Published in final edited form as:

Pharmacoepidemiol Drug Saf. 2020 November ; 29(11): 1432–1439. doi:10.1002/pds.5111.

High Concordance between Chart Review Adjudication and Electronic Medical Record Data to Identify Prevalent and Incident Diabetes Mellitus among Persons With and Without HIV

Kathleen A. McGinnis, MS, DrPH^a, Amy C. Justice, MD, PhD^{a,b,c}, Sam Bailin, MD^d, Melissa Wellons, MD^d, Matthew Freiberg, MD^{d,e}, John R. Koethe, MD^{d,e}

^aVeterans Affairs Connecticut Healthcare System, West Haven, CT, USA

^bDepartment of Internal Medicine, Yale School of Medicine, New Haven, CT, USA

^cSchool of Public Health, Yale School of Medicine, New Haven, CT, USA

^dDepartment of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

^eVeterans Affairs Tennessee Valley Healthcare System, Nashville, TN, USA

Abstract

Background: Electronic Medical Records (EMR) represent a rich source of data, but the value of EMR for health research relies on accurate ascertainment of clinical diagnoses. Identifying diabetes in EMR is complicated by the variety of accepted diagnostic criteria, some of which can be confounded by conditions such as HIV infection. We compared EMR-based criteria for estimating diabetes prevalence and incidence in the Veterans Administration (VHA), overall and by HIV status, against physician chart review and adjudication.

Research Design and Methods: We used laboratory values (serum glucose and hemoglobin A1c% [HbA1c]), ICD-9 codes, and medication records from the United States Veterans Aging Cohort Study Biomarker Cohort (VACS-BC) to identify veterans with any indication of diabetes in the EMR for subsequent physician adjudication. Sensitivity, specificity, PPV, NPV, and kappa statistics were used to evaluate agreement of EMR-based diabetes diagnoses with chart review adjudicated diagnoses.

Results: EMR entries were reviewed for 1546 persons with HIV (PWH) and 843 HIV-negative participants through 2015. Agreement was at least moderate overall (kappa = 0.42) for all pre-specified measures and among PWH vs. HIV-negative, and African-American vs. white subgroups. Having at least one HbA1c ≥ 6.5% provided substantial agreement with chart adjudication for prevalent and incident diabetes (kappa = 0.89 and 0.73).

Conclusions: Identification of those with diabetes nationally within the VHA can be used in future studies to evaluate treatments, health outcomes, and adjust for diabetes in epidemiologic

---

Corresponding Author: Kathleen McGinnis, DrPH, MS, VA Connecticut Healthcare System, 950 Campbell Avenue, Building 35a 2nd Floor (11-ACSLG), West Haven, CT 06516, Tel: 203.932.5711 x3541, Fax: 203.937.4926.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

studies. Our methodology may provide insights for other organizations seeking to use EMR data for accurate determination of diabetes.

Keywords

diabetes mellitus; electronic medical record; HIV; validation

Introduction

Electronic Medical Records (EMR) represent a rich source of data, but the value of the EMR for health research is dependent on the ability to accurately ascertain clinical disease diagnoses.^{1,2} The identification of persons with diabetes mellitus in EMR is complicated by the variety of accepted diagnostic criteria, the lack of contextual information to interpret laboratory values (e.g., fasting status or the concomitant use of medications associated with transient hyperglycemia), and the inadvertent entry of incorrect diagnostic codes by providers.^{3,4} Additionally, there can be variation in the mapping of EMR fields to source data as observed in a prior study which found numerous and differing missing values in two large electronic laboratory databases populated from the same source.⁵

In the United States, 34 million adults are estimated to have diabetes, representing over 12% of the adult population. Diabetes is the seventh leading cause of death in the United States in 2018 and a major contributor to other comorbidities.⁶ Accurate identification of diabetes is critical for epidemiologic health outcomes and other analyses, yet up to a quarter of persons with diabetes in the United States are estimated to lack a formal diagnosis.⁶ Current guidelines define diabetes according to fasting plasma glucose, oral glucose tolerance results, and hemoglobin A1c (HbA1c) thresholds.⁷ Among laboratory values, HbA1c is more often available in EMRs because it doesn't require fasting and can be routinely collected at clinical visits. HbA1c represents an indirect measure of blood glucose levels averaged over time but is susceptible to factors affecting hemoglobin glycation including age, race, hemoglobinopathies, anemia, genetics, and HIV infection, and current thresholds for disease diagnosis may not be as accurate in some subgroups.⁷⁻¹³

The United States Veterans Health Administration (VHA) maintains one of the most highly developed health information systems in the world, and VHA EMR data have been used to identify diabetes using varying combinations of ICD-9 codes, lab values, and medication data.^{14,15} The Veterans Aging Cohort Study (VACS) previously validated a composite definition employing diagnostic codes, laboratory values, and medications to identify prevalent diabetes,¹⁶ but for incident diabetes used a single HbA1c value $\geq 6.5\%$ ^{7,17} without formally validating this choice.¹⁸ In this study, we use chart adjudication to establish the validity of several EMR options for identifying prevalent and incident diabetes for participants with and without HIV in the VACS Biomarker Cohort (VACS-BC).

