BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (http://bmjopen.bmj.com).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Cohort Profile: The US Veterans Administration Diabetes Risk (VADR) National Cohort

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-039489
Article Type:	Cohort profile
Date Submitted by the Author:	07-May-2020
Complete List of Authors:	Avramovic, Sanja; George Mason University; VA New York Harbor Healthcare System Alemi, Farrokh ; George Mason University Kanchi, Rania; New York University School of Medicine Lopez, Priscilla; New York University School of Medicine Hayes, Richard; New York University School of Medicine Thorpe, Lorna; New York University School of Medicine, Department of Population Health Schwartz, Mark; New York University School of Medicine; VA New York Harbor Healthcare System
Keywords:	General diabetes < DIABETES & ENDOCRINOLOGY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, EPIDEMIOLOGY

SCHOLARONE™
Manuscripts

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

For peer review only

Cohort Profile: The US Veterans Administration Diabetes Risk (VADR) National Cohort

Sanja Avramovic,[1,3] Farrokh Alemi,[1] Rania Kanchi,[2] Priscilla Lopez,[2] Richard B Hayes,[2] Lorna E. Thorpe,[2] Mark D Schwartz[2,3]

[1] Department of Health Administration and Policy, George Mason University, Fairfax, VA
[2] Department of Population Health, NYU School of Medicine, New York, NY
[3] VA New York Harbor Healthcare System, New York, NY

Correspondence to
Dr. Sanja Avramovic; savramov@gmu.edu

Abstract

Purpose The VADR cohort facilitates studies on temporal and geographic patterns of prediabetes and diabetes, as well as targeted studies of their predictors. The cohort provides an infrastructure for examination of novel individual and community-level risk factors for diabetes and their consequences among veterans. This cohort also establishes a baseline against which to assess the impact of national or regional strategies to prevent diabetes in veterans.

Participants The VA Diabetes Risk Cohort (VADR) includes all 6,082,246 veterans in the United States who were diabetes-free as of January 1, 2008, or who subsequently enrolled in the VA for primary care and were diabetes-free at cohort entry through December 31, 2016, and who had at least 2 diabetes-free visits to a VA primary care service at least 30 days apart within any 5-year period since January 1, 2003.

Findings to date The incidence rate of type 2 diabetes in this cohort of over 6 million veterans followed for a median of 5.5 years (over 35 million person-years) was 26 per 1000 person-years. During the study period, 8.5% of the cohort were lost to follow-up and 17.7% died. Many demographic, comorbidity, and other clinical variables were more prevalent among patients with incident diabetes.

Future Plans This cohort will be used to study community-level risk factors for diabetes, such as attributes of the food environment and neighborhood socioeconomic status via geospatial linkage to residence address information.

Strengths and limitations of this study

- A strength of this national cohort is that it has a large size, a high degree of long-term follow-up, and a comprehensive set of variables.
- The VA healthcare system is the nation's largest integrated healthcare system, in which veterans are followed across all VA facilities and in-system providers.
- Documented data is restricted to that which is collected in EHRs during the course of clinical practice, leading to the possibility of confounding, selection bias and measurement error.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

- The veteran population is predominantly male and white, and so the findings may not generalize to minorities or to women.

Introduction

Diabetes mellitus (diabetes) is a chronic disease that affects 34.2 million (10.5%) of adults and children in the United States (US).[1] As of 2018, diabetes was the seventh leading cause of death and one of the major contributors to heart disease and stroke.[2] Adjusting for age and gender, all-cause mortality is 1.5 times greater for people with diabetes than for people without diabetes, and average health care costs are 2.3 times higher.[3] Another 88 million American adults (34.5%) are estimated to have prediabetes and at risk of developing diabetes.[4]

The Veterans Administration (VA) cares for more than 8 million US veterans, of whom approximately 25% have diabetes.[5,6] The annual mortality rate among veterans with diabetes is 5%—nearly double that of veterans without diabetes.[7,8] It is likely that nearly 3 million other veterans have prediabetes. These high rates compared to the general population may be due to the increased proportions of overweight (37%) and obesity (41%) among veterans,[9] their older age, lower socioeconomic status,[10] and possible exposure to herbicides such as Agent Orange.[11]

Behavioral prevention interventions can reduce the incidence of diabetes by 50–70%,[12,13] but scaling this up for population impact has been challenging due to the intensity and cost of the intervention and challenges of enrolling patients for such programs.[14-17]

In response to these challenges, we developed the Veterans Administration Diabetes Risk (VADR) Cohort, a national cohort of all US veterans enrolled at the Veterans Health Administration (VHA) since January 1, 2008 who were diabetes-free at enrollment. The cohort was developed as a part of the Diabetes Location, Environmental Attributes, and Disparities (LEAD) network; a Center for Disease Control and Prevention (CDC) funded research collaboration among Drexel University, Geisinger-Johns Hopkins, New York University School of Medicine, and University of Alabama at Birmingham with the CDC as a collaborative scientific partner in the network.[18] The VADR cohort facilitates studies on temporal and geographic patterns of prediabetes and diabetes, as well as targeted studies of their predictors. The cohort provides an infrastructure for examination of novel individual and community-level risk factors for diabetes and their consequences among veterans. This cohort also establishes a baseline against which to assess the impact of national or regional strategies to prevent diabetes in veterans.

Cohort description

VADR is the largest national cohort of diabetes-free adults in the US. Established in 2017 as a dynamic cohort enabled by the VA national electronic health record (EHR), the cohort includes diabetes-free US veterans enrolled in primary care clinics at any VA facility as early as January 1, 2008 through December 31, 2016, and followed from cohort entry through December 31, 2018. VA primary care clinics operate in 170 VA Medical Centers (VAMCs) and in more than 1,000 Community-Based Outpatient Clinics (CBOCs) across the US.[19] As a dynamic cohort, subject entry and follow-up is ongoing, but this paper reports on the cohort from 2008 through 2018.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Building on published, validated criteria in EHRs,[7,20] we defined diabetes using the following query-based definition comprised of any of three criteria: (1) at least two encounters (inpatient or outpatient) with documentation of a Type 2 diabetes ICD-9/10 code (ICD-9: 250.x; ICD-10: E11.x) or (2) a documented prescription for a diabetes medication other than metformin or acarbose alone; or (3) at least one encounter with a diabetes ICD-9/10 code and two elevated (≥ 6.5%) glycosylated hemoglobin (Hgb A1C) lab test results (see Appendix for complete definition).[21] We excluded metformin or acarbose alone from the criteria because these drugs may be used for diabetes prevention in patients with prediabetes; including them may lead to misclassifying cases of prediabetes as diabetes.[22,23] This definition was used to exclude prevalent diabetes cases prior to cohort entry and to estimate diabetes incidence during the study period.

For etiologic analyses, subjects were eligible for the cohort if they were veterans with at least 2 diabetes-free visits to a VA primary care service, occurring at least 30 days apart, from January 1st 2003 to December 31st 2018. Cohort entry (baseline) was defined as either January 1, 2008 or the date of the second diabetes-free primary care visit for subjects entering after January 1, 2008. Eligible subjects were allowed to enter the cohort through December 31, 2016 to allow at least 2 years of follow-up during which subjects may be diagnosed with diabetes. Subjects were censored when they developed diabetes, died, or were lost to follow-up (defined as having no encounters in the VA health system for more than 2 years). Once a patient was lost-to-follow-up, they were not eligible to re-enter the cohort. Encounters for follow-up included any visits to primary care, specialists, emergency departments, walk-in clinics, hospitalizations, or nursing home stays at any VA facility. Person-years (PY) of follow-up for each subject were calculated as the interval between cohort entry date and censor date.

As shown in **Figure 1**, the cohort was developed from a base total population of 8,346,180 patients seen for at least 1 primary care visit between 1999, the earliest year for which EHR data were available on patients, and the start of the study period. The cohort was then restricted to patients seen in the five years prior to the study period start date, January 1, 2008. Patients were excluded if they had fewer than 2 primary care visits, at least 30 days apart during that five year time period and less than 2 primary care visits after cohort entry. After excluding patients with prevalent diabetes, the initial diabetes-free cohort included 2,968,855 patients. Another 3,113,391 diabetes-free patients met the same eligibility criteria after the start of the study period and entered the cohort between January 1, 2008 and December 31, 2016, resulting in a diabetes-free cohort of 6,082,246 patients.

Information on subjects in the cohort was updated daily as it was drawn from EHR at all VA facilities into the VA corporate data warehouse (CDW), based on all clinical services provided and documented by the VA to subjects over time. All data in the cohort were obtained through the VA Informatics and Computing Infrastructure (VINCI), a secure, high performance interface with VA's national CDW, available through VA's Information Resource Center (VIReC).[24] The CDW contains data integrated from VA's electronic medical record (VISTA, Veterans Health Information Systems and Technology Architecture), including all administrative data (e.g. all dates of encounters and diagnostic codes for outpatient and inpatient care), patient demographic characteristics, clinical data (e.g. vital signs, health factors, pharmacy, laboratory,

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

radiological, clinical notes, etc.), and healthcare utilization factors as they accrue over time, as the CDW is refreshed daily.[25]

The main outcome variable was a new diagnosis of Type 2 diabetes, measured using the definition described earlier.

Predictor variables and covariates

All continuous variables with repeated measures, including anthropomorphic, vital signs, and laboratory values, were defined as the average of the two most recent measures, prior to or at the time of cohort entry. If only one measure was taken prior to cohort entry, that was used as the baseline measure. The rate of missing data for all variables was measured.

Demographic measures were captured at baseline, including age, gender, marital status, race/ethnicity. First address on file per patient in cohort were exported out of the VINCI environment, geocoded using ArcGIS[26] and Python[27], and mapped to show number of patients in the cohort per census tract using QGIS.[28]

Glycemia and body weight are important predictors of diabetes. We measured Hgb A1c as a continuous value, and classified as normal (<5.7%), prediabetes (5.7% to 6.4%), or diabetes (≥6.5%). We measured weight in pounds and body mass mass index (BMI), defined as (weight in kilograms) / (height in meters)^2. BMI was also classified as underweight (<18.5); normal (18.5 to <25); overweight (25.0 to <30); and obese (≥30.0).[29]

Common comorbidities measured at baseline included established risk factors for diabetes such as obesity, hypertension, gestational diabetes, cardiovascular disease, chronic kidney disease hyperuricemia, fatty liver disease, polycystic ovary syndrome, and hepatitis C. These and all other comorbidities were defined as having at least 1 ICD code in the EHR prior to entering the cohort. Hyperlipidemia was defined as at least 2 encounters with ICD codes for hyperlipidemia, total cholesterol >240 mg/dL, or the use of lipid lowering medications.[30] Hypertension was defined as at least ICD code for hypertension or at least 2 consecutive elevated BP within the last two years prior to cohort entry.[31,32] Elevated BP was included as as ≥130/80 and ≥140/90, respectively, to comply with changes in hypertension guidelines over the course of the study period.[33,34]

Other clinical variables potentially related to diabetes incidence included: Blood Pressure (BP, excluding those measured in the hospital, emergency department, or at night); Lipids (Total Cholesterol, High Density Lipoprotein, Low Density Lipoprotein, and Triglycerides; Hepatic Transaminase Enzymes (serum aspartate aminotransferase - AST or SGOT - and alanine aminotransferase - ALT or SGPT); Renal Function (measured as Estimated Glomerular Filtration Rate - eGFR); Smoking Status: (Obtained from Health Factor files within CDW at cohort entry, classified as current, ever, or never smokers); and Agent Orange Exposure (Obtained from the number of veterans with Agent Orange listed as a health factor in the medical record).[11] Beside this select list, all documented diagnoses and treatments are available for the cohort.

Findings

The total person-years (PY) for this national cohort with 6,082,246 veterans from all 50 states was 35,889,982 (median 5.5 PY, IQR: 2.6 - 9.8). As shown in **Table 1**, the mean age of the cohort

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

was 58 years at baseline, 36.4% were 65 or older, most were male (91.7%), more than two-thirds were non-Hispanic white (75.1%), 16.2% were non-Hispanic black, and 6.1% were Hispanic. The majority (55.2%) were married or living with a partner.

At baseline, the average Hgb A1C was 5.8% among the 40.7% of the cohort tested at entry, and of these, 41.5% had an Hgb A1C in the prediabetes range. The average weight was 196.9 pounds and average BMI was 28.8 (SD 5.4). At baseline, 40.6% were overweight and 36.1% were obese. Traditional clinical risk factors for diabetes were common in this cohort as 49.5% had hypertension, 44.1% had hyperlipidemia, and 42.6% were smokers. Other clinical risk factors for diabetes included ischemic heart disease (16.4%), peripheral vascular disease (4.2%), heart failure (3.0%), and chronic kidney disease (2.5%). Most of these risk factors were present at baseline at higher rates among those who developed diabetes compared with those who did not during cohort follow-up.

Figure 2 shows the number of subjects in the cohort over time, from inception January 1, 2008 through December 31, 2018. Almost half (48.8%) of the cohort entered at cohort inception in January 1, 2008, with the remainder entering during the study period through December 31, 2016. During cohort follow-up, 936,627 (15.4%) veterans developed diabetes, for an incidence rate of 26 per 1,000 PY. Additionally, 518,503 (8.5%) were lost to follow-up, and 1,077,662 (17.7%) died during the study period. **Figure 3** shows the geographic distribution of the number of patients per tract. The majority of addresses were able to be geocoded (89%); of those not geocoded, about half were PO boxes, and the other half were missing. The majority of census tracts had between 20-80 patients.

