Validation of Veterans Affairs Electronic Medical Record Smoking Data Among Iraq- and Afghanistan-Era Veterans

Patrick S. Calhoun, Ph.D.^{1,2,3,4}, Sarah M. Wilson, Ph.D.^{1,2,3}, Jeffrey S. Hertzberg, B.S.^{2,3}, Angela C. Kirby, M.S.^{1,2,3}, Scott D. McDonald, Ph.D.^{1,5}, Paul A. Dennis, Ph.D.^{2,3}, Lori A. Bastian, M.D., M.P.H.⁶, Eric A. Dedert, Ph.D.^{2,3} The VA Mid-Atlantic MIRECC Workgroup, and Jean C. Beckham, Ph.D.^{1,2,3}

¹VA Mid-Atlantic Region Mental Illness Research, Education and Clinical Center (MIRECC), Durham VA Medical Center, Durham, NC, USA;

²Durham VA Medical Center, Durham, NC, USA; ³Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, USA; ⁴Center for Health Services Research in Primary Care, Durham VA Medical Center, Durham, NC, USA; ⁵Hunter Holmes McGuire VA Medical Center, Richmond, VA, USA; ⁶VA Connecticut Healthcare System, West Haven, CT, USA.

BACKGROUND: Research using the Veterans Health Administration (VA) electronic medical records (EMR) has been limited by a lack of reliable smoking data.

OBJECTIVE: To evaluate the validity of using VA EMR “Health Factors” data to determine smoking status among veterans with recent military service.

DESIGN: Sensitivity, specificity, area under the receiver-operating curve (AUC), and kappa statistics were used to evaluate concordance between VA EMR smoking status and criterion smoking status.

PARTICIPANTS: Veterans ($N = 2025$) with service during the wars in Iraq/Afghanistan who participated in the VA Mid-Atlantic Post-Deployment Mental Health (PDMH) Study.

MAIN MEASURES: Criterion smoking status was based on self-report during a confidential study visit. VA EMR smoking status was measured by coding health factors data entries (populated during automated clinical reminders) in three ways: based on the most common health factor, the most recent health factor, and the health factor within 12 months of the criterion smoking status data collection date.

KEY RESULTS: Concordance with PDMH smoking status (current, former, never) was highest when determined by the most commonly observed VA EMR health factor ($\kappa = 0.69$) and was not significantly impacted by psychiatric status. Agreement was higher when smoking status was dichotomized: current vs. not current ($\kappa = 0.73$; sensitivity = 0.84; specificity = 0.91; AUC = 0.87); ever vs. never ($\kappa = 0.75$; sensitivity = 0.85; specificity = 0.90; AUC = 0.87). There were substantial missing Health Factors data when restricting analyses to a 12-month period from the criterion smoking status date. Current smokers had significantly more Health Factors entries compared to never or former smokers.

CONCLUSIONS: The use of computerized tobacco screening data to determine smoking status is valid and feasible. Results indicating that smokers have

significantly more health factors entries than non-smokers suggest that caution is warranted when using the EMR to select cases for cohort studies as the risk for selection bias appears high.

KEY WORDS: smoking; cigarette use; validation; measurement. *J Gen Intern Med* 32(11):1228–34

DOI: 10.1007/s11606-017-4144-5

© Society of General Internal Medicine (outside the USA) 2017

INTRODUCTION

In the US, smoking takes a heavy toll because of tobacco-related illness, death, medical expenditures, and lost productivity.^{1–3} Military service increases risk for initiation and maintenance of cigarette smoking,^{4–8} and younger veterans endorse high rates of tobacco use. Among Veterans Affairs (VA) patients who served during conflicts in Iraq and/or Afghanistan, 50% have a lifetime history of smoking and 24% currently smoke.⁹ Accurate assessment of smoking status among this relatively young group of veterans could have significant public health implications.¹⁰ The VA electronic medical record (EMR) system is designed to estimate the prevalence of health problems, assess and improve the performance of health services, and realign system resources.¹¹ Generally, smoking status is underreported when using nicotine dependence ICD-9 codes in the list of health problems in the VA EMR.¹²

In 2006, the VA implemented a national performance measure, which required screening all outpatients for tobacco use. This measure was supported by a widespread adoption of voluntary electronic clinical reminders in the EMR to prompt annual screening and provide elements of brief advice.¹³ Although the VA has one of the most advanced EMR systems in the nation,^{14–16} there is little research that validates use of VA EMR smoking information,¹⁰ especially in younger populations of veterans who generally report the highest smoking rates.