Methods

Data Source

The VACS-BC is a subset of the VACS Survey 2002,¹⁹ a prospectively enrolled, observational, longitudinal study of United States veterans with and without an HIV diagnosis matched on age, race-ethnicity (combined such that if a person was Hispanic and any race, they were included in the Hispanic group), sex, and geographic region. From September 2005 to September 2007, 1546 people with HIV (PWH) and 843 HIV-negative VACS Survey participants provided blood samples to comprise the VACS-BC specimen archive, a sub-cohort constituted to investigate innate and adaptive immune system correlates of cardiovascular and metabolic diseases.²⁰ VACS participants consented to sample and data collection, and institutional review boards at Yale University, West Haven VA Medical Center and Vanderbilt University approved this analysis.

Chart Adjudication for Diabetes Mellitus

A previously developed EMR algorithm (Table 1) scanned participant records from the time of entry into VA care through 9/30/2015 for the presence of specific pharmacy data, laboratory values, and ICD-9 codes that could signify cases of diabetes using data available through the Corporate Data Warehouse (CDW).¹⁶ This algorithm has been used in several VACS studies to identify prevalent diabetes,^{16,21} but had not previously been used to identify incident diabetes.

The EMR scan identified 865 medical records for subsequent adjudication by two physicians with expertise in endocrinology and metabolism to assess whether elements of the composite diabetes diagnosis criteria (Table 1) were present. Prevalent (i.e., present at the time of blood collection) versus incident (i.e., after blood collection) diabetes was determined, and onset date was assigned for incident cases (described below). The physician reviews were merged and any discrepancies regarding the presence of diabetes or the onset date were given a final adjudication by a third physician with expertise in internal medicine and metabolism. Some reasons identified by adjudicators for being flagged by the EMR algorithm but not validated in chart review are: 1) spurious laboratory values (e.g., glucose measurements >1000 during inpatient hospitalizations), 2) documented secondary causes of temporary hyperglycemia (e.g., short-term high-dose steroid use), and 3) the presence of ICD-9 codes in the absence of confirmatory laboratory or medication data or further documentation in the chart (see Supplementary Table 1 for example cases rejected on adjudication).

For patients who met at least one of the six criteria of the composite diabetes definition (Table 1) during chart adjudication, the date of onset was based on a series of sequential criteria: 1) the date of the first HbA1c $\geq 6.5\%$; 2) in the absence of a HbA1c value $\geq 6.5\%$ (or if no HbA1c were measured) the date of the second glucose >200 mg/dl; or 3) the date of the only glucose >200 mg/dl among those patients who also received an outpatient oral hypoglycemic or insulin for >30 days.

EMR Measures for Diabetes Mellitus

We compared several EMR-based strategies for identifying prevalent and incident diabetes against the dual-physician chart adjudication results. Our primary definition for diabetes was based on having at least one HbA1c $\geq 6.5\%$ in accordance with the 2019 American Diabetes Association guidelines.⁷ We also created four additional definitions: 1) at least two HbA1c $\geq 6.5\%$ at any time points but not on the same day; 2) at least two HbA1c $\geq 6.5\%$ less than one year apart and not on the same day; 3) one or more inpatient or two or more outpatient ICD-9 codes 250.0–250.9 or 357.2, and 4) the EMR composite diabetes definition as listed in Table 1. For the EMR composite diabetes definition, the date of onset is based on the first date that any criteria were met; each of the six criteria requires at least two events to occur and the date of the 2nd event is used as the date on which the criteria was met. As with the chart adjudication, our EMR-based strategies considered diabetes to be prevalent if the first date of diabetes for that measure occurred prior to the blood specimen collection date, and incident if it occurred on or after that date.

ICD-10 replaced ICD-9 codes in the VA in October 2015. We determined who had a diabetes ICD-10 code (E10, E11, E13, Z46.81, and Z96.41) from October 2015 to February 2018. These codes were chosen based on 1) reviewing other groups' recommendations, 2) the expert opinion of an endocrinologist, 3) comparing frequencies of different ICD-10 codes and crosstabs, 4) and a comparison of various combinations of ICD-10 codes to presence of HbA1c ≥ 6.5 .

