Accuracy of Electronic Health Record Data for the Diagnosis of Chronic Obstructive Pulmonary Disease in Persons Living with HIV and Uninfected Persons

Kristina Crothers[1], **Carla Rodriguez**[1,2], **Robin M. Nance**[1], **Kathleen Akgun**[3], **Shahida Shahrir**[1], **Joon Kim**[4], **Guy Soo Hoo**[5], **Amir Sharafkhaneh**[6], **Heidi M. Crane**[1], and **Amy C. Justice**[3]

[1]Department of Medicine, University of Washington, Seattle, WA

[2]Center for Health Research, Kaiser Permanente Mid-Atlantic Research Institute, Rockville, MD

[3]Department of Medicine, Veterans Affairs (VA) Connecticut Healthcare System and Yale University, West Haven, CT

[4]Department of Medicine, James J. Peters VA Medical Center and Icahn School of Medicine at Mt. Sinai, New York, NY

[5]Department of Medicine, VA Greater Los Angeles Healthcare System and David Geffen School of Medicine at UCLA, Los Angeles, CA

[6]Department of Medicine, Michael E. DeBakey VA Medical Center and Baylor College of Medicine, Houston, TX

Abstract

Purpose: No prior studies have addressed the performance of electronic health record (EHR) data to diagnose chronic obstructive pulmonary disease (COPD) in persons living with HIV (PLWH), in whom COPD could be more likely to be underdiagnosed or misdiagnosed, given the higher frequency of respiratory symptoms and smoking compared to HIV-uninfected (uninfected) persons.

Methods: We determined whether EHR data could improve accuracy of ICD-9 codes to define COPD when compared to spirometry in PLWH vs. uninfected, and quantified level of discrimination using the area under the receiver-operating curve (AUC). The development cohort consisted of 350 participants who completed research spirometry in the Examinations of HIV Associated Lung Emphysema (EXHALE) study, a pulmonary substudy of the Veterans Aging Cohort Study. Results were externally validated in 294 PLWH who performed spirometry for clinical indications from the University of Washington (UW) site of the Centers for AIDS Research Network of Integrated Clinical Systems cohort.

Corresponding Author: Kristina Crothers, MD, 325 9th Avenue, Harborview Medical Center, Seattle, WA 98104, crothk@uw.edu.

Results: ICD-9 codes performed similarly by HIV status, but alone were poor at discriminating cases from non-cases of COPD when compared to spirometry (AUC 0.633 in EXHALE; 0.651 in the UW cohort). However, algorithms that combined ICD-9 codes with other clinical variables available in the EHR – age, smoking and COPD inhalers – improved discrimination and performed similarly in EXHALE (AUC 0.771) and UW (AUC 0.734).

Conclusions: These data support that EHR data in combination with ICD-9 codes have moderately good accuracy to identify COPD when spirometry data are not available, and perform similarly in PLWH and uninfected individuals.

Keywords

Pulmonary Disease, Chronic Obstructive; HIV; Smoking; Electronic Health Records; Area Under Curve

Introduction

Chronic obstructive pulmonary disease (COPD) is a leading cause of death worldwide[1] and is associated with a substantial economic burden. Yet, COPD is often both underdiagnosed as well as misdiagnosed in the absence of spirometry, the gold standard to document the presence of chronic airflow limitation that is the hallmark of COPD.[1–4] Spirometry is often under-utilized, including in primary care settings.[4–8] Under-diagnosis of COPD leaves many without needed interventions, including more aggressive efforts at smoking cessation and appropriate medications.[5,9,10] Misdiagnosis can result in inappropriate use of medications, with concomitant exposure to unnecessary harms from medication side effects, excess costs, and lack of intervention for the actual cause of disease.

In larger scale epidemiologic studies, determination of spirometry results to document the presence of airflow limitation consistent with COPD is not readily feasible given the expense and difficulty in obtaining these data from the electronic health record (EHR), in addition to the infrequent clinical use of spirometry. Rather, diagnoses often rest on review of billing and claims data, typically derived from the International Classification of Diseases, 9th edition (ICD-9) codes and more recently 10th edition (ICD-10). However, while prior studies have found variable accuracy of ICD-9 codes for COPD, few have incorporated other EHR data into algorithms for diagnosis of COPD.[11–14]

Our objective was to develop and validate a model using data available in the EHR, including ICD-9 codes and clinically derived data, to accurately define COPD when compared to spirometry. We sought to compare results by HIV status, as prior studies have not addressed the performance of EHR data in persons living with HIV (PLWH), in whom COPD could be more likely to be either under-diagnosed or misdiagnosed. Greater misdiagnosis could result from the frequent presence of respiratory symptoms and high prevalence of smoking in this population, particularly if spirometry is not performed to confirm the presence of COPD.[15–17] Under-diagnosis may also be more likely to occur as prior studies have shown that spirometry is under-utilized in PLWH and providers may be less aware of smoking status in their patients with HIV.[18,19] As a result, we hypothesized that diagnoses of COPD that rely on ICD-9 codes alone in PLWH could be less accurate in

PLWH. It is important to assess the performance of EHR data to diagnose COPD in PLWH and uninfected so that bias is not introduced in studies that compare differences in COPD in these populations.