McGinnis and colleagues¹⁰ developed an algorithm to determine smoking status based on tobacco clinical reminder

Received March 8, 2017

Revised June 26, 2017

Accepted July 18, 2017

Published online August 14, 2017

data. In comparison to smoking status collected during a research visit, there was substantial agreement (kappa statistics 0.61–0.66) between EMR data and survey results when examining current, former, and never smoking categories.¹⁰ Agreement was higher when categories were collapsed into ever/never smoking and current/not current smoking.¹⁰

To date no other published studies have validated this method to utilize VA clinical reminder data to determine smoking status. The purpose of the current study is to provide an independent replication and extension of the McGinnis and colleagues’¹⁰ method to code VA EMR “Health Factors” data for smoking status. The current study represents the first to compare VA EMR data to smoking status obtained during a confidential study visit among Iraq/Afghanistan era veterans. Additionally, given that psychiatric conditions are associated with increased smoking prevalence,^{17,18} the current study aimed to validate EMR smoking data in veterans with psychiatric disorders.^{17,18}

METHODS

Participants

The sample included 2034 US veterans who participated in the Study of Post-Deployment Mental Health (PDMH),^{19–23} an ongoing multi-site study of veterans with military service since September 11, 2001. Procedures and recruitment methods for the PDMH study have been detailed elsewhere.^{19–23} The current study included all individuals who completed self-report measures and clinical interviews on the same day between December 2005 and April 2015 and had at least one primary care visit at a participating VA medical center. Nine participants were excluded because of missing PDMH data on smoking status, resulting in a final sample of 2025 participants.

Measures

Smoking Status Criterion. Smoking status at the PDMH study visit was used as a criterion to validate EMR smoking data. Participants completed a paper or electronic questionnaire that assessed smoking status during a confidential study visit.^{24,25} “Ever smoker” was defined as those who smoked at least 100 cigarettes in their lifetime.²⁶ “Never smokers” were those who never smoked or smoked <100 cigarettes in their lifetime.²⁶ “Former smoker” was defined as those who ever smoked ≥ 100 lifetime cigarettes and who reported no past-month cigarette smoking. “Current smokers” included anyone who identified as a smoker or who reported smoking all or part of a cigarette within the past 30 days for those who ever smoked >100 cigarettes.

EMR Smoking Status. EMR Health Factors data were obtained from the VA Corporate Data Warehouse (CDW) spanning a period from 2001 to 2016. Health Factors data

are collected nationally using automated clinical reminders that healthcare providers must complete on a regular basis. The exact text, frequency, and possible responses to tobacco clinical reminders may vary by site and time. Health Factors data are available for any records that exist in the VA EMR since October 1, 1999.¹⁰ Smoking Health Factors data consist of fixed text entries representing results of smoking-focused clinical reminders. These text entries (e.g., “TOBACCO MEDS OFFERED”) were coded (Current Smoker, Former Smoker, Never Smoker, Unknown) using the methods described by McGinnis et al.¹⁰ Of the 962 text entries coded by McGinnis et al.,¹⁰ only 82 unique health factor text entries were used by medical centers in VISN-6.

Following McGinnis et al.,¹⁰ smoking status was defined in three ways using (1) the most commonly recorded EMR Health Factors response code, (2) the most recent EMR Health Factors response, and (3) the Health Factors response restricted within 12 months prior to or following their study visit. If there was ever an instance where two entries were equally “most common,” the tied entry that was recorded most recently was chosen. If a participant had multiple observations during the window around their PDMH visit, the entry closest to the PDMH visit date was used to determine smoking status from the EMR.

Psychiatric Diagnoses and EMR Medical Appointments.

The Structured Clinical Interview for DSM-IV Axis I Disorders²⁷ (SCID-IV) was used to determine current psychiatric diagnoses. Outpatient service utilization was based on VA clinic “stops” defined as a patient encounter with one or more health professionals within a particular clinic. Stop codes (three-digit codes used to classify all billable patient appointments or encounters) were used to categorize encounters as primary care or mental health using established methods.^{28–30} The number of primary care and mental health appointments was counted for a window that included 12 months prior to and 12 months following the PDMH study visit.

Analyses

Agreement on smoking status was calculated between EMR smoking data and PDMH smoking status for the full sample and stratified by psychiatric status (any mental health diagnosis vs. none). Greater than chance agreement is indicated by positive kappa values. Intermediate values of kappa can be interpreted as follows: 0.00–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; 0.81–1.00, almost perfect.³¹ Both simple kappa statistics and weighted kappa values were calculated. Weighted kappa³² was determined by applying weights that account for the fact that there is greater disagreement when results are two categories apart (e.g., never smoker and current smoker) than one category (e.g., former smoker and current smoker). Diagnostic efficiency statistics,

including sensitivity (e.g., the proportion of current smokers correctly identified as smokers by EMR data), specificity (i.e., the proportion of non-smokers correctly classified), and the estimated area under the receiver-operating characteristic curve (AUC)³³ were derived for dichotomous outcomes comparing EMR data to PDMH smoking status as the reference or “gold” standard. All analyses were conducted using SAS 9.4 (SAS Institute, Cary, NC).