Statistical Analysis

Demographic and clinical characteristics were compared between veterans with and without an HIV diagnosis. Prevalent and incident diabetes rates were calculated. Corresponding standardized differences were calculated for comparison by HIV status. For EMR incident diabetes measures, we excluded those meeting the EMR composite definition for prevalent diabetes. For chart adjudicated incident diabetes, we excluded those with chart adjudicated prevalent diabetes. Operating statistics with 95% confidence intervals were generated for the EMR-based definitions using the physician-adjudicated record as the gold standard. Specifically, sensitivity represents the percent identified with EMR diabetes out of those with chart adjudicated diabetes; specificity represents the percent identified without EMR diabetes out of those without diabetes based on chart adjudication. Positive predictive value represents the percent with chart adjudicated diabetes out of all those identified with EMR diabetes. Negative predictive value represents the percent without chart adjudicated diabetes out of all those identified as not having EMR diabetes. To quantify overall agreement, we used kappa statistics which consider agreement that would be expected by chance. The kappa statistic ranges from 0 to 1, with 0 representing chance alone versus 1 representing perfect agreement. Intermediate values were interpreted as follows: 0.00—poor, 0.01 to 0.20—slight, 0.21 to 0.40—fair; 0.41 to 0.60—moderate; 0.61 to 0.80—substantial; 0.81 to 1.00—almost perfect²².

We also compared time between date of incidence based on chart adjudication versus EMR measures for those who had dates for both. For the HbA1c definitions requiring 2 or more values ≥ 6.5 , the second date was used as the incidence date. As a sensitivity analysis, we

repeated the EMR-defined incident diabetes analyses after excluding patients with prevalent diabetes based on chart adjudication rather than based on the EMR composite definition.

Lastly, we compared presence of a diabetes based on ICD-10 codes to chart adjudication. We included only those who were alive up to October 2016 to allow time for an ICD-10 code to be captured (n=1118 PWH and 635 HIV-negative).

Results

Mean age was 52 and ranged from 25 to 91. The participants were 95% male, 68% African-American, 20% white, and 8% Hispanic; 65% had HIV infection by VACS Survey 2002 design (Table 2). Of the 865 patient charts manually adjudicated, 89 (10%) did not meet the criteria for either prevalent or incident diabetes; 529 had prevalent diabetes and 247 had incident diabetes (Supplemental Figure S1). Chart adjudicated diabetes onset dates ranged from 1993 to 2017.

Diabetes Prevalence and Agreement

Chart adjudication identified prevalent diabetes in 18% of PWH and 29% of the HIV-negative, while prevalence measures based on different EMR criteria ranged from 10% to 22% in PWH and 19% to 37% in the HIV-negative (Table 2). Based on kappa statistics, two EMR measures provided the best agreement with chart adjudication: 1) the EMR composite definition (kappa = 0.90 for both PWH and HIV-negative); and 2) having at least one HbA1c $\geq 6.5\%$ (kappa = 0.87 and 0.92 for PWH and HIV-negative, respectively) (Table 3). However, all of the EMR measures provided substantial agreement or better (i.e., kappa ≥ 0.65) with chart adjudication. Based on chart adjudication, diabetes prevalence among PWH was 18% for African-Americans and 14% for white participants; among the HIV-negative, diabetes prevalence was 28% for African-Americans and 30% for white participants. The operating characteristics and kappa statistics were similar between African-American and white PWH and HIV-negative participants (Supplemental Table S2).

Diabetes Incidence and Agreement

Overall, chart adjudication identified incident diabetes in 11% of the PWH and 19% of the HIV-negative, while incidence ranged from 5% to 15% for PWH and 10% to 28% for HIV-negative using the EMR measures (Table 2). Based on kappa statistics, the EMR measures for incident diabetes providing the best agreement with chart adjudication are: 1) having at least two HbA1c $\geq 6.5\%$ (kappa = 0.76 and 0.81 for PWH and HIV-negative, respectively); 2) the EMR composite criteria (kappa = 0.76 and 0.78, respectively); and 3) having at least one HbA1c $\geq 6.5\%$ (kappa = 0.71 and 0.74 for PWH and HIV-negative, respectively) (Table 3). However, all of the measures examined provided moderate agreement or better (kappa ≥ 0.58) with chart adjudication.

Based on chart adjudication, diabetes incidence among PWH was 12% for African-Americans and 8% for white participants; and among the HIV-negative, diabetes incidence was 19% among African-Americans and 17% among whites. The operating characteristics and kappa statistics were similar between African-American and white PWH and HIV-negative participants (Supplemental Table S2).

