JAMIA Open, 5(2), 2022, 1–7
https://doi.org/10.1093/jamiaopen/ooac040
Research and Applications

Research and Applications

Using the biomarker cotinine and survey self-report to validate smoking data from United States Veterans Health Administration electronic health records

Kathleen A. McGinnis[1], Melissa Skanderson[1], Amy C. Justice [1,2], Hilary A. Tindle[3,4], Kathleen M. Akgün[2], Aleksandra Wrona[1,3,4], Matthew S. Freiberg[3,4], Matthew Bidwell Goetz[5,6], Maria C. Rodriguez-Barradas[7], Sheldon T. Brown[8], and Kristina A. Crothers[9]

[1]VA Connecticut Healthcare System, West Haven, Connecticut, USA, [2]Department of Medicine, Yale University, New Haven, Connecticut, USA, [3]Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee, USA, [4]Veterans Health Administration-Tennessee Valley Healthcare System Geriatric Research, Education and Clinical Center (GRECC), Nashville, Tennessee, USA, [5]VA Greater Los Angeles Healthcare System, Los Angeles, California, USA, [6]David Geffen School of Medicine at UCLA, Los Angeles, California, USA, [7]Infectious Diseases Section, Department of Medicine, Michael E. DeBakey VAMC, Baylor College of Medicine, Houston, Texas, USA, [8]James J. Peterson VAMC, Bronx, New York, USA, and [9]Department of Medicine, Harborview Medical Center, University of Washington, Seattle, Washington, USA

Corresponding Author: Kathleen A. McGinnis, DrPH, VA Connecticut Healthcare System, 950 Campbell Avenue, West Haven, CT 16506, USA; kathleen.mcginnis3@va.gov

Received 24 February 2022; Revised 4 May 2022; Editorial Decision 8 May 2022; Accepted 19 May 2022

ABSTRACT

Objective: Tobacco use/smoking for epidemiologic studies is often derived from electronic health record (EHR) data, which may be inaccurate. We previously compared smoking from the United States Veterans Health Administration (VHA) EHR clinical reminder data with survey data and found excellent agreement. However, the smoking clinical reminder items changed October 1, 2018. We sought to use the biomarker salivary cotinine (cotinine $\geq$30) to validate current smoking from multiple sources.

Materials and Methods: We included 323 Veterans Aging Cohort Study participants with cotinine, clinical reminder, and self-administered survey smoking data from October 1, 2018 to September 30, 2019. We included International Classification of Disease (ICD)-10 codes F17.21 and Z72.0. Operating characteristics and kappa statistics were calculated.

Results: Participants were mostly male (96%), African American (75%) and mean age was 63 years. Of those identified as currently smoking based on cotinine, 86%, 85%, and 51% were identified as currently smoking based on clinical reminder, survey, and ICD-10 codes, respectively. Of those identified as not currently smoking based on cotinine, 95%, 97%, and 97% were identified as not currently smoking based on clinical reminder, survey, and ICD-10 codes. Agreement with cotinine was substantial for clinical reminder (kappa = .81) and survey (kappa = .83), but only moderate for ICD-10 (kappa = .50).

Discussion: To determine current smoking, clinical reminder, and survey agreed well with cotinine, whereas ICD-10 codes did not. Clinical reminders could be used in other health systems to capture more accurate smoking information.

Conclusions: Clinical reminders are an excellent source for self-reported smoking status and are readily available in the VHA EHR.

Key words: smoking, cotinine, self-reported, ICD-10, Veterans Health Administration, electronic health record (EHR)

LAY SUMMARY

The Veterans Health Administration (VHA) changed the questions used to screen for tobacco use/smoking at routine clinic visits in October 2018. The tobacco use/smoking data from clinic visits prior to October 2018 was shown to agree well with confidentially self-reported cigarette smoking survey data, and the new questions beginning in October 2018 should result in even better agreement. Nonetheless, the new clinic questions needed to be evaluated and we were able to validate them using the biomarker cotinine, which detects nicotine in saliva for up to 4 days, as the "gold standard." We also compared cigarette smoking survey data and electronic health record diagnostic codes to cotinine. To identify current smoking, clinic questions and survey agreed well with cotinine (of those identified with tobacco use based on cotinine, 86% and 85% reported current tobacco use at clinic visits, and on the survey, respectively). Of those identified with tobacco use based on cotinine, only 51% had a diagnostic code for tobacco use. Tobacco use collected at VHA clinic visits is accurate, and this is important for identifying those who would benefit from tobacco use interventions and lung cancer screening.