Because cohort data were drawn from the VA EHR, which depends on documentation of services provided, some subjects had missing values for some variables at baseline. For example, the percentage of missing variables at cohort entry were: gender (<0.01%); race/ethnicity (10.1%); marital status (7.5%); BMI (4.3%); and Hgb A1C (59.3%). The missing race/ethnicity variable in VA data is widely known.[35] Screening for diabetes with Hgb A1c became more common after the recommendation was published in 2009.[36]

Strengths and limitations

A primary strength of this national cohort is its large size and long-term follow-up. The cohort includes a comprehensive set of demographic, anthropomorphic, clinical, treatment, and other administrative variables, drawn from all inpatient and outpatient encounters, each of which are automatically updated over time. In addition to the select comorbidities identified in this paper, the cohort includes data related to all comorbidities. Future work will include calculation of a multi-morbidity index to measure the impact of medical history on emergence of diabetes.[37]

As the nation's largest, integrated healthcare system, the VA follows veterans across all VA facilities, even after moving and changing VA facilities or providers within the system. Additionally, data on veterans who are Medicare or Medicaid beneficiaries and seek health care outside of the VA will be included by merging the study cohort with data from the Centers for Medicare and Medicaid (CMS). Finally, home addresses are available and were geocoded in

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

order to study the effect of community level characterstics and the impact of moving over time on incident diabetes in future work using this cohort.

The cohort has a few limitations. It relies on data documented during the course of clinical practice in EHRs and thus causal inferences face difficulties associated with unmeasured confounding, selection bias, and measurement error. Selection biases may arise as lower health care utilizers are more likely to be lost to follow-up or excluded, and higher utilizers may be more likely to meet criteria for key exposure and outcome variables. This is partially mitigated by the several-year, longitudinal follow-up.

The veteran population is predominantly male and white, and so the findings may not generalize to minorities or to women. Nonetheless, our large cohort ensures a sufficient and growing sample of women veterans (504,020) and patients from major ethnic/racial groups (886,150 NH black veterans, 331,376 Hispanic veterans), providing the ability to study diabetes incidence among these subgroups and improving the generalizability of our findings to non-veteran populations.

For peer review only

Patient and Public Involvement

This cohort study was conducted without engagement or co-production by patients or the public.

Contributorship statement

Authors contributed equally to this work.

Competing interests

The authors report no competing interests.

Funding

This study was funded by the Centers for Disease Control and Prevention (5 U01DP006299-02-00; PI: LET).

Data sharing statement

To gain access to data from the diabetes cohort described in this profile, interested researchers can contact the corresponding author. Access to VA electronic health records is limited to researchers with active, VA appointments and have an IRB-approved protocol. The process for obtaining a VA appointment without compensation (WOC) can be lengthy and varies by region, depending on the VA office processing the application. Once a researcher has a VA appointment and has IRB approval, the VA has developed a comprehensive data infrastructure to support secure and remote access to data via the VINCI platform. Additionally, deidentified datasets can be established and shared with appropriate IRB approval and data use agreements.

Further details

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Further details regarding the ability to access VA data can be found on the VA website dedicated to researchers: https://www.hsrd.research.va.gov/for_researchers/default.cfm,[38] including links to policies and guidance documents, special interest groups, funding opportunities, and a link to the VA Informatics and Computing Infrastructure (VINCI) site where access to actual data is granted once appropriate applications have been submitted and approved: https://www.hsrd.research.va.gov/for_researchers/vinci/.[24]

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Profile in a Nutshell

- The VA Diabetes Risk Cohort (VADR) includes all 6,082,246 veterans in the United States who were diabetes-free as of January 1, 2008, or who subsequently enrolled in the VA for primary care and were diabetes-free at cohort entry through December 31, 2016. Follow up is ongoing and is presented here through December 31, 2018.

- This is an ongoing, dynamic cohort enabled by the VA national electronic health record network, with passive data collection as it relies on routine medical information obtained from all inpatient and outpatient clinical encounters, updated daily.

- Subjects eligible for the cohort, either at baseline or afterwards, include all veterans who had at least 2 diabetes-free visits to a VA primary care service at least 30 days apart within any 5-year period since January 1, 2003. VA primary care clinics operate in 151 VA Medical Centers and more than 800 Community-Based Outpatient Clinics across the United States.

- The main outcome variable is incidence of type 2 diabetes, using a query-based definition comprised of at least two encounters with documentation of a diabetes ICD-9/10 code, or a documented prescription for a diabetes medication other than Metformin or Acarbose only; or at least one encounter with a diabetes ICD-9/10 code documented and two elevated (≥ 6.5%) glycosylated hemoglobin during the study period.

- The incidence rate of type 2 diabetes in this cohort of over 6 million veterans followed for a median of 5.5 years (over 35 million person-years) was 26 per 1000 person-years.

- During the study period, 8.5% of the cohort were lost to follow-up and 17.7% died.

- Many demographic, comorbidity, and other clinical variables were more prevalent among patients with incident diabetes.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 1. Cohort demographics and clinical characteristics by incident diabetes status

	All veterans^ n (%) or mean (SD)	Veterans without incidence diabetes ǂ n (%) or mean (SD)	Veterans with incident diabetes n (%) or mean (SD)
Total	6,082,246	5,145,619	936,627
Demographic characteristics			
Age	58 (17)	53 (16.1)	61 (12.2)
Age categories			
18-34	745,521 (12.3)	592,263 (16.7)	21788 (2.3)
35-49	1,009,704 (16.6)	725,443 (20.4)	130347 (13.9)
50-64	2,114,320 (34.8)	1,280,156 (36.1)	451975 (48.3)
65-79	1,499,835 (24.7)	786,222 (22.2)	261601 (27.9)
80+	712,733 (11.7)	165,266 (4.7)	70904 (7.6)
Gender			
Male	5,578,056 (91.7)	3,167,546 (89.2)	886789 (94.7)
Female	504,020 (8.3)	381,782 (10.8)	49,818 (5.3)
Race ethnicity			
Non-Hispanic white	4107,390 (75.1)	2,421,016 (74.2)	617020 (71.1)
Non-Hispanic black	886,150 (16.2)	542,362 (16.6)	172580 (19.9)
Hispanic	331,376 (6.1)	211,268 (6.5)	55245 (6.4)
Non-Hispanic Asian	55,209 (1)	37,208 (1.1)	7732 (0.9)
Non-Hispanic other	86,270 (1.6)	52,698 (1.6)	15218 (1.8)
Marital status			
Married or living with a partner	3,104,735 (55.2)	1,833,115 (55.2)	477624 (56.2)
Single	2,523,397 (44.8)	1,488,011 (44.8)	372,313 (43.8)
Clinical characteristics			
HbA1c	5.8 (0.4)	5.6 (0.4)	5.9 (0.3)
Normal (<5.7%)	1,311,814 (53)	875,704 (62.3)	128685 (25.8)
Prediabetes (5.7%-6.49%)	1,027,373 (41.5)	514,771 (36.6)	268,113 (53.8)
Diabetes range (≥6.5%)	134,321 (5.4)	15,304 (1.1)	101,749 (20.4)
Weight in pounds	196.9 (40.7)	196.6 (38.9)	214.8 (45.1)
BMI	28.8 (5.4)	28.7 (5.1)	31.3 (6.0)
Underweight (<18.5)	48,956 (0.8)	18,043 (0.5)	4,417 (0.5)
Normal weight (18.5-<25)	1,308,490 (22.5)	732,446 (21.6)	105,231 (11.7)

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Overweight (25-<30)	2,362,954 (40.6)	1,441,766 (42.5)	294,237 (32.6)
Obese (≥30)	2,101,515 (36.1)	1,197,744 (35.3)	499,091 (55.3)
Measured blood pressure*			
Systolic blood pressure	130 (14.7)	129 (14.2)	133 (14.9)
Diastolic blood pressure	76 (10.0)	77 (9.6)	78 (10.1)
Elevated blood pressure (≥130/80)	3,516,683 (60.7)	2,011,677 (59.4)	609,089 (67.9)
Elevated blood pressure (≥140/90)	1,499,531 (25.9)	796,643 (23.5)	287,673 (32.1)
Hypertension			
≥1ICD code or 2 consecutive elevated BP* (≥130/80)	3,774,345 (62.1)	1,966,277 (38.2)	721,080 (77.0)
≥1ICD code or 2 consecutive elevated BP* (≥140/90)	3,153,815 (51.9)	1,547,465 (30.1)	646,090 (69.0)
Lipids			
Total cholesterol	185.5 (38.4)	188.3 (37.7)	185.4 (40.2)
Triglyceride	140.8 (88.9)	137.6 (86.9)	168.1 (103.2)
LDL	112.6 (33.1)	115.3 (32.8)	111.4 (34.1)
HDL	46.3 (14.2)	47.3 (14.3)	42.4 (12.5)
Hyperlipidemia¥	2,681,776 (44.1)	1,371,540 (38.6)	528,320 (56.4)
Smoking status			
Current smoker	948,387 (42.6)	562,038 (41.9)	140,356 (44.3)
Not a smoker	1,280,059 (57.4)	780,210 (58.1)	176,812 (55.7)
Estimated glomerular filtration rate (eGFR)	80.1 (18.3)	82.5 (16.7)	78.6 (18.4)
eGFR≥90 (stage 1)	1,036,931 (30.8)	691,985 (33.9)	145,396 (27.9)
eGFR ≥60 to <90 (stage 2)	1,899,356 (56.5)	1,180,801 (57.8)	300,189 (57.6)
eGFR<60 (stage 3, 4, or 5)	427,148 (12.7)	169,862 (8.3)	75,169 (14.4)
Chronic kidney disease (ICD codes)	150,829 (2.5)	47,355 (0.9)	29,237 (3.1)
Ischemic heart disease (ICD codes)	999,988 (16.4)	378,843 (7.4)	209,455 (22.4)
Heart failure (ICD codes)	181,388 (3.0)	40,412 (0.8)	41,271 (4.4)
Peripheral vascular disease (ICD codes)	256,074 (4.2)	77,343 (1.5)	55,244 (5.9)
Stroke (ICD codes)	30,424 (0.5)	10,724 (0.2)	6,275 (0.7)

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Agent orange exposure	26,419 (0.4)	14,040 (0.3)	7,258 (0.8)
Chronic hepatitis C	102,535 (1.7)	51,790 (1.0)	21,382 (2.3)
Hyperuricemia	180,946 (3.0)	74,544 (1.4)	54,981 (5.9)
Polycystic ovary syndrome	4,994 (.0)	3,765 (1.0)	801 (1.6)
Gestational diabetes	157 (0.03)	131 (0.03)	22 (0.04)
Liver enzymes			
Abnormal AST	25.8 (14.3)	25.7 (14.2)	26.7 (14.7)
Abnormal ALT	29.5 (19.8)	30.3 (19.9)	32.4 (21.1)
Elevated AST	277,607 (6.4)	156,503 (6.2)	56,466 (8.3)
Elevated ALT	778,953 (17.5)	482,357 (18.6)	158,752 (22.8)
Fatty liver disease	253,139 (10.4)	136,850 (10.8)	67,367 (15.7)

* Only recent (within 2 years of cohort entry) BP measurements were used. Nighttime BP (8 PM to 7 AM) and BP measured in ER were excluded

¥ Hyperlipidemia was defined as: at least 2 ICD-9/10 codes for hyperlipidemia, total cholesterol > 240 mg/dL, or lipid lowering medication use

^ Including those lost to follow up and those died during the study period

ł Only patients who completed the follow up and were diabetes free at the end of the study