The time between chart adjudication and EMR incidence dates is lowest using the first HbA1c date with a mean of 26 days (SD=233). For comparison, using the second HbA1c date resulted in a mean difference of 365 days (SD=438) and using ICD-9 date resulted in a mean difference of 564 days (SD=749). Difference between incidence dates was lower for PWH compared to HIV-negative participants for all comparisons made (Table 4). Of note, 56% (53% of PWH and 64% HIV-negative) had an HbA1c value prior to the blood date and 89% (86% of PWH and 95% of HIV-negative) had a value ever.

Diabetes ICD-10 Codes and Agreement

Presence of diabetes was similar using ICD-10 (32%) codes and chart adjudication (31%). Agreement was substantial and similar for PWH and HIV-negative (kappa = 0.77 and 0.76), as shown in Table 3.

Discussion

We assessed agreement between EMR-based approaches for identifying presence of diabetes compared to two-physician manual chart adjudication among the 2,389 PWH and HIV-negative participants in the VACS-BC. For all definitions, diabetes prevalence and incidence is lower among veterans with HIV compared to those without, which may reflect individual reasons for seeking care in the VHA and is consistent with previous research.^{16,18}

In this study, the EMR diabetes prevalence definitions which agree best with chart review are: 1) HbA1c $\geq 6.5\%$ at least once, and 2) the composite definition. The EMR incident diabetes definitions with the best agreement with chart review, based on Kappa statistics are: 1) at least two HbA1c $\geq 6.5\%$, 2) the EMR composite criteria, and 3) having HbA1c $\geq 6.5\%$ at least once. However, the date of onset using HbA1c $\geq 6.5\%$ was closest to date of onset based on chart review, so this may be the best measure to use when date of incidence is important. Based on this finding, when multiple criteria are used to identify diabetes it is preferable to use the first date of any indication of diabetes. Operating characteristics and kappa statistics are similar for PWH and the HIV-negative, and for African-American and white subgroups. Agreement of ICD-10 codes with chart adjudicated diabetes was substantial.

While EMRs have the potential to produce high-quality research, this study highlighted some of the challenges of correctly identifying medical conditions including the lack of contextual information to interpret laboratory values, inadvertent entry of incorrect diagnostic codes,^{3,4} potential missing data,⁵ and provider variation in screening practices.²³ Our approach is designed to establish both prevalence and incidence which enables the identification of retrospective cohorts of patients with prevalent diabetes as well as analyses of relationships between factors at a given point in time (e.g., CD4+ T cell count at antiretroviral therapy initiation in PWH) with incident diabetes.

Prior studies have validated EMR diabetes definitions,^{24–27} but our study reflects the more recent addition of HbA1c criteria that was recommended by the American Diabetes Association in 2010,²⁸ which we found to be far more specific than other criteria, as observed in prior studies.²⁵ HbA1c $\geq 6.5\%$ has been used for VACS incidence diabetes in

recent years,¹⁸ and HbA1c requires fewer resources to extract from the EMR compared to the more complicated composite definition. We found at least one HbA1c $\geq 6.5\%$ to be 88% sensitive overall for adjudicated diabetes prevalence. Among those with only HbA1c value $\geq 6.5\%$, the addition of random glucose criteria only increased the proportion identified with diabetes from 20.0% (477) to 20.2% (482). Among the 351 participants who did not have an HbA1c value in the EMR, only five had a random glucose >200 mg/dL, 342 had glucose measurements but no values >200 mg/dL, and four had no recorded glucose values.

There are several limitations to this study. Participants are mostly men so results may not be generalizable to women. Results may not be generalizable to other EMR systems as the consistency of care for veterans in the VA ensures a greater opportunity to capture comorbidities compared to other systems. Because chart review was conducted only on medical records identified as potentially containing a diagnosis of diabetes, operating characteristics need to be interpreted with caution as it is possible we did not identify every occurrence of diabetes through our method of identifying potential cases. Because of the timeframe of this study, the main analysis included ICD-9 codes, but most health systems have since transitioned to using ICD-10 codes. The results of this study are relevant nonetheless because: 1) future studies that use historical data will still rely on ICD-9 codes; 2) the concept of comparing EMR ICD-9 and ICD-10 to chart adjudication is the same; and 3) our main findings found substantial agreement between HbA1c $\geq 6.5\%$ and chart adjudicated diabetes. Additionally, we compared presence of diabetes based on chart review up to October 2017 versus ICD-10 codes from October 2015 to 2017 and results are similar.