INTRODUCTION

Smoking is a leading cause of mortality[1–4] and is a major risk factor for comorbidities, such as bacterial pneumonia, chronic obstructive pulmonary disease (COPD), cardiovascular disease, and cancer.[5–10] Adjusting for smoking status in research studies is paramount for interpreting results, particularly when comparing outcomes between populations with varied smoking rates and smoking status over time. Furthermore, to evaluate efficacy of smoking cessation interventions, accurately determining smoking status over time is necessary. Although cotinine or confidential self-administered self-report may be more accurate,[11] electronic health record (EHR)-derived measures may be the only practical means of assessing smoking in many situations.

The Veterans Health Administration (VHA) benefits from one of the most highly developed health information systems in the world with over 20 years of longitudinal clinical data on over 22 million veterans. Although the EHR includes International Classification of Diseases (ICD), 9th Revision (ICD-9), and 10th Revision (ICD-10) nicotine and tobacco use diagnosis codes, these have been previously shown to be susceptible to underreporting.[9] Since 1999, smoking status has been documented in the EHR via clinical reminders which prompt clinicians to ask questions regarding smoking behavior and record patient responses in structured data fields within the VHA EHR Health Factors Tables. We previously compared self-reported tobacco use/smoking status from the clinical reminder data with confidentially collected cigarette smoking survey data and found excellent agreement.[12,13]

However, the VHA changed the questions used to screen for tobacco use/smoking at routine clinic visits via clinical reminders in October 2018. The new questions beginning in October 2018 should result in even better agreement, but have not yet been validated. As of October 1, 2018, the clinical reminder tobacco use items have been standardized and streamlined to more closely align with the Healthcare Effectiveness Data and Information Set, a set of performance measures created by the National Committee for Quality Assurance. VHA patients in primary care and mental health outpatient clinics are to be screened annually with the tobacco use clinical reminder. Although assessing tobacco use more consistently across sites should lead to improved data collection, this change needs to be evaluated. Therefore, validating the new format of the clinical reminder against cotinine, a biomarker reflecting nicotine exposure, is particularly timely.

The aims of this study were to (1) assess and compare the agreement of current smoking based on the newly revised clinical reminder, confidentially reported survey data, and ICD-10 codes with the salivary cotinine biomarker as the "reference standard"; (2) determine whether agreement of clinical reminder with survey smoking (current/former/never) is similar before and after the October 2018 clinical reminder changes; and (3) quantify agreement of ICD-9 and ICD-10 codes with survey smoking.

MATERIALS AND METHODS

Data

The Veterans Aging Cohort Study (VACS) survey sample is a cohort of people with HIV (PWH) and uninfected veterans consented for prospective, recurrent confidential self-administered surveys on cigarette smoking status, alcohol, and drug use, and other health-related behaviors. From October 2012 to February 2018, 2195 PWH and 1976 uninfected people were enrolled in the VACS survey sample.[14–16] These data are also linked to the longitudinal VHA EHR using the Corporate Data Warehouse.

VACS survey sample participants with HIV and on antiretroviral treatment were eligible for recruitment into the Medications, Alcohol, and Substance use in HIV Study (MASH), a VACS substudy. MASH enrolled 467 VACS survey sample participants from October 2018 to May 2021. Participants provided salivary cotinine samples in addition to nail clippings and blood spots for substance use assessment.

Of the 467 enrolled in the VACS MASH substudy through May 2021, 354 PWH were enrolled from October 2018 to September 2019 (FY 2019), and 323 who had survey, cotinine, and clinical reminder tobacco use data were included in the analytic sample. Nicotine replacement therapy (NRT) was identified using the VHA Pharmacy Benefits Management data set.