RESULTS

Demographic Characteristics by Smoking Status

Participants in the PDMH study were predominately men who were exposed to combat during deployment to Iraq and/or Afghanistan (see Table 1 for sample characteristics). Current smoking was reported by 28% of the sample at the PDMH visit. Fifty-two percent were lifetime non-smokers and 20% were former smokers. Current smokers were more likely to be male, younger, White, non-married, and have fewer years of education. Current PTSD, major depressive disorder (MDD), and substance abuse/dependence were all associated with current smoking.

Three-Category Smoking Status (Never, Former, Current) Comparison

In terms of agreement with PDMH smoking status, the highest kappa value was observed when smoking status was defined by the most common VA EMR Health Factors entry ($\kappa = 0.69$) (see Table 2). Of current smokers in the PDMH study, 84% were determined to be current smokers based on the most common Health Factors data while 8% were classified as former smokers and 8% were classified as never smokers. Most never smokers were correctly classified by the most common EMR Health Factors entry (90%). Only 54% of former smokers, however, were correctly identified using this method; of those incorrectly classified, 21% were classified as current smokers and 25% as never smokers. There was also substantial agreement between PDMH smoking status and VA EMR Health Factors data when using the most recent Health Factor entry ($\kappa = 0.61$). While 838 participants did not have any Health Factor entries within the window around the PDMH visit (see Table 2), agreement was substantial for those with non-missing data ($\kappa = 0.64$).

Dichotomous Smoking Status (Current/Not Current; Ever/Never) Comparison

Table 3 shows observed agreement between PDMH smoking status and EMR Health Factors data when smoking categories were collapsed into current/not-current smoking. When examining current smoking, agreement was best for the most common Health Factors entry ($\kappa = 0.73$). The AUC for the most common Health Factors entry was significantly better than the

most recent Health Factors entry ($\chi^2 = 14.85, p < 0.0001$) but did not differ from the Health Factors restricted to be within 12 months of the PDMH study ($\chi^2 = 0.76, n.s.$). Similarly, the highest agreement when examining PDMH ever smoking status (ever vs. never) was observed when using the most common Health Factors entry ($\kappa = 0.75$). The AUC for the most common Health Factors entry was significantly better than the most recent Health Factors entry ($\chi^2 = 19.28, p < 0.0001$) but did not differ from the Health Factors restricted to be within 12 months of the PDMH study ($\chi^2 = 1.48, n.s.$).

Agreement Stratified by Psychiatric Status

Approximately, 49% ($n = 987$) of the sample was diagnosed with a current mental health condition and 36% of these were current smokers. Results examining concordance between EMR Health Factors data and PDMH smoking status indicate few differences between psychiatric and non-psychiatric groups (see Table 4). Consistent with results in the full sample, agreement was highest for the most common Health Factors response for both the three-level smoking status variable (ever, never, former) and dichotomous smoking outcomes (ever/never, current/not current).

Data Missing Not at Random for EMR Health Factors

For the “most common” and “most recent” EMR Health Factors data extraction methods, the rate of missing data was low ($< 1\%$); however, there were substantial missing data (41%) when restricting Health Factors to a 12-month period from the PDMH study visit (only 1187 participants had a Health Factors entry in this chronological window). Post hoc analyses indicated that participants who did not have a Health Factor entry within 12 months of their PDMH research visit were significantly less likely to be a current smoker (13% vs. 40%; $\chi^2 = 169.99, p < 0.0001$) or former smoker (17% vs. 20%; $\chi^2 = 6.15, p = 0.013$), but were more likely to be never smokers (70% vs. 39%; $\chi^2 = 187.98, p < 0.0001$).

Regarding potential underlying reasons for missing Health Factor data, post-hoc analyses indicated that smoking status was significantly related to the number of EMR smoking-related Health Factors entries present in participants’ medical charts. Never smokers had significantly fewer EMR Health Factors entries in their medical charts ($M = 2.2; SD = 3.1$) than both current smokers ($M = 11.3, SD = 9.0; t = -23.38, p < .0001$) and former smokers ($M = 4.8, SD = 5.2; t = -9.45, p < .0001$), suggesting they are less likely to be screened. Current smokers had significantly more EMR Health Factor entries than former smokers ($t = 14.03, p < 0.0001$). The increased number of health factors entries among smokers could not be accounted for by the number of primary care visits in the year surrounding a participant’s PDMH visit. The average number of primary care visits within 12 months of the