Notably, we found a HbA1c $\geq 6.5\%$ was less sensitive for adjudicated diabetes in PWH (85%) compared to the HIV-negative (92%), which supports prior studies indicating the relationship of HbA1c and blood glucose differs in PWH. In a study of PWH and HIV-negative controls without known hemoglobinopathies, HbA1c underestimated the mean fasting glycemia by 12.3% in PWH as compared to controls, and the factors associated with HbA1c inaccuracy included low serum haptoglobin suggestive of increased hemolysis, higher mean corpuscular volume (MCV), and use of the thymidine analogue antiretroviral medications.¹¹ In a similar study, HbA1c underestimated fasting glucose by an average of 29 mg/dl in the PWH compared to HIV-negative controls, and greater discordance was associated with higher MCV and current use of abacavir, lamivudine, and zidovudine antiretrovirals.¹² Finally, the largest study to date comparing HbA1c and fasting glucose in 1357 PWH and 1500 HIV-negative controls in the Multicenter AIDS Cohort Study (MACS) found the median HbA1c among PWH was 0.21% lower than among controls at a fasting glucose of 125 mg/dL, and the magnitude of this effect increased with fasting glucose >126 mg/dL. Among the PWH, greater discordance (observed – expected HbA1c) was associated with lower CD4+ T cell count, higher MCV, higher mean corpuscular hemoglobin, or use of antiretroviral regimens containing a protease inhibitor, non-nucleoside reverse transcriptase inhibitor, or zidovudine.¹³

In summary, the utilization of EMRs for health research offers the opportunity to assess far larger patient cohorts than could be routinely enrolled in a clinical study. Given the high prevalence of diabetes risk factors within the US population, the identification of diabetes within EMRs is critical for identifying cohorts of people with diabetes, evaluating health

outcomes, and adjusting as a comorbidity in analyses. We validated the use of EMR, including a single HbA1c $\geq 6.5\%$, to identify prevalent and incident diabetes in the VA EMR and this measure can likely be extended to other sources of EMR data and in various subgroups.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments/Role of Funding

This study was funded by a grant from the National Institute on Alcohol Abuse and Alcoholism by COMpAAAS/Veterans Aging Cohort Study (U24-AA020794, U01-AA020790, U01-AA020799, U10 AA013566), and a grant from the National Institute of Diabetes and Digestive and Kidney Diseases (R56-DK108352). The funders of this study had no role in study design, data collection, analysis, interpretation and presentation, or in the decision to submit the manuscript for publication. Views presented in the manuscript are those of the authors and do not reflect those of the Department of Veterans Affairs, or the United States Government. We thank Kendall Bryant, PhD for the scientific collaboration.

References

1. 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]
2. Harris SB, Glazier RH, Tompkins JW, et al. Investigating concordance in diabetes diagnosis between primary care charts (electronic medical records) and health administrative data: a retrospective cohort study. *BMC Health Serv Res* 2010;10:347. [PubMed: 21182790]
3. 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. [PubMed: 16849965]
4. Lloyd SS, Rissing JP. Physician and coding errors in patient records. *JAMA*. 1985;254(10):1330–1336. [PubMed: 3927014]
5. McGinnis KA, Skanderson M, Levin FL, Brandt C, Erdos J, Justice AC. Comparison of two VA laboratory data repositories indicates that missing data vary despite originating from the same source. *Med Care*. 2009;47(1):121–124. [PubMed: 19106740]
6. Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2020. Atlanta, GA: Centers for Disease Control and Prevention, U.S. Dept of Health and Human Services 2020 Available at: <https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf> [Accessed March 31, 2020].
7. American Diabetes Association. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2019. *Diabetes Care*. 2019;42(Suppl 1):S13–S28. [PubMed: 30559228]
8. Ziemer DC, Kolm P, Weintraub WS, et al. Glucose-independent, black-white differences in hemoglobin A1c levels: a cross-sectional analysis of 2 studies. *Ann Intern Med*. 2010;152(12):770–777. [PubMed: 20547905]
9. Kumar PR, Bhansali A, Ravikiran M, et al. Utility of glycated hemoglobin in diagnosing type 2 diabetes mellitus: a community-based study. *J Clin Endocrinol Metab*. 2010;95(6):2832–2835. [PubMed: 20371663]
10. Herman WH, Ma Y, Uwaifo G, et al. Differences in A1C by race and ethnicity among patients with impaired glucose tolerance in the Diabetes Prevention Program. *Diabetes Care*. 2007;30(10):2453–2457. [PubMed: 17536077]
11. Diop ME, Bastard JP, Meunier N, et al. Inappropriately low glycated hemoglobin values and hemolysis in HIV-infected patients. *AIDS Res Hum Retroviruses*. 2006;22(12):1242–1247. [PubMed: 17209766]