To compare current/former/never smoking based on clinical reminder and ICD-9 and ICD-10 codes to confidential survey data before and after the October 2018 national clinical reminder changes, we also identified an analytic sample ($n = 227$) including participants with and without HIV who completed a VACS follow-up survey and had clinical reminder tobacco use data from October 2016 to September 2017 (FY 2017). We did not include clinical reminder tobacco use/smoking data from October 2017 to September 2018

because the changes in the VHA Outpatient Tobacco Use Performance Measure (clinical reminder) were partially rolled out for piloting at some sites during this time.

The study was approved by the Internal Review Boards (IRB) for the Department of Veterans Affairs (VA) and Yale University. The IRB numbers for VACS are VA IRB—AJ0001, VA IRB net—1583210, Yale IRB—0309025943; and for MASH are VA IRB—AJ0015, VA IRB net—1583214, Yale IRB—2000022127. Due to VA regulations and our ethics agreements, the analytic data sets used for this study are not permitted to leave the VA firewall without a Data Use Agreement. This limitation is consistent with other studies based on VA data.

Measures used to approximate cigarette smoking

Salivary cotinine biomarker

Cotinine is the predominant metabolite of nicotine and is used as a biomarker for nicotine exposure. Saliva samples were self-collected using oral swabs in research clinics on the same day surveys were completed. Samples were shipped to Salimetrics laboratory where they were analyzed for quantitative cotinine ng/ml using ELISA methodology. Because prior research has indicated that values between 10 and 30 could indicate a non-heavy level of smoking or passive smoke exposure,[17,18] for the validation we assessed cutoffs of both $\geq$10 and $\geq$30. We used a dichotomous variable for current versus not current nicotine exposure (smoking) because salivary cotinine has a short half-life in the body (average half-life of 20 hours depending on hepatic rate of nicotine metabolism) so cannot be used to detect nicotine exposure beyond ~72 hours (Supplementary Appendix Table SA).

Confidential self-administered survey smoking

Smoking from the VACS surveys was coded as current if a person reported that they have smoked at least 100 cigarettes in their entire life and are "still smoking" or if it has been less than a month since last smoked. Former smoking was defined as having smoked at least 100 cigarettes in one's lifetime and having smoked more than 1 month ago. Otherwise, participants were considered to have never smoked.

Clinical reminder tobacco use

The clinical reminder for tobacco use is required to be completed annually for VHA patients in primary care and mental health outpatient clinics. The clinical reminder tobacco use/smoking data in the Health Factor Tables consists of standardized text entries representing the patient responses to questions asked by healthcare personnel that vary by site and over time through September 30, 2018; the text of the questions themselves were not included in Health Factors data. There are over a thousand different text entries up to September 30, 2018.

Starting October 1, 2018, tobacco use (defined as cigarettes, cigars, pipes, or smokeless tobacco) questions have been standardized and limited in number; response options are asked in a consistent manner for all VHA sites, but are still recorded in standardized text format. There are only 22 response options for October 1, 2018 and onward (see Supplementary Appendix Table SB). Patients who report using tobacco on at least some days are advised to quit as part of the clinical reminder. The tobacco use question does not include use of electronic delivery devices, such as e-cigarettes.

Using these new clinical reminders, we updated our prior algorithms that classify pre-2018 smoking clinical reminders as never, former, and current smoking (mapping strategies, including specific SAS and SQL code, and the look up table are available at www.vacohort.org). If there was more than 1 clinical reminder entry during the FY, we used the most common entry to be consistent with our prior work.[12] We also evaluated the use of clinical reminder tobacco use closest to the survey date.

ICD codes for tobacco use/nicotine dependence

From the EHR we identified ICD-9 and ICD-10 codes indicative of smoking. ICD-10 codes began October 2015. We included ICD-9 codes 305.1 for "Current tobacco use" and V15.82 for "History of tobacco use" (Wiley et al 2013). For ICD-10 we included F17.21 for "Nicotine dependence with cigarettes," Z72.0 for "Tobacco Use," and Z87.891 for "Personal history of nicotine dependence." We defined current smoking based on the presence of nonhistory tobacco use/nicotine codes within a year prior to the survey date. We considered former smoking to have occurred if a participant had any of the above tobacco use/nicotine codes dating back to 1999, but none of the other nonhistory tobacco use/nicotine codes within the past year. Otherwise, participants were considered to have never smoked.