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

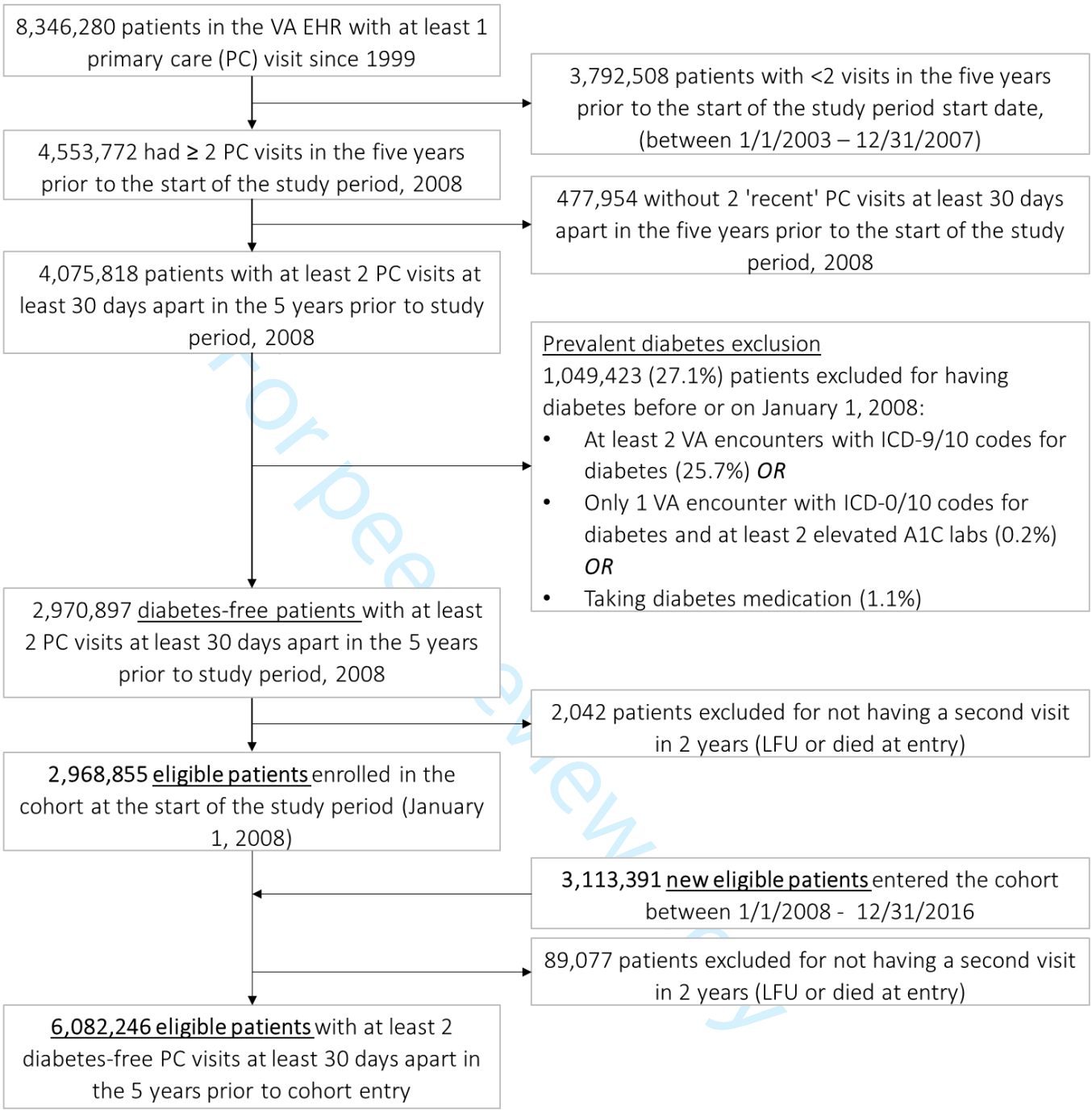
References

1. Centers for Disease Control and Prevention. *Estimates of Diabetes and its Burden in the United States.* Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2017;2017.
2. Centers for Disease Control and Prevention. *Underlying Cause of Death 1999-2008.* National Center for Health Statistics;2018.
3. American Diabetes Association. Economic costs of diabetes in the U.S. in 2012. *Diabetes care.* 2013;36(4):1033-1046.
4. Prevention CfDCa. Diabetes Quick Facts. 2019; https://www.cdc.gov/diabetes/basics/quick-facts.html. Accessed February 27, 2020.
5. Veterans Health Administration. VA Research on Diabetes. 2019.
6. Liu Y, Sayam S, Shao X, et al. Prevalence of and Trends in Diabetes Among Veterans, United States, 2005-2014. *Preventing chronic disease.* 2017;14:E135.
7. Miller D, Safford M, Pogach L. Who has diabetes? Best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data. *Diabetes care.* 2004;27 Suppl 2:B10-21.
8. Veterans Health Administration. Close to 25 Percent of VA Patients Have Diabetes. 2011; http://www.va.gov/health/NewsFeatures/20111115a.asp Accessed December 31, 2015.
9. Breland J, Phibbs C, Hoggatt K, et al. The Obesity Epidemic in the Veterans Health Administration: Prevalence Among Key Populations of Women and Men Veterans. *Journal of general internal medicine.* 2017;32(Suppl 1):11-17.
10. Kondo K, Low A, Everson T, et al. Health Disparities in Veterans: A Map of the Evidence. *Medical care.* 2017;55 Suppl 9 Suppl 2:S9-s15.
11. National Academies of Sciences E, and Medicine. *Veterans and Agent Orange.* National Academies Press;2018.
12. Prevention CfDCa. *National Diabetes Prevention Program.* 2019.
13. Group DPPR, Knowler W, Fowler S, et al. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. *Lancet (London, England).* 2009;374(9702):1677-1686.
14. Ackermann R, Finch E, Brizendine E, Zhou H, Marrero D. Translating the Diabetes Prevention Program into the community. The DEPLOY Pilot Study. *American journal of preventive medicine.* 2008;35(4):357-363.
15. Ackermann R, Marrero D. Adapting the Diabetes Prevention Program lifestyle intervention for delivery in the community: the YMCA model. *The Diabetes educator.* 2007;33(1):69, 74-65, 77-68.
16. Baker M, Simpson K, Lloyd B, Bauman A, Singh M. Behavioral strategies in diabetes prevention programs: a systematic review of randomized controlled trials. *Diabetes research and clinical practice.* 2011;91(1):1-12.
17. Schellenberg E, Dryden D, Vandermeer B, Ha C, Korownyk C. Lifestyle interventions for patients with and at risk for type 2 diabetes: a systematic review and meta-analysis. *Annals of internal medicine.* 2013;159(8):543-551.
18. Diabetes LEAD (Location, Environmental Attributes, and Disparities) Network. http://diabetesleadnetwork.org/. Accessed February 27, 2020.
19. Administration VH. About VHA. https://www.va.gov/health/aboutvha.asp. Accessed February 27, 2020.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

20. Bowe B, Xie Y, Li T, Yan Y, Xian H, Al-Aly Z. The 2016 global and national burden of diabetes mellitus attributable to PM2.5 air pollution. *The Lancet Planetary health.* 2018;2(7):e301-e312.
21. Davis PJ, Liu M, Sherman S, et al. HbA1c, lipid profiles and risk of incident type 2 Diabetes in United States Veterans. *PloS one.* 2018;13(9):e0203484.
22. Tseng E, Yeh HC, Maruthur NM. Metformin Use in Prediabetes Among U.S. Adults, 2005-2012. *Diabetes care.* 2017;40(7):887-893.
23. Hu R, Li Y, Lv Q, Wu T, Tong N. Acarbose Monotherapy and Type 2 Diabetes Prevention in Eastern and Western Prediabetes: An Ethnicity-specific Meta-analysis. *Clinical therapeutics.* 2015;37(8):1798-1812.
24. Affairs DoV. VA Informatics and Computing Infrastructure (VINCI). 2015; http://www.hsrd.research.va.gov/for_researchers/vinci/. Accessed February 27, 2020.
25. Casey J, Schwartz B, Stewart W, Adler N. Using Electronic Health Records for Population Health Research: A Review of Methods and Applications. *Annual review of public health.* 2016;37:61-81.
26. ArcGIS: The mapping and analytics platform. https://www.esri.com/en-us/arcgis/about-arcgis/overview.
27. Python. https://www.python.org/.
28. QGIS.
29. NHLBI N. Classification of Overweight and Obesity by BMI, Waist Circumference, and Associated Disease Risks. https://www.nhlbi.nih.gov/health/educational/lose_wt/BMI/bmi_dis.htm. Accessed February 27, 2020.
30. Soohoo M, Moradi H, Obi Y, Kovesdy CP, Kalantar-Zadeh K, Streja E. Serum triglycerides and mortality risk across stages of chronic kidney disease in 2 million U.S. veterans. *Journal of clinical lipidology.* 2019;13(5):744-753.e715.
31. Burg MM, Brandt C, Buta E, et al. Risk for Incident Hypertension Associated With Posttraumatic Stress Disorder in Military Veterans and the Effect of Posttraumatic Stress Disorder Treatment. *Psychosomatic medicine.* 2017;79(2):181-188.
32. Peng M, Chen G, Kaplan GG, et al. Methods of defining hypertension in electronic medical records: validation against national survey data. *Journal of public health (Oxford, England).* 2016;38(3):e392-e399.
33. Rowan CG, Turner JR, Shah A, Spaeder JA. Antihypertensive treatment and blood pressure control relative to hypertension treatment guidelines. *Pharmacoepidemiology and drug safety.* 2014;23(12):1294-1302.
34. Armstrong C. JNC8 guidelines for the management of hypertension in adults. *American family physician.* 2014;90(7):503-504.
35. Hernandez S, Sylling P, Mor M, et al. Developing an Algorithm for Combining Race and Ethnicity Data Sources in the Veterans Health Administration. *Military medicine.* 2019.
36. Gillett M. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes care.* 2009;32(7):1327-1334.
37. Alemi F, Levy C, Kheirbek R. The Multimorbidity Index: A Tool for Assessing the Prognosis of Patients from Their History of Illness. *EGEMS (Wash DC).* 2016;4(1):1235-1235.
38. Affairs DoV. Health Services Research & Development. https://www.hsrd.research.va.gov/for_researchers/default.cfm. Accessed February 27, 2020.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml



* Patients with unreliable information on date of birth and date of death and patients with year of birth <1900 were excluded (n=2,248)

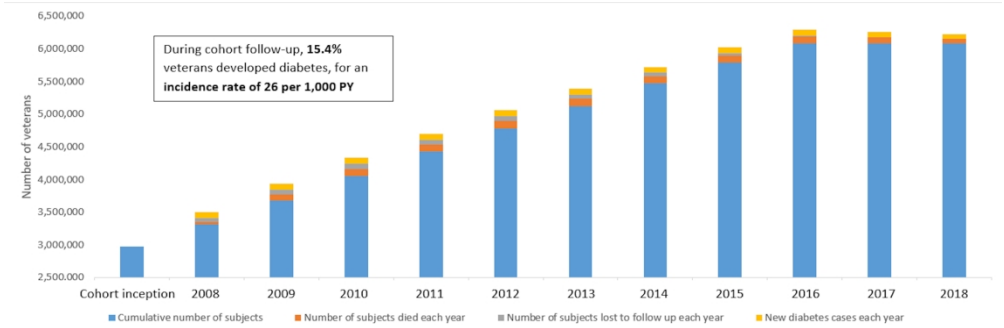


Figure 2. Cohort Trends, with cumulative numbers and percentage of patients, 2008 through 2018

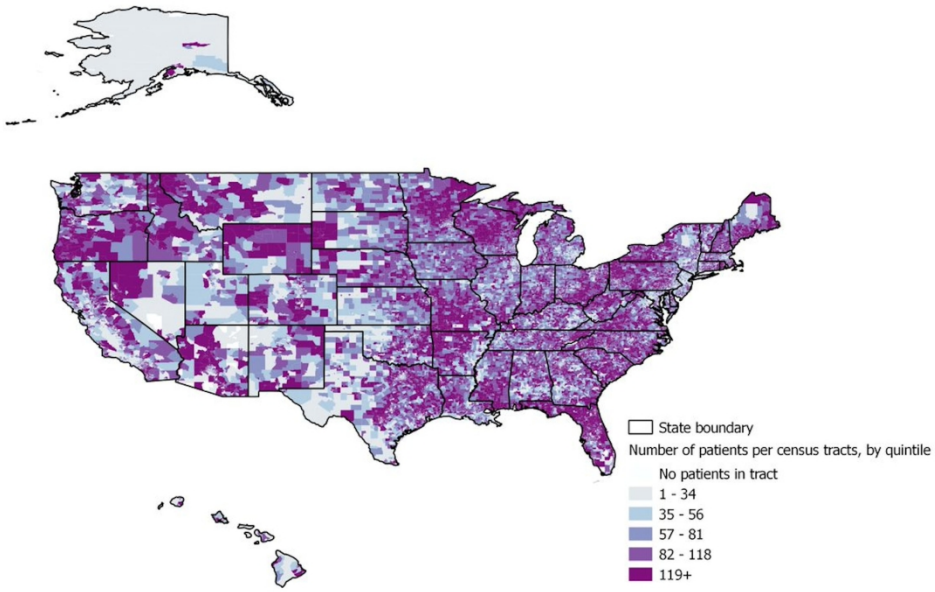


Figure 3. Geographic distribution of VADR Cohort

179x127mm (300 x 300 DPI)

Appendix. Defining chronic conditions in the VA environment

		Details				
Component	**Definition**	**ICD codes**		**Labs**	**Medications***	**Other**
		ICD-9	ICD-10			
Type 2 Diabetes Mellitus[1-4]	(1) At least two encounters (inpatient or outpatient) with documentation of a Type 2 diabetes ICD-9/10 code (ICD-9: 250.x; ICD-10: E11.x), *or* (2) a documented prescription for a diabetes medication other than metformin or acarbose alone, *or* (3) at least one encounter with a diabetes ICD-9/10 code and two elevated glycosylated hemoglobin (Hgb A1C) lab test results • Implausible A1C labs removed (range based on NHANES reporting) • Multiple labs measured on the same day or on the same day and time were averaged if they ranged within 1% • A1C labs ranging >1% were removed • If only one A1C lab was available prior to cohort entry date, that lab was used. • If more than one A1C lab was available, the average of the last two was taken	250, 250.0, 250.00, 250.02, 250.1, 250.10, 250.12, 250.2, 250.20, 250.22, 250.3, 250.30, 250.32, 250.4, 250.40, 250.42, 250.5, 250.50, 250.52, 250.6, 250.60, 250.62, 250.7, 250.70, 250.72, 250.8, 250.80, 250.82, 250.9, 250.90, 250.92	E11.00, E11.01, E11.21, E11.22, E11.29, E11.311, E11.319, E11.321, E11.3211, E11.3212, E11.3213, E11.3219, E11.329, E11.3291, E11.3292, E11.3293, E11.3299, E11.331, E11.3311, E11.3312, E11.3313, E11.3319, E11.339, E11.3391, E11.3392, E11.3393, E11.3399, E11.341, E11.3411, E11.3412, E11.3413, E11.3419, E11.349, E11.3491, E11.3492, E11.3493, E11.3499, E11.351, E11.3511, E11.3512, E11.3513, E11.3519, E11.352, E11.3521, E11.3522, E11.3523, E11.3529, E11.353, E11.3531, E11.3532, E11.3533, E11.3539, E11.354, E11.3541, E11.3542, E11.3543, E11.3549, E11.355, E11.3551, E11.3552, E11.3553, E11.3559, E11.359, E11.3591, E11.3592, E11.3593, E11.3599, E11.36, E11.37, E11.37X1, E11.37X2, E11.37X3, E11.37X9, E11.39, E11.40, E11.41, E11.42, E11.43, E11.44,	LOINC code corresponding to A1C: • 17855-8 17856-6 4548-4 4549-2	Chloropropamide, glipizide, glyburide, glimepiride, metformin, repaglinide, nateglinide, tosiglitazone, pioglitazone, sitagliptin, saxagliptin, linagliptin, alogliptin, canagliflozin, dapagliflozin, acarbose, meglitol, colesevelam, insulin	-