We tested several models in order to allow maximum flexibility for future use and to inform performance in datasets where access to certain variables may be limited. Unlike prior studies in which cohorts were derived based on patients referred to the pulmonary function laboratory for clinical indications,[11,13] we utilized data from a research cohort where all participants had spirometry performed to develop a predictive model, decreasing the likelihood of verification bias. We then performed a validation of these results in a separate cohort of PLWH in care in a different health system in whom spirometry had been obtained for clinical purposes.

Methods

Development and Validation Cohorts

For our development cohort, we utilized data from the Examinations of HIV Associated Lung Emphysema (EXHALE) study, a pulmonary substudy of the Veterans Aging Cohort Study (VACS).[20] EXHALE was an observational, longitudinal multicenter study conducted at four of the Veteran Affairs (VA) Medical Centers (VAMC) participating in VACS, namely the Atlanta, Bronx, Houston and Los Angeles sites. Outpatients in VACS were approached for enrollment, which was stratified by HIV and current smoking to obtain a similar proportion of current smokers in the HIV-uninfected (uninfected) participants as in the PLWH. Participants with known history of lung diseases other than COPD or asthma were excluded, as were those with acute respiratory infections or illness in the four weeks prior to baseline measurements. Participants were not required to have COPD or asthma to be included. Results presented here represent the cross-sectional analysis of baseline data from 189 PLWH and 161 uninfected participants who were enrolled from 2009 through 2012. Institutional Review Boards at all locations approved this study, and participants provided written informed consent.

To assess external validity, we used data from the University of Washington (UW) HIV cohort of PLWH in clinical care, which is a participating site in the Centers for AIDS Research (CFAR) Network of Integrated Clinical Systems (CNICS) study.[21] We identified 294 PLWH who were enrolled at the UW site who had spirometry performed for clinical purposes between January, 2000 and October, 2015 (referred to as the UW cohort); 95 of these patients had airflow limitation on testing that was consistent with COPD.

Data Collection

In EXHALE, demographic data, laboratory values and diagnostic codes (ICD-9) for existent medical conditions were obtained via the VA EHR and administrative databases. Variables included age, sex, race, medications and laboratory data. Any metered dose inhalers (MDIs), prescribed prior to research spirometry and for a duration of at least 90 days, were identified from the VA pharmacy databases and consisted of: short-acting beta agonists, anticholinergics, long-acting beta agonists, long-acting muscarinic antagonists,

corticosteroids, and combinations thereof (e.g. albuterol and ipratroprium bromide, or salmeterol and fluticasone). Participants also completed surveys at enrollment, from which we obtained smoking history and respiratory symptoms. Never smokers were defined as those who had smoked less than 100 cigarettes in their lifetime and current smokers as those who had smoked within the past year.

Data for UW participants was obtained from the CNICS data repository, which systematically captures demographic, clinical, medication, and laboratory data for all patients receiving care at each CNICS site from the EHR and other institutional data systems.[21] Quality assessments of data are conducted at the sites prior to data transmission and prior to insertion into the central CNICS data repository by the Data Management Core. We used the medication data to identify the MDIs that were prescribed prior to spirometry for at least 90 days, as in EXHALE. CNICS participants complete a clinical assessment of patient-reported measures and outcomes on touch-screen tablets every 4–6 months as part of routine clinic appointments.[22–24] The CNICS clinical assessment was the source of smoking status data, and was defined similarly as in EXHALE.

Pulmonary function testing (PFT)

Research spirometry in the EXHALE study was performed pre- and post-bronchodilator according to American Thoracic Society (ATS) criteria[25,26] in the clinical pulmonary function test (PFT) laboratory at each participating center. Investigators reviewed results for quality purposes within EXHALE and excluded those with tests not meeting ATS criteria for reproducibility and acceptability. Provider-ordered spirometry in the UW cohort was performed in the clinical PFT laboratory at Harborview Medical Center also in accordance with ATS standards, although we were unable to manually review the individual flow-volume loops to exclude results that might not have met all ATS standards; 7% of the UW cohort had post-bronchodilator spirometry. True COPD was considered present per the Global Obstructive Lung Disease Initiative Guidelines when spirometry confirmed airflow limitation, defined as an FEV1/FVC ratio of less than 0.70.[1] If bronchodilator testing was not performed, results of pre-bronchodilator spirometry were used to define COPD.