Analyses

Demographic characteristics and smoking measures for both the FY 2019 and FY 2017 participants were summarized. To assess and compare current smoking based on the FY 2019 clinical reminder, survey, and ICD-10 codes to cotinine $\geq$30 ng/ml and $\geq$10 ng/ml data as the "reference standard," we calculated sensitivity, specificity, positive predictive value, negative predictive value, agreement, and kappa statistics. Of those with a positive cotinine value who reported former or never smoking on the survey, we calculated the percent who were on NRT and/or reported noncigarette tobacco use on the survey. We also examined those with a negative cotinine value who reported current smoking on the survey. We collapsed categories so that no cells based on <11 individuals are reported.

To determine whether agreement of current/former/never clinical reminder and ICD-9 and ICD-10 codes with confidential survey smoking data (as "reference standard") is similar before and after the October 2018 clinical reminder changes, we calculated the percent correctly identified into the 3 groups (agreement percent) and kappa statistics for FY 2019 and FY 2017. Kappa statistics were generated in 2 ways. First, agreement was considered to be either *perfect* (1) or *not perfect* (0). The second way includes weighting that acknowledges a difference between being 2 categories apart (never smoking and currently smoking) versus being only 1 category apart (never smoking and former smoking or current smoking and former smoking). If agreement is different by only 1 category, weighting would be 0.5 rather than 0. We also compared the kappa statistics by HIV status for FY 2017.

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.[19] McHugh[20] suggests interpreting values as follows: 0–0.20—none, 0.21–0.39—minimal; 0.40–0.59—weak; 0.60–0.79—moderate; 0.80–0.90—strong; 0.90–1.00—almost perfect.

RESULTS

FY 2019 analytic sample

We identified 323 participants (all PWH) who enrolled in the VACS sub-study in FY 2019 with survey and cotinine data and also had at

Table 1. Characteristics of VACS FY 2019 and FY 2017 survey data

	FY 2019 ($n = 323$)	FY 2017 ($n = 227$)
Mean age (SD)	63 (7.6)	59 (8.9)
Female (%)	4	5
Race/ethnicity (%)		
African American	75	60
White	18	29
Hispanic/other	9	11
Electronic nicotine delivery system use in past year (%)	6	NA
Prescription for COPD or asthma medication (%)	15	NA
HIV Positive (%)	100	57
Cigarette smoking measures		
Cotinine ng/ml (%)[a]		
<30	57	NA
≥30	43	
Survey (%)		
Never	24	24
Former	38	24
Current	38	52
Clinical reminder (%)		
Never	26	27
Former	34	14
Current	39	59
ICD codes (%)		
Never	38	31
Former	38	45
Current	24	24

COPD: chronic obstructive pulmonary disease; ICD: International Classification of Disease; VACS: Veterans Aging Cohort Study.

[a] $n = 318$.

least 1 smoking clinical reminder completed in FY 2019. Participants were mostly male (96%), African American (75%), mean age was 63 years, 15% reported prescription use of COPD or asthma medications, and 6% reported electronic nicotine delivery service use in the past year. Current smoking was common (43% based on cotinine ≥30, 38% based on survey, 39% based on most common clinical reminder), but less common if based on ICD codes (24%; Table 1).

Current smoking using cotinine as the "reference standard"

Clinical reminders and survey data performed similarly when compared with the reference standard of cotinine to identify current smokers. The median difference between dates of cotinine and closest clinical reminder was 105 days (interquartile range = 22–162). Cotinine and survey data were captured on the same day. Of those identified as currently smoking based on cotinine, 86%, 85%, and 51% were identified as currently smoking based on clinical reminder, survey, and ICD-10 codes, respectively (sensitivity). Of those identified as not currently smoking based on cotinine, 95%, 97%, and 97% were identified as not currently smoking based on clinical reminder, survey, and ICD-10 codes (specificity). Agreement with cotinine was substantial for clinical reminder (kappa = .81) and survey (kappa = .83), but only moderate for ICD-10 codes (kappa = .50; Table 2).