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

			E11.49, E11.51, E11.52, E11.59, E11.610, E11.618, E11.620, E11.621, E11.622, E11.628, E11.630, E11.638, E11.641, E11.649, E11.65, E11.69, E11.8, E11.9			
Comorbidities						
A1C[4-10]	(1) Mean and standard deviation calculated after identifying labs using LOINC codes • Implausible A1C labs removed (range based on NHANES reporting) • Multiple labs measured on the same day or on the same day and time were averaged if they ranged within 1% • A1C labs ranging >1% were removed • If only one A1C lab was available prior to cohort entry date, that lab was used. • If more than one A1C lab was available, the average of the last two was taken	-	-	LOINC code corresponding to A1C: • 17855-8 • 17856-6 • 4548-4 • 4549-2	-	
BMI[11-14]	(1) Calculated as weight (Kg)/[height(m)]^2. Normal weight defined as BMI < 25, Overweight def < 30, and Obese Height: • Implausible values removed (range based on published literature)	-	-	-	-	Obtained by vital signs records

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

	• Multiple heights recorded during the same visit were averaged if they ranged within 3 inches (7.62 cm) or less • Measurements ranging more than 3 inches were deleted • If only one height measurement was available prior to cohort entry date, that height was used • If more than one height measurement was available prior to cohort entry date, the average of the last two was taken <u>Weight:</u> • Implausible values removed (range based on published literature) • Multiple weights recorded during the same visit were averaged if they ranged within 10 lb (4.536 Kg) or less. Measurements ranging more than 10 lb were deleted • If only one weight measurement was available prior to cohort entry date, that weight was used • If more than one weight measurement was available prior to cohort entry date, the average of the last two was taken					
Blood Pressure (BP)[15-19]	(1) Mean and standard deviation of systolic and diastolic BP calculated	-	-	-	-	Obtained by vital signs records

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

	• Records deleted if measured at nighttime (8pm to 7am) or if diastolic BP was greater than systolic BP • Only BP measured within the 2 years prior to cohort entry was included • BP measured on the same day was averaged • If only one BP measured was available on, or prior to, cohort entry date, it was used as the baseline BP • If more than one BP was available on, or prior to, cohort entry date, the average of the last two measurements was used as the baseline BP					
Hypertension (HTN)[20-24]	(1) At least one ICD-9/10 code for HTN	401.0, 401.1, 401.9	I10.X	-	-	-
Hyperlipidemia[25,26]	(1) Elevated total cholesterol (>240 mg/dL), *or* (2) Lipid-lowering medication use, *or* (3) at least 2 ICD-9/10 codes documenting hyperlipidemia	272.0, 272.1, 272.2, 272.3, 272.4, 272.5, 272.6, 272.7, 272.8, 272.9	E78, E78.0, E78.00, E78.01, E78.1, E78.2, E78.3, E78.4, E78.41, E78.49, E78.5, E78.6, E78.7, E78.70, E78.71, E78.72, E78.79, E78.8, E78.81, E78.89, E78.9	LOINC code corresponding to Total cholesterol: • 2093-3 • 14647-2	Generic names for class "CV350": • Atorvastatin • Cholestryamin • Colestipol • Ezetimibe • Ezetimibe/Simv-astatin • Gemfibrozil • Lomitapide • Mipomerson • Pravastatin • Rosuvastatin • Simvastatin	-

For peer review only

Lipids (HDL, LDL, Triglycerides)[25-27]	(1) Mean and standard deviation calculated after identifying labs using LOINC codes • Implausible values removed (range based on NHANES reporting) • The median of different lab values completed at the same day was taken • Lab values that are 50% > or < the median were considered outliers and removed • The median of the rest of the labs was taken as the lab value for that day • For all, if only one lab was available prior to the cohort entry date, that lab was used. If more than 1 lab value was available, the average of the last two was taken	-	-	LOINC codes corresponding to: <u>HDL:</u> • 2085-9 • 18263-4 • 9832-7 <u>LDL:</u> • 13457-7 • 18262-6 • 2089-1 • 14155-6 <u>Triglycerides:</u> • 2571-8 • 14927-8 • 12228-3 • 3049-4 • 1644-4 • 12951-0 • 3048-6	-	-
Chronic Kidney Disease (CKD)[24,28,29]	(1) At least one ICD-9/10 code documenting CKD	585.5X	N18.X	-	-	-
Ischemic Heart Disease[30]	(1) At least one ICD-9/10 code documenting ischemic heart disease	410.X, 411.X, 412.X, 413.X, 414.X	I20.X, I21.X, 122.X, 123.X, 124.X, 125.X	-	-	-
Heart Failure[30,31]	(1) At least one ICD-9/10 code documenting heart failure	428.X	I50.X	-	-	-
Peripheral Vascular Disease (PVD)[30,31]	(1) At least one ICD-9/10 code documenting PVD	440.0, 440.1, 440.2, 440.20, 440.21, 440.22, 440.23, 440.29, 440.4, 440.8, 440.9, 443.9, 557.0, 557.1, 557.9	I73.X	-	-	-
Stroke	(1) At least one ICD-9/10 code documenting stroke	346.60, 346.61, 346.62, 346.63, 432.0, 432.1, 432.9, 433.01, 433.11, 433.21, 433.31, 433.81, 433.91, 434.0, 434.00, 434.01, 434.1, 434.10,	I63.X	-	-	-

For peer review only

		434.11, 434.9, 434.90, 434.91, 436.5, 430.X, 431.X				
Agent Orange	(1) "Agent orange" flag was used to generate the number of people with this exposure	-	-	-	-	From patient problem lists
Hepatitis C	(1) At least one ICD-9/10 code documenting Hepatitis C	070.54	B18.2	-	-	-
Hyperuricemia	(1) At least one ICD-9/10 code documenting Hyperuricemia	790.6	E79.0	-	-	-
Polycystic Ovary Syndrome	(1) At least one ICD-9/10 code documenting Polycystic Ovary Syndrome	256.4	E28.2	-	-	-
Gestational Diabetes	(1) At least one ICD-9/10 code documenting Gestational Diabetes	V12.21, 648.83	Z86.32, 024.4X	-	-	-
eGFR[32,33]	(1) Mean and standard deviation calculated after identifying labs using LOINC codes • Implausible eGFR values were excluded • The median of eGFR measured on the same day was taken, then values more than 50% different than the median were excluded as outliers • The median of the rest of the measures was taken as eGFR for that day • If only one eGFR was available prior to cohort entry date, it was taken as the baseline estimate • If more than one eGFR was available prior to cohort entry date, the average of the most recent two was taken as the baseline estimate	-	-	LOINC code corresponding to <u>eGFR:</u> • 62238-1 • 48643-1 • 33914-3	-	-

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Liver enzymes	(1) Mean and standard deviation calculated after identifying labs using LOINC codes • Implausible values removed (range based on NHANES reporting) • The median of different lab values completed at the same day was taken • Lab values that are 50% > or < the median were considered outliers and removed • The median of the rest of the labs was taken as the lab value for that day For all, if only one lab was available prior to the cohort entry date, that lab was used. If more than 1 lab value was available, the average of the last two was taken	-	-	LOINC codes corresponding to <u>AST:</u> • 14409-7 • 14410-5 • 43822-6 • 88112-8 • 14412-1 • 14414-7 • 16412-9 • 1918-2 • 27344-1 • 14413-9 • 1917-4 • 1919-0 • 1920-8 • 30239-8 • 14411-3 • 44786-2 <u>ALT:</u> • 1741-8 • 25302-1 • 54491-6 • 1742-6 • 1743-4 • 44785-4 • 16324-6 • 50168-4 • 76625-3 • 1744-2 • 54492-4 • 77144-4	-	Exclusions: - Patients with hepatitis B and C (ICD-9: 070.2x, 070.3x, 070.41, 070.44, 070.51, 070.51, 070.54, 070.7x. ICD-10: B18.x) - alcohol abuse (ICD-9: 291.x, 303.0x, 303.9x, 305.0x. ICD-10 F10.x) - patients with other rare liver disease (ICD-9: 576.1, 275.03, 275.01, 275.1, 237.4, 571.42, 571.6, 275.09. ICD-10: 237.4, D44.0, D44.2, D44.9, 275.01, E83.110, 275.03, E83.118, E83.119, 275.09, E83.10, E83.19, 275.1, E83.00, E83.01, E83.09, 571.42, K75.4, 571.6, K74.3, K74.4, K74.5, 576.1, K83.0

For peer review only

	At least 2 elevated ALT (≥40 U/L) at least 6 months apart within 2 years					Exclusions: - Patients with hepatitis B and C (ICD-9: 070.2x, 070.3x, 070.41, 070.44, 070.51, 070.51, 070.54, 070.7x. ICD-10: B18.x) - alcohol abuse (ICD-9: 291.x, 303.0x, 303.9x, 305.0x. ICD-10 F10.x) - patients with other rare liver disease (ICD-9: 576.1, 275.03, 275.01, 275.1, 237.4, 571.42, 571.6, 275.09. ICD-10: 237.4, D44.0, D44.2, D44.9, 275.01, E83.110, 275.03, E83.118, E83.119, 275.09, E83.10, E83.19, 275.1, E83.00, E83.01, E83.09, 571.42, K75.4, 571.6, K74.3, K74.4, K74.5, 576.1, K83.0
Fatty liver disease[34]		-	-	-	-	

* Medications are documented prescriptions, not prescriptions filled

For peer review only

References

1. Jones CD, Greenwood RH, Misra A, Bachmann MO. Incidence and progression of diabetic retinopathy during 17 years of a population-based screening program in England. *Diabetes care.* 2012;35(3):592-596.
2. Parikh SV, Saya S, Divanji P, et al. Risk of death and myocardial infarction in patients with peripheral arterial disease undergoing percutaneous coronary intervention (from the National Heart, Lung and Blood Institute Dynamic Registry). *The American journal of cardiology.* 2011;107(7):959-964.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

3. Rassen JA, Bartels DB, Schneeweiss S, Patrick AR, Murk W. Measuring prevalence and incidence of chronic conditions in claims and electronic health record databases. *Clinical epidemiology.* 2019;11:1-15.
4. Carls G, Huynh J, Tuttle E, Yee J, Edelman SV. Achievement of Glycated Hemoglobin Goals in the US Remains Unchanged Through 2014. *Diabetes Ther.* 2017;8(4):863-873.
5. Pani LN, Korenda L, Meigs JB, et al. Effect of aging on A1C levels in individuals without diabetes: evidence from the Framingham Offspring Study and the National Health and Nutrition Examination Survey 2001-2004. *Diabetes care.* 2008;31(10):1991-1996.
6. Nicolo M, Boullata JI. Serum 25OHD concentration as a predictor of haemoglobin A1c among adults living in the USA: NHANES 2003 to 2010. *BMJ Nutrition, Prevention & Health.* 2019:bmjnph-2019-000029.
7. Davis PJ, Liu M, Sherman S, et al. HbA1c, lipid profiles and risk of incident type 2 Diabetes in United States Veterans. *PloS one.* 2018;13(9):e0203484.
8. Gillett M. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes care.* 2009;32(7):1327-1334.
9. *Laboratory Procedure Manual: Glycohemoglobin*
10. Heinemann LF, G. Quality of HbA1c measurement in the practice: The german perspective. *J Diabetes Science and Technology.* 2015;9(3):687-695.
11. Fryar CK-M, D; Gu, Q; Ogden, CL. Mean body weight, height, waist circumference, and body mass index among adults: United States, 1999-2000 through 2015-2016. *National Health Statistics Reports.* 2018;122.
12. McDowell M, Hughes J, Borrud LG. Health characteristics of U.S. adults by body mass index category: results from NHANES 1999-2002. *Public Health Rep.* 2006;121(1):67-73.
13. Martinez K, Tucker L, Bailey B, LeCheminant J. Expanded Normal Weight Obesity and Insulin Resistance in US Adults of the National Health and Nutrition Examination Survey. *J Diabetes Res.* 2017;2017:9502643-9502643.
14. Tarlov E, Silva A, Wing C, et al. Neighborhood Walkability and BMI Change: A National Study of Veterans in Large Urban Areas. *Obesity (Silver Spring, Md).* 2020;28(1):46-54.
15. Armstrong C. JNC8 guidelines for the management of hypertension in adults. *American family physician.* 2014;90(7):503-504.
16. Rowan CG, Turner JR, Shah A, Spaeder JA. Antihypertensive treatment and blood pressure control relative to hypertension treatment guidelines. *Pharmacoepidemiology and drug safety.* 2014;23(12):1294-1302.
17. Freitag MH, Vasan RS. What is normal blood pressure? *Current opinion in nephrology and hypertension.* 2003;12(3):285-292.
18. O'Brien E, Parati G, Stergiou G. Ambulatory blood pressure measurement: what is the international consensus? *Hypertension (Dallas, Tex : 1979).* 2013;62(6):988-994.
19. Agarwal R, Light RP. The effect of measuring ambulatory blood pressure on nighttime sleep and daytime activity--implications for dipping. *Clin J Am Soc Nephrol.* 2010;5(2):281-285.
20. Burg MM, Brandt C, Buta E, et al. Risk for Incident Hypertension Associated With Posttraumatic Stress Disorder in Military Veterans and the Effect of Posttraumatic Stress Disorder Treatment. *Psychosomatic medicine.* 2017;79(2):181-188.
21. Fletcher RD, Jones R, Moore H, et al. INCREASED BODY MASS INDEX (BMI) PROMOTES HYPERTENSION AND WORSENS BLOOD PRESSURE CONTROL AMONG US VETERANS. *Journal of the American College of Cardiology.* 2018;71(11 Supplement):A1817.
22. Peng M, Chen G, Kaplan GG, et al. Methods of defining hypertension in electronic medical records: validation against national survey data. *Journal of public health (Oxford, England).* 2016;38(3):e392-e399.
23. Banerjee D, Chung S, Wong EC, Wang EJ, Stafford RS, Palaniappan LP. Underdiagnosis of hypertension using electronic health records. *American journal of hypertension.* 2012;25(1):97-102.
24. *Hypertension prevalence estimator tool.*
25. Soohoo M, Moradi H, Obi Y, Kovesdy CP, Kalantar-Zadeh K, Streja E. Serum triglycerides and mortality risk across stages of chronic kidney disease in 2 million U.S. veterans. *Journal of clinical lipidology.* 2019;13(5):744-753.e715.
26. Laclaustra M, Stranges S, Navas-Acien A, Ordovas JM, Guallar E. Serum selenium and serum lipids in US adults: National Health and Nutrition Examination Survey (NHANES) 2003-2004. *Atherosclerosis.* 2010;210(2):643-648.
27. Toth PP, Potter D, Ming EE. Prevalence of lipid abnormalities in the United States: the National Health and Nutrition Examination Survey 2003-2006. *Journal of clinical lipidology.* 2012;6(4):325-330.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