12. Kim PS, Woods C, Georgoff P, et al. A1C underestimates glycemia in HIV infection. *Diabetes Care*. 2009;32(9):1591–1593. [PubMed: 19502538]
13. Slama L, Palella FJ, Jr, Abraham AG, et al. Inaccuracy of haemoglobin A1c among HIV-infected men: effects of CD4 cell count, antiretroviral therapies and haematological parameters. *J Antimicrob Chemother*. 2014;69(12):3360–3367. [PubMed: 25096078]
14. Corrigan JM, Eden J, Smith BM, Eds. *Leadership by example: Coordinating government roles in improving healthcare quality committee on enhancing federal healthcare quality programs*. Washington, DC: National Academy Press 2002.
15. McQueen L, Mittman BS, Demakis JG. Overview of the Veterans Health Administration (VHA) Quality Enhancement Research Initiative (QUERI). *J Am Med Inform Assoc* 2004;11(5):339–343. [PubMed: 15187071]
16. Butt AA, McGinnis K, Rodriguez-Barradas MC, et al. HIV infection and the risk of diabetes mellitus. *AIDS*. 2009;23(10):1227–1234. [PubMed: 19444074]
17. Tien PC, Schneider MF, Cox C, et al. Association of HIV infection with incident diabetes mellitus: impact of using hemoglobin A1C as a criterion for diabetes. *J Acquir Immune Defic Syndr* 2012;61(3):334–340. [PubMed: 22878421]
18. Herrin M, Tate JP, Akgun KM, et al. Weight gain and incident diabetes among HIV-infected veterans initiating antiretroviral therapy compared with uninfected individuals. *J Acquir Immune Defic Syndr*. 2016;73(2):228–236. [PubMed: 27171741]
19. 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]
20. Armah KA, McGinnis K, Baker J, et al. HIV status, burden of comorbid disease, and biomarkers of inflammation, altered coagulation, and monocyte activation. *Clin Infect Dis*. 2012;55(1):126–136. [PubMed: 22534147]
21. Medapalli RK, Parikh CR, Gordon K, et al. Comorbid diabetes and the risk of progressive chronic kidney disease in HIV-infected adults: data from the Veterans Aging Cohort Study. *J Acquir Immune Defic Syndr*. 2012;60(4):393–399. [PubMed: 22592587]
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174. [PubMed: 843571]
23. Winterstein AG, Kubilis P, Bird S, Cooper-DeHoff RM, Nichols GA, Delaney JA. Misclassification in assessment of diabetogenic risk using electronic health records. *Pharmacoepidemiol Drug Saf*. 2014;23(8):875–881. [PubMed: 24923707]
24. Richesson RL, Rusincovitch SA, Wixted D, et al. A comparison of phenotype definitions for diabetes mellitus. *J Am Med Inform Assoc*. 2013;20(e2):e319–326. [PubMed: 24026307]
25. Spratt SE, Pereira K, Granger BB, et al. Assessing electronic health record phenotypes against gold-standard diagnostic criteria for diabetes mellitus. *J Am Med Inform Assoc*. 2017;24(e1):e121–e128. [PubMed: 27616701]
26. Nichols GA, Desai J, Elston Lafata J, et al. Construction of a multisite DataLink using electronic health records for the identification, surveillance, prevention, and management of diabetes mellitus: the SUPREME-DM project. *Prev Chronic Dis*. 2012;9:E110. [PubMed: 22677160]
27. Raebel MA, Schroeder EB, Goodrich G, et al. Validating type 1 and type 2 diabetes mellitus in the mini-sentinel distributed database using the surveillance, prevention, and management of diabetes mellitus (SUPREMEDM) datalink. 2016 Available at https://www.sentinelinitiative.org/sites/default/files/Methods/Mini-Sentinel_Methods_Validating-Diabetes-Mellitus_MSDD_Using-SUPREME-DM-DataLink.pdf [Accessed March 31, 2020].
28. American Diabetes Association. Standards of medical care in diabetes—2010. *Diabetes Care*. 2010;33 Suppl 1:S11–61. [PubMed: 20042772]

Key points:

- Diabetes is common among those with and without human immunodeficiency virus.
- Agreement of electronic medical record based diabetes diagnosis with physician chart review was high for identifying prevalent and incident diabetes.
- Agreement was similar for human immunodeficiency virus and race/ethnicity subgroups.
- International Classification of Diseases, Ninth Revision and International Classification of Diseases, Tenth Revision codes, hemoglobin A1c, and a composite definition all provided at least substantial agreement with chart review.
- Identification of those with diabetes nationally within the Veterans Health Administration can be used in future studies to evaluate treatments, health outcomes, and adjust for diabetes in epidemiologic studies.