Selection of ICD Codes to Defined COPD

We first compared the performance of different ICD-9 code groupings and varied the time window between ICD-9 codes and research spirometry to identify the algorithm that most accurately identified true COPD. As spirometry results were obtained prior to October 2015, ICD-10 codes were not included. Methods and results of these models are described further in the online Supplement. Briefly, we compared results using 1) different ICD-9 code groupings (Figure 1)[4,11,27]; 2) varying the time window to identify ICD-9 codes prior to research spirometry (ranging from ever to 1 year prior); and 3) irrespective of primary or secondary position, requiring ≥ 1 inpatient and/or ≥ 2 outpatient occurrences[27] versus requiring ≥ 1 ICD-9 codes of any type, inpatient or outpatient.

Statistical analysis

We tested the accuracy of COPD ICD-9 codes compared to airflow limitation by spirometry (FEV1/FVC <0.70). Estimates of sensitivity, specificity, positive predictive value (PPV),

negative predictive value (NPV) and their confidence intervals were generated from logistic regression coefficients for each algorithm.[28] In adjusted models, we compared whether accuracy was different according to HIV status and other clinical factors.

As anticipated, model performance with ICD-9 codes alone was poor; thus, we generated a series of predictive models to determine whether additional variables frequently available clinically could improve accuracy to discriminate cases from non-cases of COPD. *A priori*, we hypothesized that age, smoking status, and MDI prescriptions of 90 days would be important predictors in addition to ICD-9 codes. We used Bayesian Model Averaging (BMA) as a technique to help inform variable selection, in which variables selected for model inclusion have a standard cut-off of a predicted probability of 50% or higher for being in the best model; BMA identified age and prior prescription of MDI's for model inclusion in EXHALE. Age was used as a continuous variable centered at age 50. As MDI data were missing in 35 of the EXHALE participants, these patients were considered not exposed to MDI's given the overall low prevalence of MDI prescription; results of all models were similar when these patients were excluded. Smoking status was included in additional models as ever vs. never as this is more accurately obtained from the EHR than pack-years of smoking.[29,30] In order to retain maximum flexibility for future work, we evaluated different combinations of the four predictors. We then evaluated these same models in the UW cohort. Level of discrimination was quantified using the area under the receiver-operating curve (AUC) or c-statistic, where values below <0.70, 0.70 to 0.80, and >0.80 are considered poor, acceptable, and excellent, respectively. All analyses were conducted using SAS v9.2 (Cary, NC) and STATA v13 (College Station, TX).

Results

Characteristics of Development Cohort

A total of 350 participants in EXHALE completed baseline surveys and spirometry. The sample was predominantly black, male, over 50 years old and comprised mostly of former or current smokers (Table 1). About half the participants were PLWH. Overall, 15% had an ICD-9 diagnosis of COPD, but less than half of those with an ICD-9 COPD code had a prior PFT in their VA records, without significant difference by HIV status. The prevalence of airflow limitation consistent with COPD was 20%. COPD was substantially underdiagnosed: nearly two-thirds of participants with airflow limitation did not have ICD-9 codes for COPD. Among those with airflow limitation, those who were undiagnosed tended to be less likely to have chronic cough, phlegm or wheeze compared to participants who had an ICD-9 diagnosis (Table 1). COPD was also misdiagnosed: in those with COPD ICD-9 codes, only 50% had airflow limitation on spirometry.

Validity of COPD ICD-9 codes

The best performing set of COPD ICD-9 codes included 491.x, 492.x, 493.2, and 496; excluded 490; and required 1 inpatient and/or 2 outpatient occurrences at any time prior to research spirometry (Figure 1 and eTable1). This resulted in the highest AUC (0.638, 95% confidence interval [CI] 0.578–0.698), with a sensitivity of 37%, specificity of 90%, PPV of 49% and NPV of 85%. The sensitivity, specificity, PPV and NPV of COPD ICD-9 codes

were similar in PLWH and uninfected individuals and when adjusted for other factors (Table 2).

Predictive Model for COPD in EXHALE

We next generated a series of predictive models to identify COPD cases using additional data from the EHR (Table 3). Using BMA, age and prior prescription of MDI's for 90 days had a predicted probability of 50% for being in the model (Model 2). ICD-9 codes and smoking status, however, had less than a 50% predicted probability of being included in the model. Because we hypothesized that these would be important in other datasets and for flexibility in other settings, we evaluated models with different combinations of these four variables. All models had an AUC that was significantly better than ICD-9 codes alone (p<0.01). A model with all four variables (COPD ICD-9 codes, ever smoking, age, and MDI's, Model 5) had the highest AUC at 0.772 (95% CI 0.709–0.834) but was similar to a model with age, MDIs, and ever smoking (Model 4, AUC 0.771, 95% CI 0.708–0.834). In sensitivity analyses, we excluded short-acting beta-agonists and also restricted MDI's to only long-acting COPD medications, but likely due to overall low frequency of use, the AUCs were not improved (data not otherwise shown).