28. Fleet JL, Dixon SN, Shariff SZ, et al. Detecting chronic kidney disease in population-based administrative databases using an algorithm of hospital encounter and physician claim codes. *BMC Nephrology.* 2013;14(1):81.
29. Prevention CfDCa. *Prevalence of patients with an ICD-9-CM or ICD-10-CM code indicating CKD.*
30. Quan HS, V.; Halfon, P.; et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical care.* 2005;43(11):1130-1139.
31. Lix LS, M.; Piz, M.; Ahmed, R.; Quon, H.; Griffith, J., et al. *Cancer data linkage in manitoba: Expanding the instrastructure for research.* 2016.
32. Grams ME, Juraschek SP, Selvin E, et al. Trends in the prevalence of reduced GFR in the United States: a comparison of creatinine- and cystatin C-based estimates. *Am J Kidney Dis.* 2013;62(2):253-260.
33. Hemmelgarn BR, Manns BJ, Lloyd A, et al. Relation Between Kidney Function, Proteinuria, and Adverse Outcomes. *JAMA.* 2010;303(5):423-429.
34. Husain N, Blais P, Kramer J, et al. Nonalcoholic fatty liver disease (NAFLD) in the Veterans Administration population: development and validation of an algorithm for NAFLD using automated data. *Alimentary Pharmacology & Therapeutics.* 2014;40(8):949-954.

For peer review only

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	2
Objectives	3	State specific objectives, including any prespecified hypotheses	2
Methods			
Study design	4	Present key elements of study design early in the paper	2,3
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	3
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	3
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	n/a
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	3,4
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	2
Bias	9	Describe any efforts to address potential sources of bias	n/a
Study size	10	Explain how the study size was arrived at	n/a
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	4
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	n/a
		(*b*) Describe any methods used to examine subgroups and interactions	n/a
		(*c*) Explain how missing data were addressed	n/a
		(*d*) If applicable, explain how loss to follow-up was addressed	3
		(*e*) Describe any sensitivity analyses	n/a
Results			

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	3
		(b) Give reasons for non-participation at each stage	n/a
		(c) Consider use of a flow diagram	Figure1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	4,5 Table 1
		(b) Indicate number of participants with missing data for each variable of interest	n/a
		(c) Summarise follow-up time (eg, average and total amount)	3
Outcome data	15*	Report numbers of outcome events or summary measures over time	n/a

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included (*b*) Report category boundaries when continuous variables were categorized (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n/a
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	4/5
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	5
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	5
Generalisability	21	Discuss the generalisability (external validity) of the study results	5
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	6

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

For peer review only

BMJ Open

Cohort Profile: The US Veterans Administration Diabetes Risk (VADR) National Cohort

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-039489.R1
Article Type:	Cohort profile
Date Submitted by the Author:	06-Nov-2020
Complete List of Authors:	Avramovic, Sanja; George Mason University; VA New York Harbor Healthcare System Alemi, Farrokh ; George Mason University Kanchi, Rania; New York University School of Medicine Lopez, Priscilla; New York University School of Medicine Hayes, Richard; New York University School of Medicine Thorpe, Lorna; New York University School of Medicine, Department of Population Health Schwartz, Mark; New York University School of Medicine; VA New York Harbor Healthcare System
Primary Subject Heading:	Diabetes and endocrinology
Secondary Subject Heading:	Epidemiology
Keywords:	General diabetes < DIABETES & ENDOCRINOLOGY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, EPIDEMIOLOGY

SCHOLARONE™
Manuscripts

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

For peer review only

Cohort Profile: The US Veterans Administration Diabetes Risk (VADR) National Cohort

Sanja Avramovic,[1,3] Farrokh Alemi,[1] Rania Kanchi,[2] Priscilla Lopez,[2] Richard B Hayes,[2] Lorna E. Thorpe,[2] Mark D Schwartz[2,3]

[1] Department of Health Administration and Policy, George Mason University, Fairfax, VA
[2] Department of Population Health, NYU School of Medicine, New York, NY
[3] VA New York Harbor Healthcare System, New York, NY

Correspondence to
Dr. Sanja Avramovic; savramov@gmu.edu

Abstract

Purpose The Veterans Administration Diabetes Risk (VADR) cohort facilitates studies on temporal and geographic patterns of prediabetes and diabetes, as well as targeted studies of their predictors. The cohort provides an infrastructure for examination of novel individual and community-level risk factors for diabetes and their consequences among veterans. This cohort also establishes a baseline against which to assess the impact of national or regional strategies to prevent diabetes in veterans.

Participants The VA Diabetes Risk Cohort (VADR) includes all 6,082,018 veterans in the United States who were diabetes-free as of January 1, 2008, or who subsequently enrolled in the VA for primary care and were diabetes-free at cohort entry through December 31, 2016, and who had at least 2 diabetes-free visits to a VA primary care service at least 30 days apart within any 5-year period since January 1, 2003. Cohort subjects were followed from the date of cohort entry until censure defined as date of incident diabetes, loss to follow-up of 2 years, death, or until December, 31, 2018.

Findings to Date The incidence rate of type 2 diabetes in this cohort of over 6 million veterans followed for a median of 5.5 years (over 35 million person-years) was 26 per 1000 person-years. During the study period, 8.5% of the cohort were lost to follow-up and 17.7% died. Many demographic, comorbidity, and other clinical variables were more prevalent among patients with incident diabetes.

Future Plans This cohort will be used to study community-level risk factors for diabetes, such as attributes of the food environment and neighborhood socioeconomic status via geospatial linkage to residence address information.

Strengths and Limitations of this Study

- A strength of this national cohort is that it has a large size, a high degree of long-term follow-up, and a comprehensive set of variables.
- The VA healthcare system is the nation's largest integrated healthcare system, in which veterans are followed across all VA facilities and in-system providers.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

- Documented data are restricted to that which is collected in EHRs during the course of clinical practice, leading to the possibility of confounding, selection bias and measurement error.
- The veteran population is predominantly male and white, and so the findings may not generalize to minorities or to women.

Introduction

Diabetes mellitus (diabetes) is a chronic disease that affects 34.2 million (10.5%) of adults and children in the United States (US).[1] As of 2018, diabetes was the seventh leading cause of death and one of the major contributors to heart disease and stroke.[2] Adjusting for age and gender, all-cause mortality is 1.5 times greater for people with diabetes than for people without diabetes, and average health care costs are 2.3 times higher.[3] Another 88 million American adults (34.5%) are estimated to have prediabetes and at risk of developing diabetes.[4]

The Veterans Administration (VA) cares for more than 8 million US veterans, of whom approximately 25% have diabetes.[5,6] The annual mortality rate among veterans with diabetes is 5%—nearly double that of veterans without diabetes.[7,8] It is likely that nearly 3 million other veterans have prediabetes. These high rates compared to the general population may be due to the increased proportions of overweight (37%) and obesity (41%) among veterans,[9] their older age, lower socioeconomic status,[10] and possible exposure to herbicides such as Agent Orange.[11]

Behavioral prevention interventions can reduce the incidence of diabetes by 50–70%,[12,13] but scaling this up for population impact has been challenging due to the intensity and cost of the intervention and challenges of enrolling patients for such programs.[14-17]

In response to these challenges, we developed the Veterans Administration Diabetes Risk (VADR) Cohort, a national cohort of all US veterans enrolled at the Veterans Health Administration (VHA) since January 1, 2008 who were diabetes-free at enrollment. The cohort was developed as a part of the Diabetes Location, Environmental Attributes, and Disparities (LEAD) network; a Center for Disease Control and Prevention (CDC) funded research collaboration among Drexel University, Geisinger-Johns Hopkins, New York University School of Medicine, and University of Alabama at Birmingham with the CDC as a collaborative scientific partner in the network.[18]

The VADR cohort facilitates studies on temporal and geographic patterns of prediabetes and diabetes, as well as targeted studies of their predictors. For example, the cohort currently provides the infrastructure for the nationwide study examining community-level risk factors for diabetes incidence and management among veterans described above. This cohort also establishes a baseline against which to assess the impact of national or regional strategies to prevent diabetes in veterans. It also provides an analytic cohort to examine the dynamic relationship between the COVID-19 pandemic and diabetes outcomes.

Cohort Description

VADR is the largest national cohort of diabetes-free adults in the US. Established in 2017 as a dynamic cohort enabled by the VA national electronic health record (EHR), the cohort includes

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

diabetes-free US veterans enrolled in primary care clinics at any VA facility as early as January 1, 2008 through December 31, 2016, and followed from cohort entry through December 31, 2018. VA primary care clinics operate in 170 VA Medical Centers (VAMCs) and in more than 1,000 Community-Based Outpatient Clinics (CBOCs) across the US.[19] As a dynamic cohort, subject follow-up is ongoing, but this paper reports on the cohort from January 1, 2008 through December 31, 2018.

Building on published, validated criteria in EHRs,[7,20] we defined diabetes using the following query-based definition comprised of any of three criteria: (1) at least two encounters (inpatient or outpatient) with documentation of a Type 2 diabetes ICD-9/10 code (ICD-9: 250.x; ICD-10: E11.x) or (2) a documented prescription for a diabetes medication other than metformin or acarbose alone; or (3) at least one encounter with a diabetes ICD-9/10 code and two elevated (≥ 6.5%) glycosylated hemoglobin (Hgb A1C) lab test results (see Appendix for complete definition).[21] We excluded metformin or acarbose alone from the criteria because these drugs may be used for diabetes prevention in patients with prediabetes; including them may lead to misclassifying cases of prediabetes as diabetes.[22,23] This definition for incident diabetes was used to exclude prevalent diabetes cases prior to cohort entry and to estimate diabetes incidence during the study period.

For the analytic cohort, subjects were eligible if they were veterans with at least 2 diabetes-free visits to a VA primary care service, occurring at least 30 days apart, from January 1st 2003 to December 31st 2016. Cohort entry (baseline) was defined as either January 1, 2008 or the date of the second diabetes-free primary care visit for subjects entering after January 1, 2008. Eligible subjects were allowed to enter the cohort through December 31, 2016 to allow at least 2 years of follow-up during which subjects may be diagnosed with diabetes. Subjects were censored when they developed diabetes, died, or were lost to follow-up (defined as having no encounters in the VA health system for more than 2 years). Once a patient was lost-to-follow-up, they were not eligible to re-enter the cohort. Encounters for follow-up included any visits to primary care, specialists, emergency departments, walk-in clinics, hospitalizations, or nursing home stays at any VA facility. Person-years (PY) of follow-up for each subject were calculated as the interval between cohort entry date and censor date.

As shown in **Figure 1**, the cohort was developed from a base total population of 8,346,180 patients seen for at least 1 primary care visit between 1999, the earliest year for which EHR data were available on patients, and the start of the study period. The cohort was then restricted to patients seen in the five years prior to the study period start date, January 1, 2008. Patients were excluded if they had fewer than 2 primary care visits, at least 30 days apart during that five year time period and less than 2 primary care visits after cohort entry. After excluding patients with prevalent diabetes, the initial diabetes-free cohort included 2,968,763 patients. Another 3,113,255 diabetes-free patients met the same eligibility criteria after the start of the study period and entered the cohort between January 1, 2008 and December 31, 2016, resulting in a diabetes-free cohort of 6,082,018 patients.

Information on subjects in the cohort was updated daily as it was drawn from EHR at all VA facilities into the VA corporate data warehouse (CDW), based on all clinical services provided and documented by the VA to subjects over time. All data in the cohort were obtained through

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

the VA Informatics and Computing Infrastructure (VINCI), a secure, high performance interface with VA's national CDW, available through VA's Information Resource Center (VIReC).[24] The CDW contains data integrated from VA's electronic medical record (VISTA, Veterans Health Information Systems and Technology Architecture), including all administrative data (e.g. all dates of encounters and diagnostic codes for outpatient and inpatient care), patient demographic characteristics, clinical data (e.g. vital signs, health factors, pharmacy, laboratory, radiological, clinical notes, etc.), and healthcare utilization factors as they accrue over time, as the CDW is refreshed daily.[25]

The main outcome variable was a new diagnosis of Type 2 diabetes, measured using the definition described earlier.