Table 1 VA Mid-Atlantic Post-Deployment Mental Health (PDMH) Study Sample Characteristics by Criterion Smoking Status

	Total N = 2025	Never smoker n = 1058	Former smoker n = 397	Current smoker n = 570	Test statistic
Age, mean, (SD)	37.7 (10.3)	38.5 (9.9)	40.3 (11.3)	34.3 (9.3)	$F = 49.32^{***}$
Years of education, mean (SD)	13.4 (3.8)	13.7 (3.9)	13.9 (3.7)	12.8 (3.3)	$F = 14.03^{***}$
Gender, n (%)					$\chi^2 = 24.15^{***}$
Male	1600 (79.0%)	791 (74.8%)	331 (83.4%)	478 (83.9%)	
Female	425 (21.0%)	267 (25.2%)	66 (16.6%)	92 (16.1%)	
Race, n (%)					$\chi^2 = 68.86^{***}$
African-American	994 (49.1%)	611 (57.8%)	157 (39.5%)	226 (39.6%)	
Caucasian	924 (45.6%)	397 (37.5%)	221 (55.7%)	306 (53.7%)	
Other	107 (5.3%)	50 (4.7%)	19 (4.8%)	38 (6.7%)	
Hispanic ethnicity, n (%)	113 (5.7%)	56 (5.3%)	25 (6.3%)	32 (5.6%)	$\chi^2 = 0.64$
Marital status, n (%)					$\chi^2 = 37.44^{****}$
Married	1063 (52.5%)	568 (53.7%)	247 (62.2%)	248 (43.5%)	
Divorced	497 (24.5%)	245 (23.2%)	86 (21.7%)	166 (29.1%)	
Never married	454 (22.4%)	239 (22.6%)	61 (15.4%)	154 (27.0%)	
Combat exposed, n (%)	1510 (74.6%)	777 (73.4%)	291 (73.3%)	442 (77.5%)	$\chi^2 = 3.71$
Axis I disorders, n (%)					$\chi^2 = 65.02^{****}$
No Axis I disorder	1035 (51.2%)	618 (59.7%)	204 (19.7%)	213 (20.6%)	
Any Axis I disorder	987 (48.8%)	439 (44.5%)	193 (19.6%)	355 (36.0%)	
MDD, n (%)	429 (21.2%)	209 (48.7%)	79 (18.4%)	141 (32.9%)	$\chi^2 = 6.11^*$
PTSD, n (%)	650 (32.2%)	287 (44.2%)	121 (18.6%)	242 (37.2%)	$\chi^2 = 48.43^{***}$
Other anxiety dx, n (%)	250 (12.4%)	121 (48.4%)	51 (20.4%)	78 (31.2%)	$\chi^2 = 1.84$
Substance abuse, n (%)	167 (8.3%)	46 (27.5%)	32 (19.2%)	89 (53.3%)	$\chi^2 = 62.16^{***}$

Percentages may not add up to 100% because of missing data. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

MDD major depressive disorder; PTSD posttraumatic stress disorder; Dx diagnosis

PDMH study visit did not differ by smoking status in this relatively young cohort. The average number of primary care visits by smoking status was 5.9 for smokers (median = 4; $IQR = 2-8$), 6.2 for former smokers (median = 5; $IQR = 2-8$), and 6.0 for never smokers (median = 4; $IQR = 2-8$).

While analyses did not indicate differences by smoking status in the number of primary care visits, there were significant differences by smoking status in the number of mental health visits within 12 months of the PDMH study visit. Smokers had an average of 10.8 mental health visits (median = 4; $IQR = 0-12$) within 12 months of their PDMH study visit, whereas never and former smokers averaged 5.4 (median = 1; $IQR = 1-7$) and 5.7 (median = 1, $IQR = 0-6$) mental health visits, respectively. Results of a Kruskal-Wallis test for group differences indicated that smokers had significantly more mental health visits than either never or former smokers ($H = 75.88$, $df = 2$, $p < 0.0001$).

DISCUSSION

The current study replicates McGinnis and colleagues¹⁰ and extends this work to veterans of the Iraq and Afghanistan era and veterans with psychiatric disorders. Consistent with previous research,¹⁰ results suggested substantial agreement between EMR data and smoking status collected during a confidential study visit. EMR Health Factors data (spanning a period of 15 years) were available for almost all of the veterans in the current study. PDMH smoking status showed the highest agreement with smoking status defined by the most common Health Factors entry, and there was no evidence that validity was lower among those with psychiatric disorders.¹⁰

While there was substantial agreement between PDMH and EMR Health Factors data, the current study suggested that the

use of EMR Health Factors data warrants caution. If the EMR is used to select a cohort of patients in a relatively narrow chronological window (e.g., selecting patients with smoking Health Factors data in a given fiscal year), the resulting sample will likely not be representative of the entire population.