Table 1.**Diabetes Mellitus Composite Definition****Diabetes mellitus present if 1 of the following criteria are met:**

1. Serum glucose >200 mg/dl on two separate occasions (any duration apart)
2. HbA1c ≥ 6.5% on two separate occasions (any duration apart)
3. ICD-9 codes (two outpatient OR one inpatient) and treatment with an oral hypoglycemic or insulin for >30 days
4. ICD-9 codes (two outpatient OR one inpatient) and glucose >126 mg/dl on two separate occasions
5. Glucose >200 mg/dl on one occasion and treatment with an outpatient oral hypoglycemic or insulin for >30 days
6. HbA1c ≥ 6.5% on one occasion and treatment with an outpatient oral hypoglycemic or insulin for >30 days

Abbreviations: HbA1c, hemoglobin A1c; ICD-9, International Classification of Diseases, Ninth Revision; mg/dl, milligrams per deciliter

Table 2.

Demographic Characteristics for the VACS Biomarker Cohort, by HIV Status

	All	PWH	HIV-negative	Standardized Differences
N	2,389	1,546	843	
Mean age (SD)	52 (8.9)	52 (8.2)	54 (9.3)	0.203
Male sex (%)	95	97	90	-0.289
Race/ethnicity (%)				
African-American, non-Hispanic	68	69	67	-0.026
White, non-Hispanic	20	19	21	
Hispanic	8	8	8	
Other/unknown	4	4	4	
HCV+ (%)	36	40	28	-0.271
Prevalent Diabetes (%)				
Chart adjudication	22	18	29	0.266
EMR ICD-9 Code	27	22	37	0.319
EMR HbA1c 6.5% at least once	21	16	28	0.276
EMR HbA1c 6.5% two or more times	17	13	24	0.299
EMR HbA1c 6.5% two or more times within a year	13	10	19	0.273
EMR composite definition *	22	18	29	0.268
'n' without EMR prevalent diabetes	1,871	1,272	599	
Incident Diabetes* (%)				
ICD-9 Code	19	15	28	0.321
HbA1c 6.5% at least once	15	12	23	0.288
HbA1c 6.5% two or more times	10	7	15	0.242
HbA1c 6.5% two or more times within a year	6	5	10	0.219
EMR composite definition	17	14	24	0.266
'n' without Chart Review prevalent diabetes	1,860	1,265	595	
Incident Diabetes* (%)				
Chart adjudication	13	11	19	0.238
Presence of Diabetes** (%)				
ICD-10 Code	32	26	41	0.308
Chart adjudication	31	24	41	0.338

* EMR composite definition described in Table 1

** Diabetes presence is based on chart review with a last date of October 2017 and ICD-10 codes from October 2015 to February 2018; limited to those alive as of October 2016 (n=1,118 PWH and 635 HIV-negative)

Abbreviations: VACS, Veterans Aging Cohort Study; HIV, human immunodeficiency virus; PWH, persons with HIV; HCV, Hepatitis C Virus; EMR, electronic medical record; ICD-9, International Classification of Diseases, Ninth Revision; HbA1c, hemoglobin A1c; ICD-10, International Classification of Diseases, Tenth Revision

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.
Diabetes from Electronic Medical Record Data Compared with Chart Adjudication in the VACS Biomarker Cohort, by HIV Status

		n	Sensitivity	Specificity	PPV	NPV	Kappa
Prevalence							
<i>HbA1c 6.5% at least once</i>	All	2,389	88 (85–91)	99 (98–99)	95 (93–97)	97 (96–98)	0.89
	PWH	1,546	85 (80–89)	99 (98–99)	94 (90–96)	97 (96–98)	0.87
	HIV-negative	843	92 (89–95)	99 (98–100)	97 (94–99)	97 (95–98)	0.92
<i>HbA1c 6.5% two or more times</i>	All	2,389	76 (72–79)	100 (100–100)	100 (98–100)	94 (92–95)	0.83
	PWH	1,546	70 (64–75)	100 (100–100)	100 (97–100)	94 (92–95)	0.79
	HIV-negative	843	82 (77–87)	100 (99–100)	100 (97–100)	93 (91–95)	0.86
<i>HbA1c 6.5% two or more times within a year</i>	All	2,389	59 (54–63)	100 (100–100)	99 (98–100)	90 (88–91)	0.69
	PWH	1,546	53 (47–59)	100 (100–100)	99 (96–100)	91 (90–92)	0.65
	HIV-negative	843	65 (59–71)	100 (99–100)	99 (97–100)	87 (85–90)	0.72
<i>EMR Composite Definition</i>	All	2,389	92 (89–94)	98 (98–99)	93 (91–95)	98 (97–98)	0.90
	PWH	1,546	91 (87–94)	99 (98–99)	93 (89–96)	98 (97–99)	0.90
	HIV-negative	843	92 (88–95)	98 (96–99)	94 (90–97)	97 (95–98)	0.90
<i>ICD-9 code only</i>	All	2,389	90 (87–93)	91 (89–92)	73 (70–77)	97 (96–98)	0.75
	PWH	1,546	88 (83–91)	92 (91–94)	72 (67–76)	97 (96–98)	0.74
	HIV-negative	843	93 (89–96)	87 (84–90)	75 (70–80)	97 (95–98)	0.75
Incidence							
<i>HbA1c 6.5% at least once</i>	All	1,824	85 (80–89)	95 (94–96)	70 (67–78)	98 (97–98)	0.73
	PWH	1,245	81 (73–88)	96 (94–97)	69 (61–76)	98 (97–99)	0.71
	HIV-negative	579	89 (82–94)	92 (89–94)	72 (64–79)	97 (95–99)	0.74
<i>HbA1c 6.5% two or more times</i>	All	1,868	72 (65–77)	99 (99–100)	93 (88–96)	96 (95–97)	0.79
	PWH	1,271	68 (59–76)	99 (99–100)	93 (85–97)	97 (95–98)	0.76
	HIV-negative	597	76 (67–84)	99 (97–100)	93 (86–98)	95 (93–97)	0.81
<i>HbA1c 6.5% two or more times within a year</i>	All	1,868	48 (41–54)	99 (99–100)	93 (86–97)	93 (92–94)	0.59