Predictor variables and covariates

All continuous variables with repeated measures, including anthropomorphic, vital signs, and laboratory values, were defined as the average of the two most recent measures, prior to or at the time of cohort entry. If only one measure was taken prior to cohort entry, that was used as the baseline measure. The rate of missing data for all variables was measured.

Demographic measures were captured at baseline, including age, gender, marital status, race/ethnicity. First address on file per patient in cohort were exported out of the VINCI environment, geocoded using ArcGIS[26] and Python[27], and mapped to show number of patients in the cohort per census tract using QGIS.[28]

Glycemia and body weight are important predictors of diabetes. We measured Hgb A1c as a continuous value, and classified as normal (<5.7%), prediabetes (5.7% to 6.4%), or diabetes (≥6.5%). We measured weight in pounds and body mass mass index (BMI), defined as (weight in kilograms) / (height in meters)^2. BMI was also classified as underweight (<18.5); normal (18.5 to <25); overweight (25.0 to <30); and obese (≥30.0).[29]

Common comorbidities measured at baseline included established risk factors for diabetes such as obesity, hypertension, gestational diabetes, cardiovascular disease, chronic kidney disease hyperuricemia, fatty liver disease, polycystic ovary syndrome, and hepatitis C. These and all other comorbidities were defined as having at least 1 ICD code in the EHR prior to entering the cohort. Hyperlipidemia was defined as at least 2 encounters with ICD codes for hyperlipidemia, total cholesterol >240 mg/dL, or the use of lipid lowering medications.[30] Hypertension was defined as at least ICD code for hypertension or at least 2 consecutive elevated BP within the last two years prior to cohort entry.[31,32] Elevated BP was included as as ≥130/80 and ≥140/90, respectively, to comply with changes in hypertension guidelines over the course of the study period.[33,34]

Other clinical variables potentially related to diabetes incidence included: Blood Pressure (BP, excluding those measured in the hospital, emergency department, or at night); Lipids (Total Cholesterol, High Density Lipoprotein, Low Density Lipoprotein, and Triglycerides; Hepatic Transaminase Enzymes (serum aspartate aminotransferase - AST or SGOT - and alanine aminotransferase - ALT or SGPT); Renal Function (measured as Estimated Glomerular Filtration Rate - eGFR); Smoking Status: (Obtained from Health Factor files within CDW at cohort entry, classified as current, ever, or never smokers); and Agent Orange Exposure (Obtained from the

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

number of veterans with Agent Orange listed as a health factor in the medical record).[11] Beside this select list, all documented diagnoses and treatments are available for the cohort.

Access to Cohort Data

Access to VA electronic health records data is limited to researchers with active, VA appointments and an IRB-approved protocol. Once a researcher has a VA appointment and IRB approval, the VA has a comprehensive data infrastructure to support secure and remote access to data via the VINCI platform. Additionally, deidentified datasets can be established and shared with appropriate IRB approval and data use agreements. The authors encourage collaborations to leverage this cohort to examine how national or regional natural experiments may be related to diabetes incidence or diabetes outcomes.

Findings to Date

The total person-years (PY) for this national cohort with 6,082,018 veterans from all 50 states was 35,889,183 (median 5.5 PY, IQR: 2.6 - 9.8). As shown in **Table 1**, the mean age of the cohort was 58 years at baseline, 36.4% were 65 or older, most were male (91.7%), more than two-thirds were non-Hispanic white (74.8%), 16.3% were non-Hispanic black, and 6.1% were Hispanic. The majority (55.2%) were married or living with a partner.

At baseline, the average Hgb A1C was 5.7% among the 40.7% of the cohort tested at entry, and of these, 41.5% had an Hgb A1C in the prediabetes range. The average weight was 196.9 pounds and average BMI was 28.8 (SD 5.4). At baseline, 40.6% were overweight and 36.1% were obese. Traditional clinical risk factors for diabetes were common in this cohort as 46.1% had hypertension, 44.1% had hyperlipidemia, and 42.6% were smokers. Other clinical risk factors for diabetes included ischemic heart disease (16.4%), peripheral vascular disease (4.2%), heart failure (3.0%), and chronic kidney disease (2.5%). Most of these risk factors were present at baseline at higher rates among those who developed diabetes compared with those who did not during cohort follow-up.

Figure 2 shows the number of subjects in the cohort over time, from inception January 1, 2008 through December 31, 2018. Almost half (48.8%) of the cohort entered at cohort inception in January 1, 2008, with the remainder entering during the study period through December 31, 2016. During cohort follow-up, 936,596 (15.4%) veterans developed diabetes, for an incidence rate of 26 per 1,000 PY. Additionally, 518,489 (8.5%) were lost to follow-up, and 1,077,572 (17.7%) died during the study period. **Figure 3** shows the geographic distribution of the number of patients per tract. The majority of addresses were able to be geocoded (89%); of those not geocoded, about half were PO boxes, and the other half were missing. The majority of census tracts had between 20-80 patients.

Because cohort data were drawn from the VA EHR, which depends on documentation of services provided, some subjects had missing values for some variables at baseline. For example, the percentage of missing variables at cohort entry were: gender (<0.01%); race/ethnicity (10.1%); marital status (7.5%); BMI (4.3%); and Hgb A1C (59.3%). The missing

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

race/ethnicity variable in VA data is widely known.[35] Screening for diabetes with Hgb A1c became more common after the recommendation was published in 2009.[36]

Strengths and Limitations

A primary strength of this national cohort is its large size and long-term follow-up. The cohort includes a comprehensive set of demographic, anthropomorphic, clinical, treatment, and other administrative variables, drawn from all inpatient and outpatient encounters, each of which are automatically updated over time. In addition to the select comorbidities identified in this paper, the cohort includes data related to all comorbidities. Future work will include calculation of a multi-morbidity index to measure the impact of medical history on emergence of diabetes.[37]

As the nation's largest, integrated healthcare system, the VA follows veterans across all VA facilities, even after moving and changing VA facilities or providers within the system. Additionally, data on veterans who are Medicare or Medicaid beneficiaries and seek health care outside of the VA will be included by merging the study cohort with data from the Centers for Medicare and Medicaid (CMS). Finally, home addresses are available and were geocoded in order to study the effect of community level characterstics and the impact of moving over time on incident diabetes in future work using this cohortL

The cohort has a few limitations. It relies on data documented during the course of clinical practice in EHRs and thus causal inferences face difficulties associated with unmeasured confounding, selection bias, and measurement error. Selection biases may arise as lower health care utilizers are more likely to be lost to follow-up or excluded, and higher utilizers may be more likely to meet criteria for key exposure and outcome variables. This is partially mitigated by the several-year, longitudinal follow-up.

The veteran population is predominantly male and white, and so the findings may not generalize to minorities or to women. Nonetheless, our large cohort ensures a sufficient and growing sample of women veterans (504,002) and patients from major ethnic/racial groups (889,465 NH black veterans, 331,817 Hispanic veterans), providing the ability to study diabetes incidence among these subgroups and improving the generalizability of our findings to non-veteran populations.

Conclusion

The VA Diabetes Risk Cohort (VADR) is an important example of how large retospecitve cohorts can be developed using electronic health records, designed with methodologic and statistical approaches to increase generalizability and validity. The benefits of such large cohorts are that they can offer more information and ability to examine associations in substrata than smaller cohorts. Follow up is ongoing and presented here through December 31, 2018. While the main outcome of interest was incidence of type 2 diabetes in this cohort, the infrastructure is well-suited to support studies of diabetes management and management of other chronic conditions using incident cases of diabetes, particularly as retention has been shown to be

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

good. During the study period, only 8.5% of the cohort were lost to follow-up and 17.7% died. Additional methodologic work is needed to address biases unique to EHR-based observational studies, including cohort selection bias and nonignorable missing data

Patient and Public Involvement

This cohort study was conducted without engagement or co-production by patients or the public.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 1. Cohort demographics and clinical characteristics at cohort entry by incident diabetes status

	All veterans^ n (%) or mean (SD)	Veterans without incident diabetes ǂ n (%) or mean (SD)	Veterans with incident diabetes n (%) or mean (SD)	Not measured or Missing n (%)
Total	6,082,018	5,145,422	936,596	
Demographic characteristics				
Age	58.3 (17)	53.9 (16.1)	61.1 (12.2)	
Age categories				
18-34	745511 (12.3)	592256 (16.7)	21788 (2.3)	
35-49	1009677 (16.6)	725427 (20.4)	130345 (13.9)	
50-64	2114275 (34.8)	1280132 (36.1)	451964 (48.3)	
65-79	1499787 (24.7)	786213 (22.2)	261592 (27.9)	
80+	712681 (11.7)	165264 (4.7)	70900 (7.6)	
Gender				124 (0)
Male	5577892 (91.7)	3167502 (89.2)	886763 (94.7)	
Female	504002 (8.3)	381768 (10.8)	49818 (5.3)	
Race ethnicity				612210 (10.07)
Non-Hispanic white	4092942 (74.8)	2412225 (73.8)	614511 (70.8)	
Non-Hispanic black	889465 (16.3)	544434 (16.7)	173057 (19.9)	
Hispanic	331817 (6.1)	211459 (6.5)	55301 (6.4)	
Non-Hispanic Asian	55564 (1)	37545 (1.2)	7705 (0.9)	
NH-Native Hawaiian or other Pacific Islandar	50426 (0.9)	30228 (0.9)	9020 (1)	
NH-American Indian or Alaska Native	49594 (0.9)	31107 (1)	8474 (1)	
Marital status				453512 (7.46)
Married or living with a partner	3104312 (55.2)	1832844 (55.2)	477337 (56.2)	
Single	2524194 (44.9)	1488661 (44.8)	372630 (43.8)	
Clinical characteristics				
HbA1c	5.7 (0.6)	5.5 (0.4)	6.1 (0.9)	3608600 (59.33)
HbA1c categories				
Normal (<5.7%)	1311768 (53)	875677 (62.3)	128681 (25.8)	

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Prediabetes (5.7%-6.49%)	1027336 (41.5)	514758 (36.6)	268105 (53.8)	
Diabetes range (≥6.5%)	134314 (5.4)	15302 (1.1)	101746 (20.4)	
Weight in pounds	196.9 (40.7)	196.6 (38.9)	214.7 (45.1)	260313 (4.28)
BMI	28.8 (5.4)	28.7 (5.1)	31.3 (6)	260313 (4.28)
Underweight (<18.5)	48954 (0.8)	18042 (0.5)	4417 (0.5)	
Normal weight (18.5-<25)	1308427 (22.5)	732429 (21.6)	105224 (11.7)	
Overweight (25-<30)	2362867 (40.6)	1441730 (42.5)	294224 (32.6)	
Obese (≥30)	2101457 (36.1)	1197716 (35.3)	499082 (55.3)	
Measured blood pressure*				
Systolic blood pressure	130.4 (14.7)	129.3 (14.2)	133.2 (14.9)	290882 (4.78)
Diastolic blood pressure	76.8 (10)	77.6 (9.6)	78 (10.1)	290882 (4.78)
Elevated blood pressure (≥130/80)	3516559 (60.7)	2011621 (59.4)	609072 (67.9)	290882 (4.78)
Elevated blood pressure (≥140/90)	1499481 (25.9)	796617 (23.5)	287665 (32.1)	290882 (4.78)
Hypertension				
≥1ICD code or 2 consecutive elevated BP* (≥130/80)	2032490 (33.4)	2137040 (60.2)	760631 (81.2)	
≥1ICD code or 2 consecutive elevated BP* (≥140/90)	2805063 (46.1)	1617389 (45.6)	667469 (71.3)	
Lipids				
Total cholesterol	185.5 (38.4)	188.3 (37.7)	185.4 (40.2)	1101271 (18.11)
Triglyceride	140.8 (88.9)	137.6 (86.9)	168.1 (103.2)	1118227 (18.39)
LDL	112.6 (33.1)	115.3 (32.8)	111.4 (34.1)	1151308 (18.93)
HDL	46.3 (14.2)	47.3 (14.3)	42.4 (12.5)	1126777 (18.53)
Hyperlipidemia¥	2681683 (44.1)	1371518 (38.6)	408288 (43.6)	
Smoking status				3853822 (63.36)
Current smoker	948272 (42.6)	561968 (41.9)	140333 (44.3)	
Not a smoker	1279924 (57.4)	780134 (58.1)	176777 (55.8)	
Estimated glomerular filtration rate (eGFR)	80.1 (18.3)	82.5 (16.7)	78.6 (18.4)	2718664 (44.7)
eGFR categories				

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

eGFR≥90 (stage 1)	1036900 (30.8)	691968 (33.9)	145390 (27.9)	
eGFR ≥60 to <90 (stage 2)	1899314 (56.5)	1180774 (57.8)	300183 (57.7)	
eGFR<60 (stage 3, 4, or 5)	427140 (12.7)	169856 (8.3)	75168 (14.4)	
Chronic kidney disease (ICD codes)	150823 (2.5)	47353 (1.3)	29236 (3.1)	
Ischemic heart disease (ICD codes)	999927 (16.4)	378832 (10.7)	209448 (22.4)	
Heart failure (ICD codes)	181375 (3)	40411 (1.1)	41272 (4.4)	
Peripheral vascular disease (ICD codes)	256054 (4.2)	77339 (2.2)	55242 (5.9)	
Stroke (ICD codes)	30423 (0.5)	10724 (0.3)	6275 (0.7)	
Agent orange exposure (exposed)				
Chronic hepatitis C	102534 (1.7)	51789 (1.5)	21382 (2.3)	
Hyperuricemia	180941 (3)	74543 (2.1)	54981 (5.9)	
Polycystic ovary syndrome	4970 (1)	3748 (1)	798 (1.6)	
Gestational diabetes	157 (0.03)	131 (0.03)	22 (0.04)	
Liver enzymes				
AST	25.7 (14.3)	25.7 (14.2)	26.7 (14.7)	1731271 (28.47)
ALT	29.5 (19.8)	30.3 (19.9)	32.4 (21.1)	1634812 (26.88)
Elevated AST (>40 U/L)	277,603 (6.4)	156,503 (6.2)	56,466 (8.3)	1731271 (28.47)
Elevated ALT (>40 U/L)	778,931 (17.5)	482,342 (18.6)	158,747 (22.8)	1634812 (26.88)
Fatty liver disease	253134 (10.4)	136847 (10.8)	67367 (15.7)	3654047 (60.08)

* Only recent (within 2 years of cohort entry) BP measurements were used. Nighttime BP (8 PM to 7 AM) and BP measured in ER were excluded

¥ Hyperlipidemia was defined as: at least 2 ICD-9/10 codes for hyperlipidemia, total cholesterol > 240 mg/dL, or lipid lowering medication use

^ Including those lost to follow up and those died during the study period

ł Only patients who completed the follow up and were diabetes free at the end of the study

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Contributorship statement

Dr. Avramovic has full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Schwartz, Thorpe, Hayes, Avramovic, Alemi.
Acquisition, analysis, or interpretation of data: Avramovic, Lopez, Kanchi, Schwartz.
Drafting of the manuscript: Avramovic, Schwartz, Kanchi, Lopez, Thorpe
Critical revision of the manuscript for important intellectual content: All authors, Avramovic, Alemi, Kanchi, Lopez, Hayes, Thorpe, and Schwartz.
Statistical analysis: Kanchi

All authors, Avramovic, Alemi, Kanchi, Lopez, Hayes, Thorpe, and Schwartz, attest that they meet the full authorship criteria.