Table 2 Smoking Status (Current, Former, Never): Comparison of Electronic Medical Record (EMR) Health Factor to VA Mid-Atlantic Post-Deployment Mental Health Study (PDMH) Data

	PDMH smoking N = 2025	EMR most common smoking health factor N = 2025	EMR most recent smoking health factor N = 2025	EMR health factor within 1 year of PDMH study N = 1187
Smoking status, n (%)				
Current	570 (28%)	615 (30%)	512 (25%)	478 (40%)
Former	397 (20%)	313 (15%)	366 (18%)	242 (20%)
Never	1058 (52%)	1094 (54%)	1137 (56%)	459 (39%)
Unknown*	—	3 (0.2%)	10 (0.5%)	8 (0.7%)
Agreement with PDMH				
Kappa	—	0.69 (0.66–0.71)	0.61 (0.58–0.64)	0.64 (0.61–0.68)
(95% CI)	—	0.74 (0.72–0.77)	0.68 (0.65–0.71)	0.70 (0.66–0.73)
Weighted kappa				
(95% CI)				

*Health Factor data that were coded as unknown were treated as missing data for all analyses. Reduced N in EMR Health Factor within 1 Year of PDMH is due to missing data (i.e., many participants did not have any Health Factors entries in this chronological window). PDMH = Smoking status based upon Post-Deployment Mental Health Study data. Weighted kappa was determined by applying weights that account for the fact that there is greater disagreement when results are two categories apart (e.g., never smoker and current smoker) than one category (e.g., former smoker and current smoker)

Table 3 Current Smoking and Ever Smoking: Comparison of VA Electronic Medical Record (EMR) to VA Mid-Atlantic Post-Deployment Mental Health Study (PDMH) data

Health Factor	n	PDMH	VA EMR	Sensitivity	Specificity	AUC	Kappa
% Current smoker							
Most common	2022	28%	30%	0.84 (0.83–0.86)	0.91 (0.89–0.92)	0.87 (0.86–0.89)	0.73 (0.69–0.76)
Most recent	2015	28%	25%	0.72 (0.70–0.74)	0.93 (0.92–0.94)	0.83 (0.81–0.85)	0.67 (0.63–0.71)
Within 1 year of PDMH	1179	39%	41%	0.83 (0.81–0.85)	0.87 (0.85–0.89)	0.85 (0.83–0.87)	0.69 (0.65–0.73)
% Ever smoker							
Most common	2022	48%	46%	0.85 (0.83–0.87)	0.90 (0.88–0.91)	0.87 (0.86–0.89)	0.75 (0.72–0.78)
Most recent	2019	48%	44%	0.79 (0.77–0.81)	0.89 (0.88–0.90)	0.84 (0.82–0.86)	0.68 (0.65–0.72)
Within 1 year of PDMH	1179	61%	61%	0.89 (0.87–0.90)	0.81 (0.79–0.84)	0.85 (0.83–0.87)	0.70 (0.66–0.74)

Among participants in the current study who had Health Factors data within 12 months of their research visit, the smoking rate was substantially higher than the smoking rate observed in the entire sample (39% vs. 28%). Smokers had approximately five times the Health Factors smoking entries than non-smokers and thus may be more likely to be sampled. The implication of this finding is that missing Health Factor data in any given time frame are likely not missing at random. Since smokers have more smoking Health Factors entries, they have a higher probability of being sampled in a given time frame than non-smokers.

Post-hoc analyses indicated this disparity in Health Factors entries was not due to differences in the number of primary care appointments; however, smokers did have more mental health visits. Since mental health providers are also responsible for completing smoking clinical reminders in VA, current smokers likely had more opportunities to be screened for current smoking. It is also possible that some sites stop screening patients who have consistently reported they are lifetime

non-smokers. Future research could examine the use of additional scoring algorithms to determine smoking status using Health Factors data (e.g., carrying forward entries of lifetime non-smokers).

McGinnis and colleagues¹⁰ identified several potential limitations to using EMR Health Factors data. Although the current smoking coding scheme excluded Health Factors data that specified smokeless tobacco use, it is possible that some patients identified as smokers are smokeless tobacco users. Additionally, while using the most common Health Factor entry yielded the highest concordance with the PDMH criterion, this strategy can lead to misclassification because recent quitters would likely be classified as smokers.¹⁰ This may be acceptable for health services research for two reasons: (1) it takes significant time for the long-term health benefits of smoking cessation to be realized,¹⁰ and (2) many people who have recently quit smoking subsequently relapse.³⁴ While the methods used in this study follow the original validation procedure,¹⁰ it is worth noting that the validation criterion for smoking status was based on self-report rather than biological assay and could be subject to under-reporting. Under-reporting could also occur in the EMR Health Factors data, which are collected during face-to-face interviews with clinicians. Future work could examine whether incorporating other information in the EMR (e.g., problem list) improves identification of smoking status.

