of smokers (current vs. non-current and ever vs. never). With these results, researchers who use the Health Factors smoking data in the future can make a more informed decision on whether and how to create smoking groups based on what is relevant to their particular study.

Using Health Factors data to determine smoking status can substantially improve our estimates of smoking among VHA populations. "The State of Care for Veterans with HIV/AIDs" reports that 44% of HIV-infected veterans in care have a diagnosis of tobacco use ever and 24% have a diagnosis of tobacco use in 2008 (Center for Quality Management in Public Health, 2009). This information is based on ICD-9 diagnosis codes, which appear to underreport smoking when compared with the prevalence of current smoking by self-report on the VACS-8 survey. ICD-9 codes underestimate smoking likely because many providers do not assign these codes for smoking. These ICD-9 diagnosis estimates of ever and current smoking are also much lower than what we calculated based on EMR Health Factors data from 2008 for HIV-infected veterans in care (77% ever and 62% current smokers). A similar inconsistency between ICD-9 codes and Health Factors data was reported in a single center study of chronic obstructive pulmonary disease and tobacco use involving patients seen at the Boise VA Medical Center. The authors reported a current smoking prevalence of 14% based on ICD-9 codes compared with 39% based on Health Factors data (Thompson & St-Hilaire, 2010).

In summary, the agreement of the national VHA EMR Health Factors smoking data with previously collected self-completed survey data surpasses our expectations. Smoking information is now available for 80% of the veterans in care from fiscal years 1997 to 2008 based on our VC dataset and an even higher percent of the veterans in care if limited to more recent years. Based on kappa statistics, agreement between the EMR Health Factors smoking data and self-completed smoking data from two survey sources is substantial.

Finally, other studies can benefit from using EMR data to determine accurate smoking status. For example, these data can be used to generate a cohort of current smokers, to track change in smoking status over time, to assess the impact of smoking interventions, and to measure performance for quality improvement initiatives. Whereas VHA smoking data have been previously limited to cross-sectional data derived from time-consuming and costly manual chart reviews or surveys (Sherman, 2008), EMR Health Factors data can be retrieved efficiently, longitudinally, for less cost, and in a more comprehensive cohort of patients. In addition, this methodology for using EMR data can serve as a useful model for other health care organizations as they transition to the EMR. As new incentives and/or interventions for smoking cessation are used, EMR smoking data are an inexpensive source for evaluating subsequent changes in smoking.

Funding

Supported by VAHS HSR&D RCD 04-125-1, NIA K08 AG00826, an interagency agreement between the National Institute on Aging and the National Institute of Mental Health, and the

National Institute on Alcohol Abuse and Alcoholism (Grant No. U01-13566; ACJ); NIH/NHLBI 1R01 HL090342 (KC).

Declaration of Interests

None declared.

Acknowledgments

The contents do not represent the views of the Department of Veterans Affairs or the United States Government.

References

Bedimo, R., McGinnis, K. A., Dullap, M., Rodriguez-Barradas, M., & Justice, A. C. (2009). Incidence of non-AIDS-defining malignancies in the HIV-infected vs. non-infected Veterans in the HAART era: Impact of immunosuppression. *Journal of Acquired Immune Deficiency Syndromes*, *52*, 203–208. doi:10.1097/QAI.0b013e3181b033ab

Center for Quality Management in Public Health. (2009). *The state of care for veterans with HIV/AIDS*. Palo Alto, CA: U.S. Department of Veterans Affairs, Public Health Strategic Health Care Group, Center for Quality Management in Public Health.

Corrigan, J. M., Eden, J., & Smith, B. M. (2002). *Leadership by example: Coordinating government roles in improving healthcare quality committee on enhancing federal healthcare quality programs*. Washington, DC: National Academy Press.

Crothers, K., Butt, A. A., Gibert, C. L., Rodriguez-Barradas, M. C., Crystal, S., Justice, A. C., et al. (2006). Increased COPD among HIV-positive compared to HIV-negative veterans. *Chest*, *130*, 1326–1333. doi:10.1378/chest.130.5.1326

Crothers, K., Goulet, J. L., Rodriguez-Barradas, M. C., Gibert, C. L., Oursler, K. A., Goetz, M. B., et al. (2009). Impact of cigarette smoking on mortality in HIV-positive and HIV-negative Veterans. *AIDS Education and Prevention*, *21*, 40–53. doi:10.1521/aeap.2009.21.3_supp.40

Diaz, P. T., King, M. A., Pacht, E. R., Wewers, M. D., Gadek, J. E., Nagaraja, H. N., et al. (2000). Increased susceptibility to pulmonary emphysema among HIV-seropositive smokers. *Annals of Internal Medicine*, *132*, 369–372.