External Validation in the UW Cohort

Within the UW cohort (N=4126), we identified 294 PLWH who had spirometry for clinical indications. The mean age of patients was 49 (SD 9), and 80% were ever smokers, 84% were male, 67% were white, 21% were black, and 6% were Hispanic. Similar to the prevalence in the development cohort, 27% had ICD-9 diagnoses of COPD prior to spirometry; 23% were on COPD medications. Overall, 95 (32%) had confirmed airflow limitation on spirometry; of these, 45 (47%) had an ICD-9 code for COPD prior to spirometry. Of the 79 patients (27%) who had an ICD-9 code for COPD prior to spirometry, 45 (57%) were found to have airflow obstruction on testing.

We generated similar models to determine the accuracy to diagnose COPD. As within EXHALE, model performance by AUC was poor for ICD-9 codes alone (0.651, 95% CI 0.595–0.708, Model 1). The best discrimination occurred when including age, prior MDI's, COPD ICD-9 codes, and ever vs. never smoking status to identify COPD, with an AUC of 0.734 (95% CI 0.675–0.792, Model 5, statistically significantly higher AUC when compared to Model 1). A model with age, COPD ICD-9 codes, and smoking status (Model 3, AUC 0.716, 95% CI 0.654–0.779) also performed statistically significantly better than Model 1 to identify cases of COPD. As a sensitivity analysis we evaluated the inclusion of COPD ICD-9 codes from prior to and up until 12 months after spirometry, as per Cooke et al.;[11] the resulting AUC increased to 0.758 (95% CI 0.702–0.813) when also including age, prior MDI's, and smoking status.

Discussion

In this study, we found that ICD-9 codes were poor at discriminating cases from non-cases of COPD in both PLWH and uninfected populations when compared to the gold standard of spirometry to detect airflow limitation. However, when combined with or substituted for

other clinical variables that are obtainable within the EHR – namely age, ever smoking, and prior prescription of at least 90 days of any MDI's used for COPD, discrimination was adequate both within the VA-based EXHALE study (AUC 0.772) and within the UW cohort of PLWH in care (AUC 0.734). These data support that EHR data can be used to identify COPD cases in PLWH and uninfected individuals with acceptable and similar accuracy in both groups when results of spirometry are not available.

Overall, we found a lower accuracy of administrative data to diagnose COPD in our VA cohort compared to an optimal AUC of 0.79 in work by Cooke et al,[11] although our results are similar to a Canadian study.[13] The study by Cooke et al. consisted of a cohort of patients who had been clinically referred for pulmonary function tests within the VA with a 47% prevalence of true COPD, potentially explaining these differences. Unlike the algorithm by Cooke et al., ours includes smoking status and uses prescriptions for 90-days or more for MDIs rather than counting cannisters prescribed; in our primary approach, we also restricted our analyses to using variables that were recorded prior to spirometry. We found that a COPD definition that included ICD-9 codes for chronic obstructive asthma, chronic bronchitis, and emphysema – but excluded non-specific bronchitis (490) – had the best discrimination of the COPD ICD-9 code groupings. Requiring ≥1 inpatient and/or ≥2 outpatient ICD-9 codes for COPD in any position generally resulted in better model performance, and using ICD-9 codes at any time or within 5 years prior to research spirometry compared to limiting to one or two years prior to PFT resulted in statistically significantly better AUC. This could potentially reflect a lack of clinical activity centered on COPD care for these patients at recent appointments, yet nonetheless many of these patients had true COPD. Overall, the best performing ICD-9 code algorithm had a good specificity (90%), but a poor sensitivity for COPD diagnosis (37%). Similar to other studies,[11,13] the PPV of ICD-9 codes in our cohort was poor, though somewhat higher when restricted to ICD-9 codes within the previous 1–2 years (PPV ranging from 49–63%). Notably, we found no difference in the performance of ICD-9 codes to diagnose COPD by HIV status, despite our concern that COPD may be more likely to be misdiagnosed or under-diagnosed in PLWH. Notably, misdiagnosis of COPD did occur in approximately 50% of PLWH based on having a COPD ICD-9 code but no airflow obstruction on spirometry. Under-diagnosis of COPD was even more common: two-thirds of those who had airflow limitation in EXHALE and half of those at UW did not have a COPD ICD-9 code prior to spirometry.

Given the poor discrimination of ICD-9 codes for COPD, we evaluated several other variables that are associated with COPD to improve diagnostic accuracy. In a model that included age, ever smoking, prescription of MDI's for at least 90 days, and ICD-9 codes for COPD, overall discrimination improved to an adequate range both in our development and validation cohorts. Notably, AUC's were also adequate in models that excluded ICD-9 codes or MDI's, pointing to ways that models might be adapted to availability of local data or for different analytic purposes.

This is the first study to assess the performance of EHR data in addition to standardly used ICD-9 codes to diagnose COPD in a diverse cohort that included PLWH. A major strength of this study is that we included both development and validation cohorts, and that the validation cohort consisted of a different population derived from a different healthcare

system, increasing external generalizability. Further, all participants in our development cohort completed spirometry and data were not based on results from clinical referrals to the pulmonary function laboratory, thereby limiting verification bias. Additionally, we included a large sample of minority individuals from several geographic regions.