Despite these limitations, the current study found substantial agreement between VA EMR Health Factors data and study-reported smoking status. This is the first study that has examined the performance of the Health Factor smoking entries in a cohort of veterans with service during the wars in Iraq and Afghanistan. Strengths of the current study include a large, racially diverse sample with a high prevalence of psychiatric conditions including PTSD, depression, and substance use and a collection of EMR Health Factors data over a long period.

While this study focused on the validity of smoking status captured in screening data in the VA healthcare system (the single largest healthcare system in the US), the findings have implications for other healthcare systems.³⁵ The implementation and capture of population-based screening results are likely to provide a more accurate alternative to use of ICD-9 and procedure codes to assess the burden of smoking. Studies of VA and non-VA

Table 4 Kappa Values for Comparison of the VA Electronic Medical Record (EMR) to the VA Mid-Atlantic Post-Deployment Mental Health Study (PDMH) Data by Psychiatric Status

Psychiatric status	EMR health factor	Kappa for smoking status		
		Current vs. former vs. never*	Current vs. not current	Ever vs. never
No current psychiatric diagnosis	Most common	0.73 (0.69–0.77)	0.72 (0.66–0.77)	0.74 (0.69–0.78)
	Most recent	0.68 (0.64–0.73)	0.68 (0.62–0.73)	0.69 (0.64–0.73)
	Within 1 year of PDMH	0.68 (0.62–0.73)	0.67 (0.61–0.74)	0.68 (0.61–0.74)
Any current psychiatric diagnosis	Most common	0.74 (0.70–0.77)	0.73 (0.68–0.77)	0.75 (0.71–0.79)
	Most recent	0.66 (0.62–0.70)	0.65 (0.60–0.70)	0.67 (0.62–0.71)
	Within 1 year of PDMH	0.70 (0.65–0.75)	0.69 (0.63–0.75)	0.71 (0.65–0.77)

*Weighted kappa shown. Weighted kappa was determined by applying weights that account for the fact that there is greater disagreement when results are two categories apart (e.g., never smoker and current smoker) than one category (e.g., former smoker and current smoker)

EMR data indicate that ICD-9 codes and medical procedure codes considerably underestimate smoking status.^{36,37} Current findings indicate that the use of screening data is a feasible and valid approach to determine smoking status. While the VA mandated universal population-based screening, it allowed local variation in how clinical reminders were implemented (resulting in more than 900 different Health Factor text entries) negatively impacting data quality. The potential utility of screening results for research would be strengthened by the application of a standardized screen and outcome assessment strategy. Other aspects of EMR data quality (e.g., completeness, plausibility, currency) should also be considered before reuse for research.³⁸ Results from this study contribute to the growing literature documenting the potential utility of leveraging EMR systems to algorithmically assess risk for smoking,³⁶ prompt providers to collect smoking status, and provide recommendations for empirically supported treatments for smoking cessation.³⁹

In summary, VA EMR Health Factors smoking data can be used to accurately determine smoking status for Iraq/Afghanistan era veterans. Lack of smoking data has been a limitation in many studies that have used the VA EMR. Caution is warranted, however, when using EMR Health Factors data to select cases for cross-sectional or prospective cohort studies. Results of the current study suggest that selecting cases with available Health Factor smoking data in a relatively narrow chronological window may result in a sample with an inflated smoking rate compared to the population.

Acknowledgements: The VA Mid-Atlantic Mental Illness Research, Education and Clinical Center (MIRECC) Registry Workgroup for this manuscript includes: John A. Fairbank, PhD, Mira Brancu, PhD, Eric B. Elbogen, PhD, Kimberly T. Green, PhD, Jason D. Kilts, PhD, Angela Kirby, MS, Christine E. Marx, MD, MS, Scott D. Moore, MD, PhD, Rajendra Morey, MD, MS, Jennifer C. Naylor, PhD, Jennifer J. Runnals, PhD, Kristy A. Straits-Tröster, PhD, Steven T. Szabo, MD, PhD, Larry A. Tupler, PhD, Elizabeth E. Van Voorhees, PhD, H. Ryan Wagner, PhD, Durham VA Medical Center, Durham, North Carolina; Treven Pickett, PsyD, Hunter Holmes McGuire Department of Veterans Affairs Medical Center, Richmond, Virginia; Robin A. Hurley, MD, Jared Rowland, PhD, Katherine H. Taber, PhD, and Ruth Yoash-Gantz, PsyD, W. G. (Bill) Hefner VA Medical Center, Salisbury, North Carolina; John Mason, PsyD, and Marinell Miller-Mumford, PhD, Hampton VA Medical Center, Hampton, VA; and Gregory McCarthy, PhD, Yale University.