Competing interests

The authors report no competing interests.

Funding

This study was funded by the Centers for Disease Control and Prevention (5 U01DP006299-02-00; PI: LET).

Data Sharing Statement

To gain access to data from the diabetes cohort described in this profile, interested researchers can contact the corresponding author.Further details

Further details regarding the ability to access VA data can be found on the VA website dedicated to researchers: https://www.hsrd.research.va.gov/for_researchers/default.cfm, including links to policies and guidance documents, special interest groups, funding opportunities, and a link to the VA Informatics and Computing Infrastructure (VINCI) site where access to actual data is granted once appropriate applications have been submitted and approved: https://www.hsrd.research.va.gov/for_researchers/vinci/.

For peer review only

References

1. Centers for Disease Control and Prevention. *Estimates of Diabetes and its Burden in the United States.* Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2017;2017.
2. Centers for Disease Control and Prevention. *Underlying Cause of Death 1999-2008.* National Center for Health Statistics;2018.
3. American Diabetes Association. Economic costs of diabetes in the U.S. in 2012. *Diabetes care.* 2013;36(4):1033-1046.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

4. Prevention CfDCa. Diabetes Quick Facts. 2019; https://www.cdc.gov/diabetes/basics/quick-facts.html. Accessed February 27, 2020.
5. Veterans Health Administration. VA Research on Diabetes. 2019.
6. Liu Y, Sayam S, Shao X, et al. Prevalence of and Trends in Diabetes Among Veterans, United States, 2005-2014. *Preventing chronic disease.* 2017;14:E135.
7. Miller D, Safford M, Pogach L. Who has diabetes? Best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data. *Diabetes care.* 2004;27 Suppl 2:B10-21.
8. Veterans Health Administration. Close to 25 Percent of VA Patients Have Diabetes. 2011; http://www.va.gov/health/NewsFeatures/20111115a.asp Accessed December 31, 2015.
9. Breland J, Phibbs C, Hoggatt K, et al. The Obesity Epidemic in the Veterans Health Administration: Prevalence Among Key Populations of Women and Men Veterans. *Journal of general internal medicine.* 2017;32(Suppl 1):11-17.
10. Kondo K, Low A, Everson T, et al. Health Disparities in Veterans: A Map of the Evidence. *Medical care.* 2017;55 Suppl 9 Suppl 2:S9-s15.
11. National Academies of Sciences E, and Medicine. *Veterans and Agent Orange.* National Academies Press;2018.
12. Prevention CfDCa. *National Diabetes Prevention Program.* 2019.
13. Group DPPR, Knowler W, Fowler S, et al. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. *Lancet (London, England).* 2009;374(9702):1677-1686.
14. Ackermann R, Finch E, Brizendine E, Zhou H, Marrero D. Translating the Diabetes Prevention Program into the community. The DEPLOY Pilot Study. *American journal of preventive medicine.* 2008;35(4):357-363.
15. Ackermann R, Marrero D. Adapting the Diabetes Prevention Program lifestyle intervention for delivery in the community: the YMCA model. *The Diabetes educator.* 2007;33(1):69, 74-65, 77-68.
16. Baker M, Simpson K, Lloyd B, Bauman A, Singh M. Behavioral strategies in diabetes prevention programs: a systematic review of randomized controlled trials. *Diabetes research and clinical practice.* 2011;91(1):1-12.
17. Schellenberg E, Dryden D, Vandermeer B, Ha C, Korownyk C. Lifestyle interventions for patients with and at risk for type 2 diabetes: a systematic review and meta-analysis. *Annals of internal medicine.* 2013;159(8):543-551.
18. Diabetes LEAD (Location, Environmental Attributes, and Disparities) Network. http://diabetesleadnetwork.org/. Accessed February 27, 2020.
19. Administration VH. About VHA. https://www.va.gov/health/aboutvha.asp. Accessed February 27, 2020.
20. Bowe B, Xie Y, Li T, Yan Y, Xian H, Al-Aly Z. The 2016 global and national burden of diabetes mellitus attributable to PM2.5 air pollution. *The Lancet Planetary health.* 2018;2(7):e301-e312.
21. Davis PJ, Liu M, Sherman S, et al. HbA1c, lipid profiles and risk of incident type 2 Diabetes in United States Veterans. *PloS one.* 2018;13(9):e0203484.
22. Tseng E, Yeh HC, Maruthur NM. Metformin Use in Prediabetes Among U.S. Adults, 2005-2012. *Diabetes care.* 2017;40(7):887-893.
23. Hu R, Li Y, Lv Q, Wu T, Tong N. Acarbose Monotherapy and Type 2 Diabetes Prevention in Eastern and Western Prediabetes: An Ethnicity-specific Meta-analysis. *Clinical therapeutics.* 2015;37(8):1798-1812.
24. Affairs DoV. VA Informatics and Computing Infrastructure (VINCI). 2015; http://www.hsrd.research.va.gov/for_researchers/vinci/. Accessed February 27, 2020.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

25. Casey J, Schwartz B, Stewart W, Adler N. Using Electronic Health Records for Population Health Research: A Review of Methods and Applications. *Annual review of public health.* 2016;37:61-81.
26. ArcGIS: The mapping and analytics platform. https://www.esri.com/en-us/arcgis/about-arcgis/overview.
27. Python. https://www.python.org/.
28. QGIS.
29. NHLBI N. Classification of Overweight and Obesity by BMI, Waist Circumference, and Associated Disease Risks. https://www.nhlbi.nih.gov/health/educational/lose_wt/BMI/bmi_dis.htm. Accessed February 27, 2020.
30. Soohoo M, Moradi H, Obi Y, Kovesdy CP, Kalantar-Zadeh K, Streja E. Serum triglycerides and mortality risk across stages of chronic kidney disease in 2 million U.S. veterans. *Journal of clinical lipidology.* 2019;13(5):744-753.e715.
31. Burg MM, Brandt C, Buta E, et al. Risk for Incident Hypertension Associated With Posttraumatic Stress Disorder in Military Veterans and the Effect of Posttraumatic Stress Disorder Treatment. *Psychosomatic medicine.* 2017;79(2):181-188.
32. Peng M, Chen G, Kaplan GG, et al. Methods of defining hypertension in electronic medical records: validation against national survey data. *Journal of public health (Oxford, England).* 2016;38(3):e392-e399.
33. Rowan CG, Turner JR, Shah A, Spaeder JA. Antihypertensive treatment and blood pressure control relative to hypertension treatment guidelines. *Pharmacoepidemiology and drug safety.* 2014;23(12):1294-1302.
34. Armstrong C. JNC8 guidelines for the management of hypertension in adults. *American family physician.* 2014;90(7):503-504.
35. Hernandez S, Sylling P, Mor M, et al. Developing an Algorithm for Combining Race and Ethnicity Data Sources in the Veterans Health Administration. *Military medicine.* 2019.
36. Gillett M. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes care.* 2009;32(7):1327-1334.
37. Alemi F, Levy C, Kheirbek R. The Multimorbidity Index: A Tool for Assessing the Prognosis of Patients from Their History of Illness. *EGEMS (Wash DC).* 2016;4(1):1235-1235.

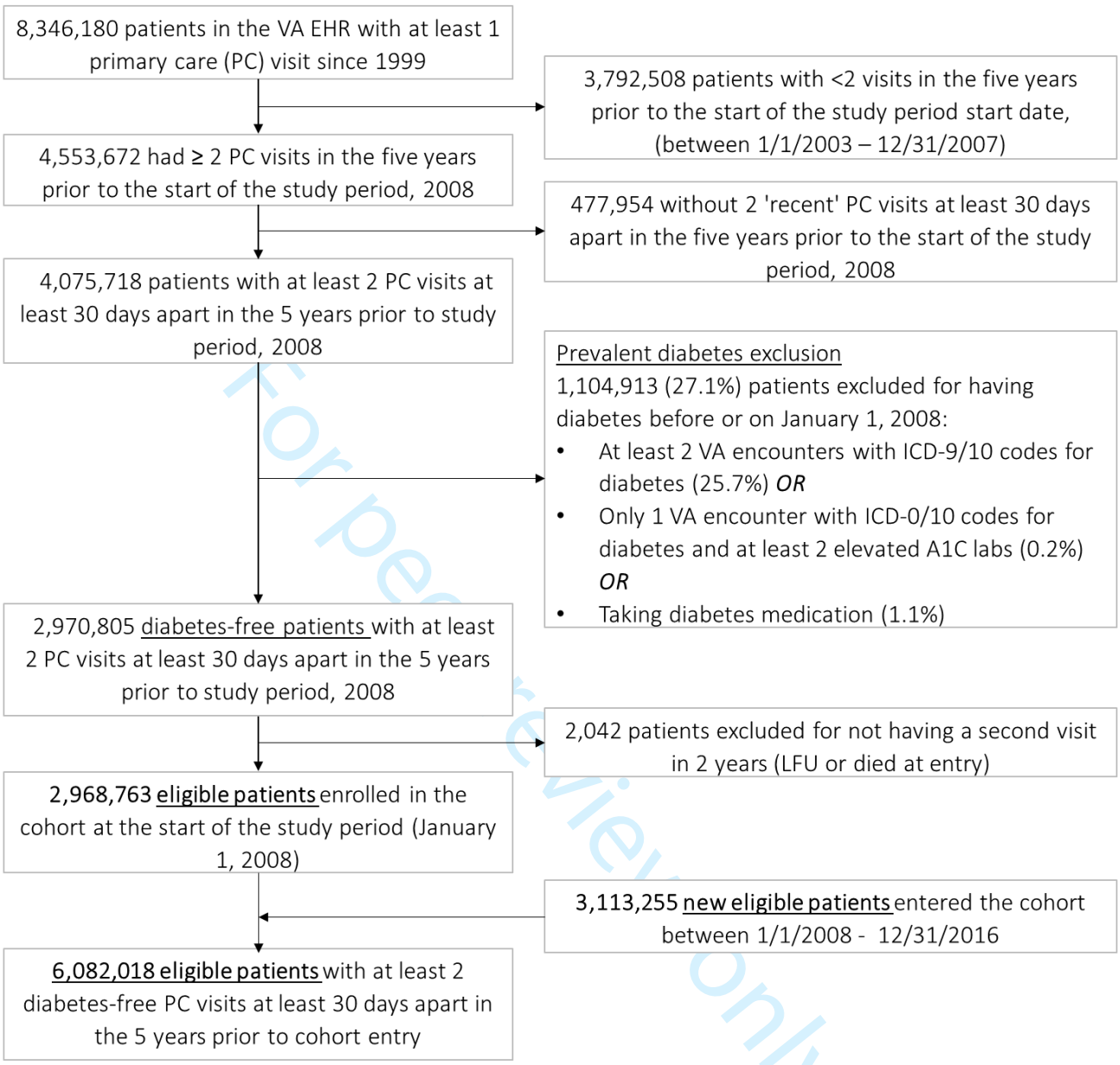
For peer review only

Figure 1. Cohort Flow Diagram of Diabetes-Free Cohort of US Veterans, 2008–2016

Figure 2. Cohort Trends, with cumulative numbers and percentage of patients, 2008 through 2018

Figure 3. Geographic distribution of VADR Cohort

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml



* Patients with unreliable information on date of birth and date of death and patients with year of birth <1900 were excluded (n=2,248)