Sensitivity and kappa statistics were slightly lower and specificity is slightly higher when the cutoff for cotinine was ≥10 rather than ≥30 (Table 2). Results were similar when we used the closest clinical reminder to the cotinine test for smoking status in place of the most common clinical reminder during the FY. There were 27 participants with cotinine ≥10 who reported former or never smoking on the survey. Of these 27, none were on NRT within 60 days prior, 16 smoked in the past based on the clinical reminder, and 13 reported current smoking based on the clinical reminder or e-cigarette use in the past month. Eleven or less participants had cotinine values of <10 (ranging from 2.5 to 8.5) but reported current smoking on the survey.

Current/former/never clinical reminder compared with survey smoking as the "reference standard"

Clinical reminders performed well to identify smoking status as current, former, or never. Of those who reported current smoking on the survey, 91% were identified as currently smoking based on the clinical reminder; of those who reported on the survey never smoking, 79% were identified as never smoking based on the most common clinical reminder; and of those who reported former smoking on the survey, 75% reported former smoking based on the clinical reminder. Agreement was substantial between clinical reminder and survey data based on kappa statistics without and with weighting for FY 2019 (.73 and .76; Table 3).

Current/former/never ICD-9 and ICD-10 compared with survey smoking as the "reference standard"

ICD codes did not perform as well as to identify current, former, or never smoking. Of those who reported current smoking on the survey, 54%, respectively were identified as currently smoking-based ICD codes; of those who reported on the survey never smoking, 79% were identified as never smoking based on ICD codes; and of those who reported former smoking on the survey, 52% were identified with former smoking based on ICD codes. Agreement was weak between survey and ICD-code data based on kappa statistics without and with weighting (.39 and .49).

FY 2017 analytic sample

To determine whether agreement of clinical reminder and ICD codes with confidential survey smoking data is similar before and after the October 2018 change in clinical reminder items, we identified 227 participants (129 PWH and 98 uninfected) in VACS who completed a follow-up survey in FY 2017 and also had at least 1 smoking clinical reminder completed in FY 2017. There were 33 participants who overlapped with the FY 2019 group. Participants were mostly male (95%), African American (60%) and mean age was 59 years. Current smoking was common (52% based on survey; 59% based on clinical reminder), but less common if based on ICD codes (24%; Table 1).

Current/former/never clinical reminder compared with survey smoking as the "reference standard"

Clinical reminders performed well to identify smoking status as current, former, or never, although agreement was lower using FY 2017 compared with FY 2019 data. Of those who reported current smoking on the survey, 92% were identified as currently smoking based on the clinical reminder; of those who reported on the survey never smoking, 71% were identified as never smoking based on the most common clinical reminder; and of those who reported former smoking on the survey, 42% reported former smoking based on the clinical reminder. Agreement was moderate between clinical reminder and survey data based on kappa statistics without and with weighting (.58 and .63; Table 3).

Table 2. Smoking from clinical reminder, survey, and ICD codes compared with cotinine (≥30 and ≥10) as reference standard ($n = 318$), FY 2019

	Percent						
	Prevalence	Agreement	Sensitivity	Specificity	PPV	NPV	Kappa
Cotinine ≥30	42		—	—	—	—	—
Clinical reminder	39	91	86	95	92	90	.81
Survey	38	92	85	97	95	90	.83
ICD-10	24	77	51	97	92	73	.50
Cotinine ≥10	46		—	—	—	—	—
Clinical reminder	39	90	82	97	95	86	.79
Survey	38	91	82	99	98	86	.81
ICD-10	24	74	48	97	93	69	.46

ICD: International Classification of Disease.