This work was supported by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Health Services Research and Development (I01HX001109); Rehabilitation Research and Development (I01RX001301), and by the National Cancer Institute (R01CA196304). This work was also supported by the VA Mid-Atlantic Mental Illness Research, Education, and Clinical Center; the Department of Veterans Affairs Office of Academic Affiliations Advanced Fellowship Program in Mental Illness Research and Treatment (Dr. Wilson), a VA Research Scientist Award from the Clinical Sciences Research and Development Service (CSR&D) of VA Office of Research and Development (ORD) (Dr. Beckham), a VA Career Development Award from the Rehabilitation Research and Development Service of VA ORD (IK2RX000703) (Dr. McDonald), and a VA Career Development Award from the CSR&D of VA ORD (IK2CX000718) (Dr. Dedert).

Corresponding Author: Patrick S. Calhoun, Ph.D.; VA Mid-Atlantic Region Mental Illness Research, Education and Clinical Center (MIRECC), Durham VA Medical Center, Durham, NC, USA (e-mail: patrick.calhoun2@va.gov).

Compliance with Ethical Standards:

Conflict of Interest: The authors have no conflicts of interest to declare. The Department of Veterans Affairs had no involvement in the study design, collection, analysis and interpretation of data, in the writing of the report, and in the decision to submit the paper for publication. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the VA or the United States government or any of the institutions with which the authors are affiliated. Since the authors are employees of the United States government and contributed to this work as part of their official duties, the work is not subject to US copyright.

REFERENCES

- Centers for Disease Control and Prevention (CDC). Annual smoking-attributable mortality years of potential life lost and economic costs—United States, 1995–1999. *MMWR Morb Mortal Wkly Rep*. 2002;51(14):300–303.
- U.S. Department of Health and Human Services. *The Health Consequences of Smoking: 50 years of Progress. A report of the Surgeon General*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2014. Printed with corrections, January 2014.
- Centers for Disease Control and Prevention. Smoking-attributable mortality, years of potential life lost, and productivity losses—United States, 2000–2004. *MMWR*. 2008;57(45):1226–1228.
- Smith B, Ryan MA, Wingard DL, et al. Cigarette smoking and military deployment: A prospective evaluation. *Am J Prev Med*. 2008;35(6):539–546.
- Bergman HE, Hunt YM, Augustson E. Smokeless tobacco use in the United States military: A systematic review. *Nicotine Tob Res*. 2012;14(5):507–515.
- Feigelman W. Cigarette smoking among former military service personnel: A neglected social issue. *Prev Med*. 1994;23:235–241.
- Hermes ED, Wells TS, Smith B, et al. Smokeless tobacco use related to military deployment, cigarettes and mental health symptoms in a large, prospective cohort study among US service members. *Addiction*. 2012;107(5):983–994.
- McKinney WP, McIntire DD, Carmody TJ, Joseph A. Comparing the smoking behavior of veterans and nonveterans. *Public Health Rep*. 1997;112:212–217.
- Acheson SK, Straits-Troster K, Calhoun PS, Beckham JC, Hamlett-Berry K. Characteristics and correlates of tobacco use among US veterans returning from Iraq and Afghanistan. *Mil Psychology*. 2011;23(2):297–314.
- McGinnis KA, Brandt CA, Skanderson M, et al. Validating smoking data from the Veteran's Affairs Health Factors dataset, an electronic data source. *Nicotine Tob Res*. 2011;13(12):1233–1239.
- Department of Veterans Affairs. VA informatics and computing infrastructure: Corporate data warehouse. 2016; http://www.hsrd.research.va.gov/for_researchers/vinci/cdw.cfm. Accessed May 16, 2016.
- Thompson WH, St-Hilaire S. Prevalence of chronic obstructive pulmonary disease and tobacco use in veterans at Boise Veterans Affairs Medical Center. *Respir Care*. 2010;55(5):555–560.
- Smith MW, Chen S, Siroka AM, Hamlett-Berry K. Using policy to increase prescribing of smoking cessation medications in the VA healthcare system. *Tob Control*. 2010;19:507–511.
- Brown SA, Lincoln MJ, Groen PJ, Kolodner RM. VISTA-US Department of Veterans Affairs national-scale HIS. *Int J Med Inform*. 2003;69:135–156.
- Institute of Medicine. *Preventing Medical Errors*. Washington, DC: National Academies Press; 2005.
- Jha AK, Perlin JB, Kizer KW, Dudley RA. Effect of the transformation of the Veterans Affairs Health Care System on the quality of care. *N Engl J Med*. 2003;348(22):2218–2227.
- McClernon FJ, Calhoun PS, Hertzberg JS, Dedert EA, Beckham JC. Associations between smoking and psychiatric comorbidity in US Iraq and Afghanistan-era veterans. *Psychol Addict Behav*. 2013;27:182–188.
- Ziedonis D, Hitsman B, Beckham JC, et al. Tobacco use and cessation in psychiatric disorders: National Institute of Mental Health report. *Nicotine Tob Res*. 2008;10(12):1691–1715.