A limitation to our study was that our development sample included few women. In the UW study, 16% were women; while consistent with the HIV epidemic in the US, which is predominantly male, it limits ability to make conclusions regarding women. The sample size of both cohorts was relatively small, though nonetheless we did find statistically significantly different results in model performance when comparing AUCs. Given limitations in our ability to detect significant differences by demographic characteristics, researchers may wish to further validate our results in larger cohorts in other diverse settings. Our validation cohort consisted only of PLWH who had been referred for spirometry for clinical indications, potentially introducing verification bias; while this might have introduced bias towards better model performance, this nonetheless mimics the clinical scenarios and data available through the EHR for analyses – namely, outside of a research study, patients are only typically referred for spirometry if there is clinical suspicion of pulmonary limitation. Post-bronchodilator spirometry was rarely performed in the validation cohort, potentially resulting in an over-estimation of the true prevalence of airflow limitation, and we were unable to review the individual flow-volume curves to exclude maneuvers that did not meet standard ATS quality criteria. Individuals with asthma may also have been included as true COPD, but given the high prevalence of smoking in both cohorts, concomitant COPD would be clinically difficult to exclude. In addition, although we used self-reported smoking status in our models, smoking status is increasingly available in many EHR systems; we have previously validated self-report of smoking status within the VA EHR.[30] Finally, while these results are valid to inform many ongoing studies, future work will require validation of ICD-10 codes, a process that can be informed by these analyses.

In conclusion, we found that ICD-9 codes for COPD are poor in unadjusted models for predicting airflow limitation as detected by spirometry. Notably, performance of ICD-9 codes was not significantly better or worse in PLWH compared to uninfected individuals. However, a model including age, ever smoking, prescription of at least 90 days of any MDI's, and COPD ICD-9 codes resulted in significantly improved discrimination to diagnose COPD. The AUC in our development and validation cohort were adequate. Larger scale epidemiologic studies may consider use of these algorithms to diagnose COPD with acceptable accuracy when spirometry results are not available, and may wish to perform validation of these algorithms within their own data prior to use. Nonetheless, our findings underscore the need to develop resources to obtain results of spirometry, such as by including test results as searchable data fields in the EHR. Future work can also consider text searching or natural language processing efforts to identify spirometry results in large electronic databases. Finally, our work highlights the need to improve the diagnostic evaluation for COPD given the under-diagnosis and misdiagnosis of COPD in PLWH as in uninfected individuals.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Global Strategy for the Diagnosis, Management and Prevention of COPD, Global Initiative for Chronic Obstructive Lung Disease (GOLD). 2017.

2. Lamprecht B, Soriano JB, Studnicka M, et al. Bold Collaborative Research Group, the E. P. I. Scan Team the Platino Team Prepocol Study Group, Bold Collaborative Research Group the EPI-SCAN Team the PLATINO Team Prepocol Study Group. Determinants of underdiagnosis of COPD in national and international surveys. Chest. 2015;148(4):971–985. [PubMed: 25950276]

3. Collins BF, Feemster LC, Rinne ST, Au DH. Factors predictive of airflow obstruction among veterans with presumed empirical diagnosis and treatment of COPD. Chest. 2015;147(2):369–376. [PubMed: 25079684]

4. Prieto-Centurion V, Rolle AJ, Au DH, et al. Multicenter study comparing case definitions used to identify patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2014;190(9):989–995. [PubMed: 25192554]

5. Ferguson GT, Enright PL, Buist AS, Higgins MW. Office spirometry for lung health assessment in adults: A consensus statement from the National Lung Health Education Program. Chest. 2000;117(4):1146–1161. [PubMed: 10767253]

6. Joo MJ, Sharp LK, Au DH, Lee TA, Fitzgibbon ML. Use of spirometry in the diagnosis of COPD: a qualitative study in primary care. COPD. 2013;10(4):444–449. [PubMed: 23537230]

7. Wu H, Wise RA, Medinger AE. Do Patients Hospitalized With COPD Have Airflow Obstruction? Chest. 2017;151(6):1263–1271. [PubMed: 28089815]

8. Lee TA, Bartle B, Weiss KB. Spirometry use in clinical practice following diagnosis of COPD. Chest. 2006;129(6):1509–1515. [PubMed: 16778268]

9. Qaseem A, Wilt TJ, Weinberger SE, et al. Diagnosis and management of stable chronic obstructive pulmonary disease: a clinical practice guideline update from the American College of Physicians, American College of Chest Physicians, American Thoracic Society, and European Respiratory Society. Ann Intern Med. 2011;155(3):179–191. [PubMed: 21810710]

10. Konstantikaki V, Kostikas K, Minas M, et al. Comparison of a network of primary care physicians and an open spirometry programme for COPD diagnosis. Respiratory medicine. 2011;105(2):274–281. [PubMed: 20663655]

11. Cooke CR, Joo MJ, Anderson SM, et al. The validity of using ICD-9 codes and pharmacy records to identify patients with chronic obstructive pulmonary disease. BMC health services research. 2011;11:37. [PubMed: 21324188]

12. Gershon AS, Wang C, Guan J, Vasilevska-Ristovska J, Cicutto L, To T. Identifying individuals with physcian diagnosed COPD in health administrative databases. COPD. 2009;6(5):388–394. [PubMed: 19863368]

13. Lacasse Y, Daigle JM, Martin S, Maltais F. Validity of chronic obstructive pulmonary disease diagnoses in a large administrative database. Canadian respiratory journal : journal of the Canadian Thoracic Society. 2012;19(2):e5–9.