Ezzati, M., & Lopez, A. D. (2003). Estimates of global mortality attributable to smoking in 2000. *The Lancet*, *362*, 847–852. doi:10.1016/S0140-6736(03)14338-3

Fultz, S. L., McGinnis, K. A., Skanderson, M., Ragni, M. V., & Justice, A. C. (2004). Association of venous thromboembolism with human immunodeficiency virus and mortality in veterans. *American Journal of Medicine*, *116*, 420–423. doi:10.1016/j.amjmed.2003.10.011

Fultz, S. L., Skanderson, M., Mole, L. A., Gandhi, N., Bryant, K., Crystal, S., et al. (2006). Development and verification of a "virtual" cohort using the National VA Health Information System. *Medical Care*, *44*(Suppl. 2), S25–S30. doi:10.1097/01.mlr.0000223670.00890.74

Iqbal, S. U., Rogers, W., Selim, A., Qian, S., Lee, A., Ren, X. S., et al. (2008). *The veterans RAND 12 item health survey (VR-12): What is it and how it is used*. Section for Pharmaco-Outcomes and Epidemiology Center for Health Quality, Outcomes and Economic Research CHQOERs Veterans Administration Medical Center, Bedford MA and Center for the Assessment of Pharmaceutical Practices (CAPP) Boston University School of Public Health. Retrieved from: http://www.chqoer.research.va.gov/CHQOER/docs/VR12.pdf. www.chqoer.research.va.gov/CHQOER/docs/VR12.pdf.

Justice, A. C., Dombrowski, E., Conigliaro, J., Fultz, S. L., Gibson, D., Madenwald, T., et al. (2006). Veterans Aging Cohort Study (VACS): Overview and description. *Medical Care*, *44*(Suppl. 2), S13–S24. doi:10.1097/01.mlr.0000223741.02074.66

Justice, A. C., Erdos, J., Brandt, C., Conigliaro, J., Tierney, W., & Bryant, K. (2006). The Veterans Affairs Healthcare System: A unique laboratory for observational and interventional research. *Medical Care*, *44*(Suppl. 2), S7–S12. doi:10.1097/01.mlr.0000228027.80012.c5

Justice, A. C., Landefeld, C. S., Asch, S. M., Gifford, A. L., Whalen, C. C., & Covinsky, K. E. (2001). Justification for a new cohort study of people aging with and without HIV infection. *Journal of Clinical Epidemiology*, *54*, S3–S8. doi:10.1016/S0895-4356(01)00440-1

Kirk, G. D., Merlo, C., O'Driscoll, P., Mehta, S. H., Galai, N., Vlahov, D., et al. (2007). HIV infection is associated with an increased risk for lung cancers, independent of smoking. *Clinical Infectious Diseases*, *45*, 103–110. doi:10.1086/518606

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. doi:10.2307/2529310

McAfee, T., Grossman, R., Dacey, S., & McClure, J. (2002). Capturing tobacco status using an automated billing system: Steps toward a tobacco registry. *Nicotine & Tobacco Research*, *4*, 31–37. doi:10.1080/14622200210128009

McGinnis, K. A., Fultz, S. L., Skanderson, M., Conigliaro, J., Bryant, K., & Justice, A. C. (2006). Hepatocellular carcinoma and non-Hodgkin's lymphoma: The roles of HIV, hepatitis C infection, and alcohol abuse. *Journal of Clinical Oncology*, *24*, 5005–5009. doi:10.1200/JCO.2006.05.7984

McQueen, L., Mittman, B. S., & Demakis, J. G. (2004). Overview of the Veterans Health Administration (VHA) quality enhancement research initiative (QUERI). *Journal of the American Medical Informatics Association*, *11*, 339–343. doi:10.1197/jamia.M1499

Mokdad, A. H., Marks, J. S., Stroup, D. F., & Gerberding, J. L. (2004). Actual causes of death in the United States, 2000. *Journal of the American Medical Association*, *291*, 1238–1245. doi:10.1001/jama.293.3.293

Sherman, S. E. (2008). A framework for tobacco control: Lessons learnt from Veterans Health Administration. *British Medical Journal*, *336*, 1016–1019. doi:10.1136/bmj.39510.805266

Sudano, I., Spicker, L. E., Noll, G., Corti, R., Weber, R., & Luscher, T. F. (2010). Cardiovascular disease in HIV infection. *American Heart Journal*, *151*, 1147–1155. doi:10.1016/j.ahj.2005.07.030

Thompson, W. H., & St-Hilaire, S. (2010). Prevalence of chronic obstructive pulmonary disease and tobacco use in veterans at Boise Veterans Affairs Medical Center. *Respiratory Care*, *55*, 555–560.

VA Information Resource Center. (2010). "*The researchers guide to VA data*" *VHA corporate data warehouse—Description*, Retrieved from http://vaww.virec.research.va.gov/DataSources Name/CDW/CDW.htm