	n	Sensitivity	Specificity	PPV	NPV	Kappa
EMR Composite Definition	PWH	1,271	47 (38–56)	100 (99–100)	92 (83–97)	94 (93–96)
	HIV-negative	597	49 (39–58)	99 (98–100)	93 (83–98)	90 (87–92)
	All	1,871	95 (92–98)	94 (93–95)	70 (65–75)	99 (99–100)
	PWH	1,272	95 (89–98)	95 (94–96)	68 (61–75)	99 (99–100)
ICD-9 code only	HIV-negative	599	96(91–99)	92 (89–94)	72 (64–80)	99 (98–100)
	All	1,871	83 (78–88)	90 (89–91)	55 (49–60)	97 (96–98)
	PWH	1,272	81 (73–88)	92 (91–94)	54 (47–61)	98 (97–99)
	HIV-negative	599	85 (77–91)	85 (81–88)	55 (48–63)	96 (94–98)
Presence of Diabetes**						
ICD-10 code	All	1,753	84 (80–87)	93 (92–95)	85 (82–88)	93 (91–94)
	PWH	1,118	81 (76–86)	95 (93–96)	85 (80–89)	93 (92–95)
	HIV-negative	635	87 (82–90)	90 (87–93)	86 (81–90)	91 (87–93)
						0.76

* Prevalence and incidence are calculated at time of VACS blood draw date (2005 to 2007)

** Presence is based on chart review with a last date of October 2017 and ICD-10 codes from October 2015 to February 2018; limited to those alive as of October 2016 (n=1,118 PWH and 635 HIV-negative)

Abbreviations: VACS, Veterans Aging Cohort Study; HIV, human immunodeficiency virus; HbA1c, hemoglobin A1c; PWH, persons with human immunodeficiency virus; EMR, electronic medical record; ICD-9, International Classification of Diseases, Ninth Revision; ICD-10, International Classification of Diseases, Tenth Revision

Table 4.

Comparison of Chart Adjudication versus EMR-assigned Diabetes Incidence Dates

EMR Measure		Days Between Chart Review and EMR Incidence Date			
		N	Mean	SD	Median (IQR)
HbA1c 6.5% at least once	All	200	26	233	0 (0–0)
	PWH	103	23	230	0 (0–0)
	HIV-negative	97	29	237	0 (0–0)
HbA1c 6.5% two or more times	All	169	365	438	213 (80–521)
	PWH	86	300	329	179 (51–440)
	HIV-negative	83	434	520	228 (98–560)
HbA1c 6.5% two or more times within a year	All	112	152	225	111 (46–219)
	PWH	59	127	105	98 (38–224)
	HIV-negative	53	180	307	125 (74–213)
EMR Composite Definition	All	216	306	478	104 (7–437)
	PWH	115	273	467	74 (7–325)
	HIV-negative	101	344	489	160 (6–526)
ICD-9 code only	All	164	564	749	222 (28–879)
	PWH	83	514	653	207 (21–920)
	HIV-negative	82	614	835	254 (30–869)

Abbreviations: EMR, electronic medical record; HbA1c, hemoglobin A1c; PWH, persons with human immunodeficiency virus; ICD-9, International Classification of Diseases, Ninth Revision