14. McKnight J, Scott A, Menzies D, Bourbeau J, Blais L, Lemiere C. A cohort study showed that health insurance databases were accurate to distinguish chronic obstructive pulmonary disease from asthma and classify disease severity. J Clin Epidemiol. 2005;58(2):206–208. [PubMed: 15680756]

15. Drummond MB, Kirk GD, Astemborski J, et al. Prevalence and risk factors for unrecognized obstructive lung disease among urban drug users. Int J Chron Obstruct Pulmon Dis. 2011;6:89–95. [PubMed: 21407821]

16. Gingo MR, George MP, Kessinger CJ, et al. Pulmonary Function Abnormalities in HIV-infected Patients During the Current Antiretroviral Therapy Era. Am J Respir Crit Care Med. 2010.

17. Crothers K, McGinnis K, Kleerup E, et al. HIV infection is associated with reduced pulmonary diffusing capacity. J Acquir Immune Defic Syndr. 2013;64(3):271–278. [PubMed: 23979001]

18. Gingo MR, Balasubramani GK, Rice TB, et al. Pulmonary symptoms and diagnoses are associated with HIV in the MACS and WIHS cohorts. BMC Pulm Med. 2014;14:75. [PubMed: 24884738]

19. Crothers K, Goulet JL, Rodriguez-Barradas MC, et al. Decreased awareness of current smoking among health care providers of HIV-positive compared to HIV-negative veterans. Journal of general internal medicine. 2007;22(6):749–754. [PubMed: 17503106]

20. 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. [PubMed: 16849964]

21. Kitahata MM, Rodriguez B, Haubrich R, et al. Cohort profile: the Centers for AIDS Research Network of Integrated Clinical Systems. Int J Epidemiol. 2008;37(5):948–955. [PubMed: 18263650]

22. Cropsey KL, Willig JH, Mugavero MJ, et al. Cigarette Smokers are Less Likely to Have Undetectable Viral Loads: Results From Four HIV Clinics. J Addict Med. 2016;10(1):13–19. [PubMed: 26656939]

23. Fredericksen R, Crane P, Tufano J, et al. Integrating a web-based patient assessment into primary care for HIV-infected adults. Journal of AIDS and HIV Research. 2012;4(2):47–55. [PubMed: 26561537]

24. Crane HM, Lober W, Webster E, et al. Routine collection of patient-reported outcomes in an HIV clinic setting: the first 100 patients. Curr HIV Res. 2007;5(1):109–118. [PubMed: 17266562]

25. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. Eur Respir J. 2005;26(2):319–338. [PubMed: 16055882]

26. Miller MR, Crapo R, Hankinson J, et al. General considerations for lung function testing. Eur Respir J. 2005;26(1):153–161. [PubMed: 15994402]

27. Justice AC, Lasky E, McGinnis KA, et al. Medical disease and alcohol use among veterans with human immunodeficiency infection: A comparison of disease measurement strategies. Med Care. 2006;44(8 Suppl 2):S52–60. [PubMed: 16849969]

28. Coughlin SS, Trock B, Criqui MH, Pickle LW, Browner D, Tefft MC. The logistic modeling of sensitivity, specificity, and predictive value of a diagnostic test. J Clin Epidemiol. 1992;45(1):1–7. [PubMed: 1738006]

29. Modin HE, Fathi JT, Gilbert CR, et al. Pack-Year Cigarette Smoking History for Determination of Lung Cancer Screening Eligibility. Comparison of the Electronic Medical Record versus a Shared Decision-making Conversation. Annals of the American Thoracic Society. 2017;14(8):1320–1325. [PubMed: 28406708]

30. McGinnis KA, Brandt CA, Skanderson M, et al. Validating smoking data from the Veteran's Affairs Health Factors dataset, an electronic data source. Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco. 2011;13(12):1233–1239. [PubMed: 21911825]

Take home points

- Persons living with HIV infection (PLWH) have an increased risk of chronic obstructive pulmonary disease (COPD); no prior studies have evaluated whether electronic health record (EHR) data performs similarly to identify PLWH who have COPD.

- On their own, ICD-9 diagnostic codes have poor accuracy for the diagnosis of COPD compared to the gold standard of spirometry.