Table 3. Smoking based on clinical reminder and ICD-9 and ICD-10 data compared with survey as "reference standard"

	Survey smoking					
	FY 2019 ($n = 323$)			FY 2017 ($n = 227$)		
Smoking	Never	Former	Current	Never	Former	Current
Clinical reminder						
Never (%)	79			71		
Former (%)		75			42	
Current (%)			91			92
Kappa statistic		.73			.58	
Agreement		82%			75%	
Weighted kappa statistic		.76			.63	
Weighted agreement		90%			84%	
ICD-9 and ICD-10						
Never (%)	79			82		
Former (%)		52			58	
Current (%)			54			42
Kappa statistic		.39			.36	
Agreement		59%			56%	
Weighted kappa statistic		.49			.47	
Weighted agreement		78%			76%	

Note: Only agreement cell percents are shown because some of the other cell sizes are too small to display because they contain ≤10 individuals. Column percents are shown.

ICD: International Classification of Disease.

Kappa statistics without and with weighting were similar for PWH (.57 and .61) and without HIV (.56 and .61), respectively, although current smoking based on all measures was less common among PWH than those without HIV (Supplementary Appendix Table SC). Results were similar when we used the closest clinical reminder smoking status in place of the most common clinical reminder during the FY.

Current/former/never ICD-9 and ICD-10 compared with survey smoking as the "reference standard"

ICD codes did not perform as well to identify current, former, or never smoking. Of those who reported current smoking on the survey, 42% were identified as currently smoking-based ICD codes; of those who reported on the survey never smoking, 82% were identified as never smoking based on ICD codes; and of those who reported former smoking on the survey, 58% were identified with former smoking based on ICD codes. Agreement was weak between survey and ICD-code data based on kappa statistics without and with weighting (.36 and .47; Table 3). Kappa statistics without and with weighting were similar but slightly better for PWH (.38 and .47) and without HIV (.31 and .45), respectively (Supplementary Appendix Table SC).

DISCUSSION

In this study, we utilized 2 sources of EHR data to define smoking, namely clinical reminders, and ICD codes, and compared these sources to the reference standards of cotinine to define current smoking, and confidential survey self-report to define current, former, or never smoking. We found that current smoking from the new clinical reminder implemented starting October 2018 agrees well with the smoking biomarker cotinine. Current/former/never smoking from the new clinical reminder also agrees well with self-administered confidential survey smoking. Moreover, the agreement between clinical reminder and confidential survey self-report is even stronger after the VHA implemented changes to streamline and standardize smoking questions across sites in October 2018 based on comparison of FY 2017 and FY 2019 results. Identifying former smoking is particularly better in FY 2019 than FY 2017. Of those with former smoking based on the survey, 75% and 42% were identified with former smoking based on the clinical reminder in FY 2019 and FY 2017, respectively.

Notably, we found that using ICD-9 and ICD-10 codes for smoking resulted in much lower current smoking prevalence, sensitivity, and agreement than the clinical reminder smoking data when considering both the biomarker cotinine and confidential self-report as "reference standards". Despite being commonly used in research,[21] these results convincingly show that when using ICD codes, half of those with current nicotine use were misclassified with no use. Therefore, studies that adjust for nicotine use based on ICD codes will not fully address any potential confounding. However, we found higher sensitivity and accuracy compared with some prior non-VHA studies.[21,22]

We used the biomarker cotinine as the reference standard to define current smoking. We found that self-reported survey had better agreement than the clinical reminder smoking with cotinine, although the difference was minimal. This could be because self-reported survey smoking was collected on the same day as the saliva for cotinine.

Although the intention of this study was to evaluate measures of cigarette smoking, only self-reported smoking on the survey is specific to cigarette smoking. In contrast, cotinine detects nicotine from