19. **Schry AR, Rissling MB, Gentes EL, et al.** The relationship between posttraumatic stress symptoms and physical health in a survey of US veterans of the Iraq and Afghanistan era. *Psychosomatics*. 2015;56(6):674–684.
20. **Calhoun PS, Levin HS, Dedert EA, Johnson YC,** Mid-Atlantic Research Education and Clinical Center Workgroup, Beckham JC. The relationship between posttraumatic stress disorder and smoking outcome expectancies among US military veterans who served since September 11, 2001. *J Trauma Stress*. 2011;24(3):303–308.
21. **Calhoun PS, McDonald SD, Guerra VS, Eggleston AM, Beckham JC, Straits-Troster K.** Clinical utility of the Primary Care-PTSD Screen among US veterans who served since September 11, 2001. *Psychiatry Res*. 2010;178(2):330–335.
22. **Gentes EL, Dennis PA, Kimbrel NA, et al.** DSM-5 posttraumatic stress disorder: Factor structure and rates of diagnosis. *J Psychiatr Res*. 2014;59:60–67.
23. **McDonald SD, Beckham JC, Morey R, Marx C, Tupler LA, Calhoun PS.** Factorial invariance of posttraumatic stress disorder symptoms across three veteran samples. *J Trauma Stress*. 2008;21:309–317.
24. **Brancu M, Wagner HR, Morey RA, et al.** The Post-Deployment Mental Health (PDMH) study and repository: A multisite study of US Afghanistan and Iraq era veterans. *Int J Methods Psychiatr Res*. 2017. doi:10.1002/mpr.1570.
25. Centers for Disease Control and Prevention. Current cigarette smoking among adults in the United States. 2014; http://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking/.
26. Schoenborn CA, Adams PF. Health behaviors of adults: United States, 2005–2007. National Center for Health Statistics. *Vital Health Stat* 2010;10(245).
27. **First MB, Spitzer RL, Gibbon M, Williams JBW.** Structured Clinical Interview for DSM-IV Axis I Disorders. New York, NY: Biometrics Research, New York State Psychiatric Institute; 1996.
28. **Frayne SM, Chiu VY, Iqbal S, et al.** Medical care needs of returning veterans with PTSD: Their other burden. *J Gen Intern Med*. 2011;26(1):33–39.
29. **Mattocks KM, Skanderson M, Goulet JL, et al.** Pregnancy and mental health among women veterans returning from Iraq and Afghanistan. *J Womens Health*. 2010;19(12):2159–2166.
30. **Calhoun PS, Bosworth HB, Grambow SC, Dudley TK, Beckham JC.** Medical service utilization by veterans seeking help for posttraumatic stress disorder. *Am J Psychiatry*. 2002;159:2081–2086.
31. **Landis JR, Koch GG.** The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
32. **Cohen J.** Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. 1968;70(4):213–220.
33. **McFall RM, Treat TA.** Quantifying the information value of clinical assessments with signal detection theory. *Annu Rev Psychol*. 1999;50:215–241.
34. **Yudkin P, Hey K, Roberts S, Welch S, Murphy M, Walton R.** Abstinence from smoking eight years after participation in randomised controlled trial of nicotine patch. *Br Med J*. 2003;327(7405):28–29.
35. **Sherman SE.** A framework for tobacco control: Lessons learnt from Veterans Health Administration. *Br Med J*. 2008;336(7651):1016–1019.
36. **Chen LH, Quinn V, Xu L, et al.** The accuracy and trends of smoking history documentation in electronic medical records in a large managed care organization. *Subst Use Misuse*. 2013;48(9):731–742.
37. **Wiley LK, Shah A, Xu H, Bush WS.** ICD-9 tobacco use codes are effective identifiers of smoking status. *J Amer Med Inform Assoc*. 2013;20(4):652–658.
38. **Weiskopf NG, Weng C.** Methods and dimensions of electronic health record data quality assessment: Enabling reuse for clinical research. *J Amer Med Inform Assoc*. 2013;20(1):144–151.
39. **Bae J, Ford EW, Huerta TR.** The electronic medical record's role in support of smoking cessation activities. *Nicotine Tob Res*. 2016;18(5):1019–1024.