- However, the addition of other data available in EHR – namely age, smoking status, and prescription of inhalers for COPD – can substantially improve the ability to correctly identify individuals with COPD when compared to spirometry.

- EHR data performs similarly in PLWH and in individuals without HIV to identify COPD.

COPD1: Consists of the following diagnoses:
490   **Bronchitis**
491.0 Simple chronic bronchitis
491.1 Mucoprulent chronic bronchitis
491.2 Obstructive chronic bronchitis
491.2 W/out mention of acute exacerbation
491.21 W/acute exacerbation
491.8 Other chronic exacerbation
491.9 Unspecified chronic bronchitis
492.0 Emphysematous bleb
492.8 Emphysema nec
496   Chronic airway obstruction – nec

COPD2: Consists of the following diagnoses:
491.0 Simple chronic bronchitis
491.1 Mucoprulent chronic bronchitis
491.2 Obstructive chronic bronchitis
491.2 W/out mention of acute exacerbation
491.21 W/acute exacerbation
491.8 Other chronic exacerbation
491.9 Unspecified chronic bronchitis
492.0 Emphysematous bleb
492.8 Emphysema nec
493.2 **Chronic obstructive asthma w/o stat asth**
496   Chronic airway obstruction – nec

Differences between definitions in bold.

Figure 1.
ICD-9 Definitions of COPD

Table 1.

Characteristics of EXHALE study population, by best ICD-9 definition of COPD or airflow limitation on spirometry

		Total	ICD-9 COPD+	ICD-9 COPD-	p-value	AFL+	AFL-	p-value
		350	53	297		70	280	
Age					0.02			<0.01
	<50	92	7 (8)	85 (92)		5 (5)	87 (95)	
	50	258	46 (18)	212 (82)		65 (25)	193 (75)	
Race/Ethnicity					<0.01			<0.01
	Black	239	28 (13)	211 (88)		36 (15)	203 (85)	
	White	58	18 (31)	40 (69)		20 (34)	38 (66)	
	Hispanic	53	7 (13)	46 (87)		14 (26)	39 (74)	
Sex					0.99			0.57
	Male	330	50 (15)	280 (85)		67 (20)	263 (80)	
	Female	20	3 (15)	17 (85)		3 (15)	17 (85)	
Smoking Status					0.08			0.03
	Never	56	4 (7)	52 (93)		4 (7)	52 (93)	
	Former	80	10 (13)	70 (88)		20 (25)	60 (75)	
	Current	211	39 (18)	172 (82)		45 (21)	166 (79)	
HIV-infected					0.85			0.96
	Yes	189	28 (15)	161 (85)		38 (20)	151 (80)	
	No	161	25 (16)	136 (84)		32 (20)	129 (80)	
COPD by ICD-9 code								<0.01
	Yes	53				26 (49)	27 (51)	
	No	297				44 (15)	253 (85)	
Asthma by ICD-9 code					<0.01			<0.01
	Yes	65	27 (42)	38 (58)		21 (32)	44 (68)	
	No	285	26 (9)	259 (91)		49 (17)	236 (83)	
Prior PFT for clinical indications					<0.01			<0.01
	Yes	58	22 (38)	36 (62)		23 (40)	35 (60)	
	No	292	31 (11)	261 (89)		47 (16)	245 (84)	
Research PFT results								

	Total	ICD-9 COPD+	ICD-9 COPD-	p-value	AFL+	AFL-	p-value
FEV1, % pred, mean (SD)	91.0 (18.3)	76.0 (20.6)	93.6 (16.6)	<0.01	76.4 (17.8)	94.7 (16.5)	<0.01
FVC, % pred, mean (SD)	94.4 (16.1)	90.7 (15.4)	95.0 (16.1)	0.08	96.7 (15.9)	93.8 (16.1)	0.17
FEV1/FVC<0.7 pre-BD, %				<0.01			<0.01
Yes	96	33 (34)	63 (66)				
No	254	20 (8)	234 (92)				
FEV1/FVC<0.7 post-BD, %				<0.01			
Yes	69	25 (36)	44 (64)				
No	273	25 (9)	248 (91)				
DLCO, % pred, mean (SD)	55.9 (16.4)	51.1 (7.7)	56.7 (16.0)	0.02	49.1 (15.7)	57.6 (16.1)	<0.01
Chronic cough/phlegm				<0.01			0.06
Yes	163	36 (22)	127 (78)		39 (24)	124 (76)	
No	184	17 (9)	167 (91)		29 (16)	155 (84)	
Wheeze (Ever)				<0.01			<0.01
Yes	154	38 (25)	116 (75)		43 (28)	111 (72)	
No	190	15 (8)	175 (92)		26 (14)	164 (86)	
Wheeze (Last year)				<0.01			0.15
Yes	131	29 (22)	102 (78)		31 (24)	100 (76)	
No	214	24 (11)	190 (89)		37 (17)	177 (83)	
Dyspnea				<0.01			<0.01
Yes	133	33 (25)	100 (75)		37 (28)	96 (72)	
No	173	15 (9)	158 (91)		24 (14)	149 (86)	
Short-acting inhaler				<0.01			<0.01
Yes	39	24 (62)	15 (38)		22 (56)	17 (44)	
No	276	19 (7)	257 (93)		43 (16)	233 (84)	
Long-acting inhaler				<0.01			<0.01
Yes	14	11 (79)	3 (21)		8 (57)	6 (43)	
No	301	32 (11)	269 (89)		57 (19)	244 (81)	