any source, including e-cigarettes, vaping, and NRT such as the patch. Based on VACS surveys from 2016 to 2020, only 5% of participants reported using electronic cigarettes in the past year and less than 2% reported exclusive use of electronic cigarettes (unpublished data). There were 27 participants with cotinine $\geq$10 who reported former or never smoking on the survey. Of those, none were on NRT within 60 days prior, and only 11 or less reported e-cigarette use in the past month, but 11 or fewer smoked currently and 16 smoked in the past based on the clinical reminder; this discrepancy might be mostly due to survey underreporting. Conversely, cotinine may miss capturing those who are currently smoking, but not smoking daily, especially in those with faster nicotine metabolism. In these data, of the 120 who reported current smoking on the survey, only 11 or fewer had a cotinine of <10 and these were detectable values (ranging from 2.5 and 8.5 ng/ml). In addition, while cotinine can be used as a reference standard measure for current smoking, it cannot provide information on former smoking as a single measure. The clinical reminder data prior to October 2018 do not specify the question asked of the patients; just the response. The clinical reminder data after October 2018 reflect "tobacco use" including cigarettes, cigars, pipes, and smokeless tobacco, but does not include electronic delivery methods. ICD-9 codes reflect "tobacco use" and ICD-10 codes reflect "tobacco use" and "nicotine dependence." Despite these differences, we still found good agreement between the different sources with the exception of ICD codes.

This study has additional strengths and weaknesses. Although we have cotinine and confidential survey as self-reported reference standards to carefully define smoking status, our comparison incorporates data from 9 VHA sites and very few are women. In the FY 2019 data, all participants are people living with HIV. However, when comparing the performance of clinical reminders and survey by HIV status, we found that the performance of the clinical reminders and survey was similar in those with or without HIV. A limitation to EHR-derived data is that smoking health factors and ICD codes data do not contain information on pack-years. Understanding the impact of the COVID-19 pandemic on screening for tobacco use will also be important as a prior study found that tobacco use screening decreased from 74% in the year before to 60% during the first year of the pandemic for patients with HIV in care in the VHA.[23]

We limited our analyses to 1 year of clinical reminder smoking data, which may not be the approach that every study should take. Although smoking clinical reminders are intended to be asked once a year, not every patient is seen in clinic yearly so a limitation of 1 year may result in missing data. For this analysis, the 1-year limitation was made because the clinical reminder changed as of October 2018, and data from October 2018 to September 2019 represents the most recent FY available prior to the start of the COVID-19 pandemic. For a valid comparison with a prior year, we also needed to limit the prior time to just 1 year. We previously found that using the most common value for smoking (compared with closest) over time agrees best with survey smoking data.[12] In this analysis, we found similar results using most common and closest smoking values. If researchers elect to use a longer amount of time to identify clinical reminders, differences between most common and closest clinical reminder compared with survey could be more pronounced.

In summary, this study strongly supports the validity of the clinical reminder data to determine smoking status when compared with confidential survey self-report both as previously collected and importantly with the changes that were implemented in VHA in FY 2019. Clinical reminders are an excellent source for self-reported current, former, or never smoking status, are readily available in the VHA EHR, and perform substantially better than ICD codes. To determine current smoking, both clinical reminder and confidential survey data agreed well with cotinine, whereas ICD-10 codes did not. Future work should explore whether there are differences in agreement by demographic characteristics, comorbidities, and or medication use. Future work should also assess whether changes in smoking can be accurately determined using clinical reminder data. Capturing the use of e-cigarettes and vaping using clinical reminders could also be useful for understanding the long-term impacts of these exposures.

FUNDING

This work was supported by National Institute on Alcohol Abuse and Alcoholism grants: U01-AA026224, U01-020790, and U24-AA020794.

AUTHOR CONTRIBUTIONS

KAM, MS, and ACJ substantially contributed to the conception and design. MS, AW, MSF, MBG, MCR, and STB were involved critically with the data acquisition. KAM and MS ran the analyses. KAM, MS, ACJ, HAT, KMA, and KAC contributed substantially to the interpretation of results. All authors contributed important intellectual content to drafts and revisions, approved the final version, and agree to be accountable for this work.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *JAMIA Open* online.

ACKNOWLEDGEMENTS

The content is solely the responsibility of the authors and does not necessarily represent the official views of the Veterans Affairs or the National Institutes of Health.

CONFLICT OF INTEREST STATEMENT

Dr. HAT has served as Principal Investigator for smoking cessation studies using FDA-approved medications that were donated by the manufacturer. The other coauthors have no competing interests to declare.

DATA AVAILABILITY

Due to US Department of Veterans Affairs (VA) regulations and our ethics agreements, the analytic data sets used for this study are not permitted to leave the VA firewall without a Data Use Agreement. This limitation is consistent with other studies based on VA data. However, VA data are made freely available to researchers with an approved VA study protocol. For more information, please visit https://www.virec.research.va.gov.

REFERENCES

1. U.S. Department of Health and Human Services. *Smoking Cessation. A Report of the Surgeon General*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Na-

tional Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2020.
2. Crothers K, Goulet JL, Rodriguez-Barradas MC, *et al*. Impact of cigarette smoking on mortality in HIV-positive and HIV-negative veterans. *AIDS Educ Prev* 2009; 21 (3 Suppl): 40–53.
3. Ezzati M, Lopez AD. Estimates of global mortality attributable to smoking in 2000. *Lancet* 2003; 362 (9387): 847–52.
4. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. *JAMA* 2004; 291 (10): 1238–45.
5. Crothers K, Butt AA, Gibert CL, *et al*.; Veterans Aging Cohort 5 Project Team. Increased COPD among HIV-positive compared to HIV-negative veterans. *Chest* 2006; 130 (5): 1326–33.
6. Diaz PT, King MA, Pacht ER, *et al*. Increased susceptibility to pulmonary emphysema among HIV-seropositive smokers. *Ann Intern Med* 2000; 132 (5): 369–72.
7. Kirk GD, Merlo C, O'Driscoll P, *et al*. HIV infection is associated with an increased risk for lung cancer, independent of smoking. *Clin Infect Dis* 2007; 45 (1): 103–10.
8. Sudano I, Spieker LE, Noll G, Corti R, Weber R, Lüscher TF. Cardiovascular disease in HIV infection. *Am Heart J* 2006; 151 (6): 1147–55.
9. 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): 555–60.
10. 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.
11. Jarvis MJ, Fidler J, Mindell J, Feyerabend C, West R. Assessing smoking status in children, adolescents and adults: cotinine cut-points revisited. *Addiction* 2008; 103 (9): 1553–61.
12. 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): 1233–9.
13. Calhoun PS, Wilson SM, Hertzberg JS, *et al*.; VA Mid-Atlantic MIRECC Workgroup. Validation of Veterans affairs electronic medical record smoking data among Iraq- and Afghanistan-era veterans. *J Gen Intern Med* 2017; 32 (11): 1228–34.
14. Gordon KS, McGinnis K, Dao C, *et al*. Differentiating types of self-reported alcohol abstinence. *AIDS Behav* 2020; 24 (2): 655–65.
15. Fultz SL, Skanderson M, Mole LA, *et al*. Development and verification of a "virtual" cohort using the National VA Health Information System. *Med Care* 2006; 44 (8 Suppl 2): S25–30.
16. Justice AC, Dombrowski E, Conigliaro J, *et al*. Veterans Aging Cohort Study (VACS): overview and description. *Med Care* 2006; 44 (8 Suppl 2): S13–24.
17. Benowitz NL, Bernert JT, Foulds J, *et al*. Biochemical verification of tobacco use and abstinence: 2019 update. *Nicotine Tob Res* 2020; 22 (7): 1086–97.
18. Kim S. Overview of cotinine cutoff values for smoking status classification. *Int J Environ Res Public Health* 2016; 13 (12): 1236.
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33 (1): 159–74.
20. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012; 22 (3): 276–82.
21. Wiley LK, Shah A, Xu H, Bush WS. ICD-9 tobacco use codes are effective identifiers of smoking status. *J Am Med Inform Assoc* 2013; 20 (4): 652–8.
22. Wang L, Ruan X, Yang P, Liu H. Comparison of three information sources for smoking information in electronic health records. *Cancer Inform* 2016; 15: 237–42.
23. McGinnis KA, Skanderson M, Justice AC, *et al*. HIV care using differentiated service delivery during the COVID-19 pandemic: a nationwide cohort study in the US Department of Veterans Affairs. *J Int AIDS Soc* 2021; 24(Suppl 6): e25810.