Results given as n (%) unless otherwise indicated, excluding those with missing data for that variable. Best ICD-9 definition of COPD requires 1 COPD inpatient ICD-9 and/or 2 COPD outpatient ICD-9 codes ever prior to research PFT, using 491.x, 492.x, 493.2, and 496.

Missing data: Smoking status missing in 3; post-bronchodilator spirometry in 8; chronic cough/phlegm in 3; wheeze (ever) in 6; wheeze (last year) 5; dyspnea in 44; and inhaler data in 35 participants.

Abbreviations:

AFL+ = airflow limitation consistent with COPD (FEV1/FVC<0.7)

AFL- = airflow limitation not consistent with COPD (FEV1/FVC 0.7)

BD = bronchodilator

DLCO = lung diffusing capacity

EXHALE = Examinations of HIV Associated Lung Emphysema

FEV1 = forced expiratory volume in one second

FVC = forced vital capacity

Pred = predicted

Table 2.

Adjusted Validity Estimates of best ICD-9 diagnosis of COPD compared to FEV1/FVC <0.7 by spirometry

Adjustment Characteristic	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV
Unadjusted model	0.638 (0.578 – 0.698)	37%	90%	49%	85%
Age	0.714 (0.655 – 0.772)				
50 yrs		40%	90%	57%	82%
< 50 yrs *	0.698 (0.629 – 0.767)	0%	92%	0%	94%
Race/Ethnicity	0.698 (0.629 – 0.767)				
Black		33%	92%	43%	89%
White *		45%	76%	50%	73%
Hispanic		36%	95%	71%	80%
Smoking Status	0.678 (0.608 – 0.748)				
Current		36%	86%	41%	83%
Former		45%	98%	90%	84%
Never *		25%	94%	25%	94%
Inhalers	0.667 (0.603 – 0.732)				
Prescription of >90 days, ever		78%	59%	72%	67%
No prescription of >90 days *		17%	92%	29%	86%
HIV Status	0.639 (0.564 – 0.714)				
Positive		37%	91%	50%	85%
Negative *		38%	90%	48%	85%

*
reference for comparison

Abbreviations:

AUC = area under the receiver operating curve

CI = confidence interval

NPV = negative predictive value

PPV = positive predictive value

Table 3.

Predictive models to diagnose COPD: Development and Validation Cohorts

Cohort	Variables Included					AUC	SD	95% CI	
EXHALE – Development Cohort (N=348)	**ICD-9 codes**	**Age**	**MDI's**	**Ever Smoker**		**AUC**	**SD**	**LL**	**UL**
Model 1	X					0.633	0.032	0.570	0.696
Model 2 *		X	X			0.766	0.031	0.705	0.827
Model 3	X	X		X		0.748	0.033	0.684	0.812
Model 4		X	X	X		0.771	0.032	0.708	0.834
Model 5	X	X	X	X		0.772	0.032	0.709	0.834
UW – Validation Cohort (N=294)	**ICD-9 codes**	**Age**	**MDI's**	**Ever Smoker**		**AUC**	**SD**	**LL**	**UL**
Model 1	X					0.651	0.029	0.595	0.708
Model 2		X	X			0.675	0.033	0.612	0.739
Model 3	X	X		X		0.716	0.032	0.654	0.779
Model 4		X	X	X		0.714	0.030	0.654	0.773
Model 5	X	X	X	X		0.734	0.030	0.675	0.792

*
BMA identified model

MDIs = metered dose inhalers

LL – Lower limit

UL – Upper limit

EXHALE:

P-value that Model 1 AUC is different than Model 2 (**<0.01**), Model 3 (**<0.01**), Model 4 (**<0.01**) and Model 5 (**<0.01**)

P-value that Model 3 AUC is different than Model 2 (0.46), Model 4 (0.26) and Model 5 (0.09)

P-value that Model 4 AUC is different than Model 2 (0.70) and Model 5 (0.96)

P-value that Model 5 AUC is different than Model 2 (0.70)

UW Cohort:

P-value that Model 1 AUC is different than Model 2 (0.54), Model 3 (**<0.01**), Model 4 (0.09) and Model 5 (**<0.01**)

P-value that Model 3 AUC is different than Model 2 (0.22), Model 4 (0.93) and Model 5 (0.34)

P-value that Model 4 AUC is different than Model 2 (**0.02**) and Model 5 (0.11)

P-value that Model 5 AUC is different than Model 2 (**<0.01**)