For peer review only

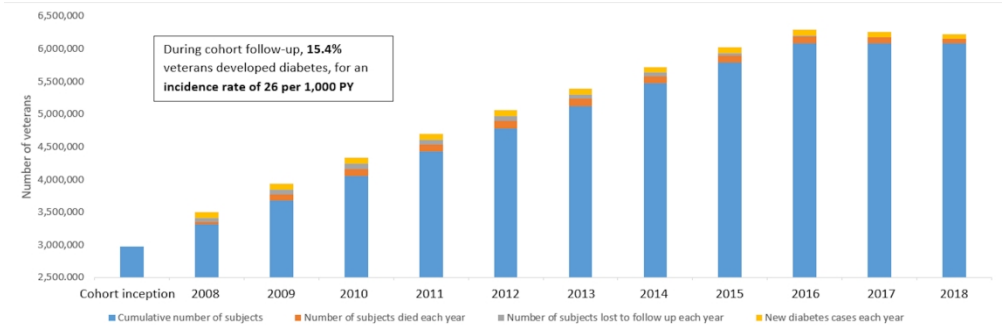


Figure 2. Cohort Trends, with cumulative numbers and percentage of patients, 2008 through 2018

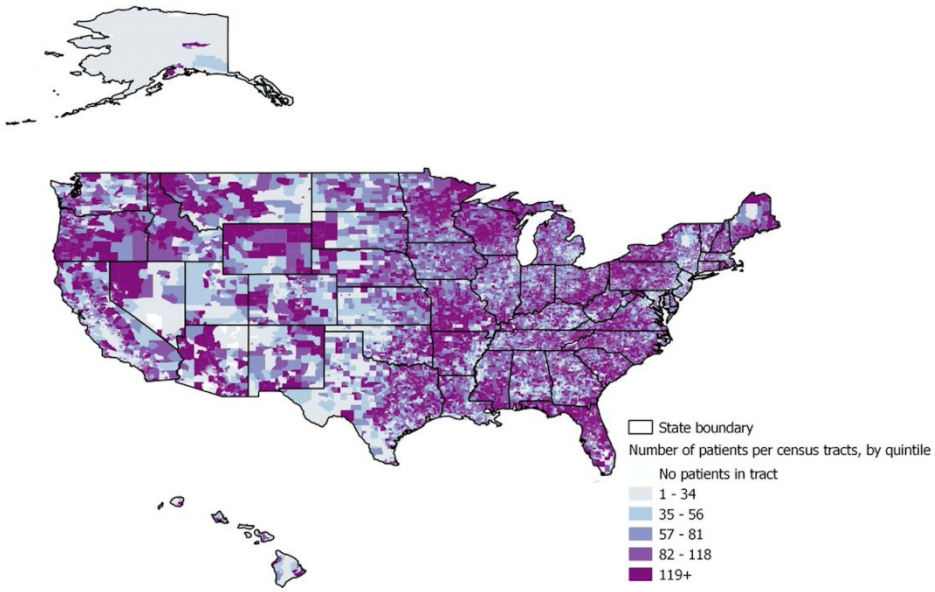


Figure 3. Geographic distribution of VADR Cohort

179x127mm (300 x 300 DPI)

Appendix. Defining chronic conditions in the VA environment

		Details				
Component	**Definition**	**ICD codes**		**Labs**	**Medications***	**Other**
		ICD-9	ICD-10			
Type 2 Diabetes Mellitus[1-4]	(1) At least two encounters (inpatient or outpatient) with documentation of a Type 2 diabetes ICD-9/10 code (ICD-9: 250.x; ICD-10: E11.x), *or* (2) a documented prescription for a diabetes medication other than metformin or acarbose alone, *or* (3) at least one encounter with a diabetes ICD-9/10 code and two elevated (≥ 6.5%) glycosylated hemoglobin (Hgb A1C) lab test results • Implausible A1C labs removed (range based on NHANES reporting) • Multiple labs measured on the same day or on the same day and time were averaged if they ranged within 1% • A1C labs ranging >1% were removed • If only one A1C lab was available prior to cohort entry date, that lab was used. • If more than one A1C lab was available, the average of the last two was taken	250, 250.0, 250.00, 250.02, 250.1, 250.10, 250.12, 250.2, 250.20, 250.22, 250.3, 250.30, 250.32, 250.4, 250.40, 250.42, 250.5, 250.50, 250.52, 250.6, 250.60, 250.62, 250.7, 250.70, 250.72, 250.8, 250.80, 250.82, 250.9, 250.90, 250.92	E11.00, E11.01, E11.21, E11.22, E11.29, E11.311, E11.319, E11.321, E11.3211, E11.3212, E11.3213, E11.3219, E11.329, E11.3291, E11.3292, E11.3293, E11.3299, E11.331, E11.3311, E11.3312, E11.3313, E11.3319, E11.339, E11.3391, E11.3392, E11.3393, E11.3399, E11.341, E11.3411, E11.3412, E11.3413, E11.3419, E11.349, E11.3491, E11.3492, E11.3493, E11.3499, E11.351, E11.3511, E11.3512, E11.3513, E11.3519, E11.352, E11.3521, E11.3522, E11.3523, E11.3529, E11.353, E11.3531, E11.3532, E11.3533, E11.3539, E11.354, E11.3541, E11.3542, E11.3543, E11.3549, E11.355, E11.3551, E11.3552, E11.3553, E11.3559, E11.359, E11.3591, E11.3592, E11.3593, E11.3599, E11.36, E11.37, E11.37X1, E11.37X2, E11.37X3, E11.37X9, E11.39, E11.40, E11.41, E11.42, E11.43, E11.44,	LOINC code corresponding to A1C: • 17855-8 17856-6 4548-4 4549-2	Chloropropamide, glipizide, glyburide, glimepiride, metformin, repaglinide, nateglinide, tosiglitazone, pioglitazone, sitagliptin, saxagliptin, linagliptin, alogliptin, canagliflozin, dapagliflozin, acarbose, meglitol, colesevelam, insulin	-

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

			E11.49, E11.51, E11.52, E11.59, E11.610, E11.618, E11.620, E11.621, E11.622, E11.628, E11.630, E11.638, E11.641, E11.649, E11.65, E11.69, E11.8, E11.9			
Comorbidities						
A1C[4-10]	(1) Mean and standard deviation calculated after identifying labs using LOINC codes • Implausible A1C labs removed (range based on NHANES reporting) • Multiple labs measured on the same day or on the same day and time were averaged if they ranged within 1% • A1C labs ranging >1% were removed • If only one A1C lab was available prior to cohort entry date, that lab was used. • If more than one A1C lab was available, the average of the last two was taken	-	-	LOINC code corresponding to A1C: • 17855-8 • 17856-6 • 4548-4 • 4549-2	-	
BMI[11-14]	(1) Calculated as weight (Kg)/[height(m)]^2. Normal weight defined as BMI < 25, Overweight defined as ≥25 and <30, and Obese as ≥30 Height: • Implausible values removed (range based on published literature)	-	-	-	-	Obtained by vital signs records

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

	• Multiple heights recorded during the same visit were averaged if they ranged within 3 inches (7.62 cm) or less • Measurements ranging more than 3 inches were deleted • If only one height measurement was available prior to cohort entry date, that height was used • If more than one height measurement was available prior to cohort entry date, the average of the last two was taken <u>Weight:</u> • Implausible values removed (range based on published literature) • Multiple weights recorded during the same visit were averaged if they ranged within 10 lb (4.536 Kg) or less. Measurements ranging more than 10 lb were deleted • If only one weight measurement was available prior to cohort entry date, that weight was used • If more than one weight measurement was available prior to cohort entry date, the average of the last two was taken					
Blood Pressure (BP)[15-19]	(1) Mean and standard deviation of systolic and diastolic BP calculated	-	-	-	-	Obtained by vital signs records

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

	• Records deleted if measured at nighttime (8pm to 7am) or if diastolic BP was greater than systolic BP • Only BP measured within the 2 years prior to cohort entry was included • BP measured on the same day was averaged • If only one BP measured was available on, or prior to, cohort entry date, it was used as the baseline BP • If more than one BP was available on, or prior to, cohort entry date, the average of the last two measurements was used as the baseline BP					
Hypertension (HTN)[20-24]	(1) At least one ICD-9/10 code for HTN	401.0, 401.1, 401.9	I10.X	-	-	-
Hyperlipidemia[25,26]	(1) Elevated total cholesterol (>240 mg/dL), *or* (2) Lipid-lowering medication use, *or* (3) at least 2 ICD-9/10 codes documenting hyperlipidemia	272.0, 272.1, 272.2, 272.3, 272.4, 272.5, 272.6, 272.7, 272.8, 272.9	E78, E78.0, E78.00, E78.01, E78.1, E78.2, E78.3, E78.4, E78.41, E78.49, E78.5, E78.6, E78.7, E78.70, E78.71, E78.72, E78.79, E78.8, E78.81, E78.89, E78.9	LOINC code corresponding to Total cholesterol: • 2093-3 • 14647-2	Generic names for class "CV350": • Atorvastatin • Cholestryamin • Colestipol • Ezetimibe • Ezetimibe/Simv-astatin • Gemfibrozil • Lomitapide • Mipomerson • Pravastatin • Rosuvastatin • Simvastatin	-

For peer review only

Lipids (HDL, LDL, Triglycerides)[25-27]	(1) Mean and standard deviation calculated after identifying labs using LOINC codes • Implausible values removed (range based on NHANES reporting) • The median of different lab values completed at the same day was taken • Lab values that are 50% > or < the median were considered outliers and removed • The median of the rest of the labs was taken as the lab value for that day • For all, if only one lab was available prior to the cohort entry date, that lab was used. If more than 1 lab value was available, the average of the last two was taken	-	-	LOINC codes corresponding to: <u>HDL:</u> • 2085-9 • 18263-4 • 9832-7 <u>LDL:</u> • 13457-7 • 18262-6 • 2089-1 • 14155-6 <u>Triglycerides:</u> • 2571-8 • 14927-8 • 12228-3 • 3049-4 • 1644-4 • 12951-0 • 3048-6	-	-
Chronic Kidney Disease (CKD)[24,28,29]	(1) At least one ICD-9/10 code documenting CKD	585.5X	N18.X	-	-	-
Ischemic Heart Disease[30]	(1) At least one ICD-9/10 code documenting ischemic heart disease	410.X, 411.X, 412.X, 413.X, 414.X	I20.X, I21.X, 122.X, 123.X, 124.X, 125.X	-	-	-
Heart Failure[30,31]	(1) At least one ICD-9/10 code documenting heart failure	428.X	I50.X	-	-	-
Peripheral Vascular Disease (PVD)[30,31]	(1) At least one ICD-9/10 code documenting PVD	440.0, 440.1, 440.2, 440.20, 440.21, 440.22, 440.23, 440.29, 440.4, 440.8, 440.9, 443.9, 557.0, 557.1, 557.9	I73.X	-	-	-
Stroke	(1) At least one ICD-9/10 code documenting stroke	346.60, 346.61, 346.62, 346.63, 432.0, 432.1, 432.9, 433.01, 433.11, 433.21, 433.31, 433.81, 433.91, 434.0, 434.00, 434.01, 434.1, 434.10,	I63.X	-	-	-

For peer review only

		434.11, 434.9, 434.90, 434.91, 436.5, 430.X, 431.X				
Agent Orange	(1) "Agent orange" flag was used to generate the number of people with this exposure	-	-	-	-	From patient problem lists
Hepatitis C	(1) At least one ICD-9/10 code documenting Hepatitis C	070.54	B18.2	-	-	-
Hyperuricemia	(1) At least one ICD-9/10 code documenting Hyperuricemia	790.6	E79.0	-	-	-
Polycystic Ovary Syndrome	(1) At least one ICD-9/10 code documenting Polycystic Ovary Syndrome	256.4	E28.2	-	-	-
Gestational Diabetes	(1) At least one ICD-9/10 code documenting Gestational Diabetes	V12.21, 648.83	Z86.32, 024.4X	-	-	-
eGFR[32,33]	(1) Mean and standard deviation calculated after identifying labs using LOINC codes • Implausible eGFR values were excluded • The median of eGFR measured on the same day was taken, then values more than 50% different than the median were excluded as outliers • The median of the rest of the measures was taken as eGFR for that day • If only one eGFR was available prior to cohort entry date, it was taken as the baseline estimate • If more than one eGFR was available prior to cohort entry date, the average of the most recent two was taken as the baseline estimate	-	-	LOINC code corresponding to <u>eGFR:</u> • 62238-1 • 48643-1 • 33914-3	-	-

For peer review only

Liver enzymes	(1) Mean and standard deviation calculated after identifying labs using LOINC codes • Implausible values removed (range based on NHANES reporting) • The median of different lab values completed at the same day was taken • Lab values that are 50% > or < the median were considered outliers and removed • The median of the rest of the labs was taken as the lab value for that day For all, if only one lab was available prior to the cohort entry date, that lab was used. If more than 1 lab value was available, the average of the last two was taken			LOINC codes corresponding to <u>AST:</u> • 14409-7 • 14410-5 • 43822-6 • 88112-8 • 14412-1 • 14414-7 • 16412-9 • 1918-2 • 27344-1 • 14413-9 • 1917-4 • 1919-0 • 1920-8 • 30239-8 • 14411-3 • 44786-2 <u>ALT:</u> • 1741-8 • 25302-1 • 54491-6 • 1742-6 • 1743-4 • 44785-4 • 16324-6 • 50168-4 • 76625-3 • 1744-2 • 54492-4 • 77144-4		Exclusions: - Patients with hepatitis B and C (ICD-9: 070.2x, 070.3x, 070.41, 070.44, 070.51, 070.51, 070.54, 070.7x. ICD-10: B18.x) - alcohol abuse (ICD-9: 291.x, 303.0x, 303.9x, 305.0x. ICD-10 F10.x) - patients with other rare liver disease (ICD-9: 576.1, 275.03, 275.01, 275.1, 237.4, 571.42, 571.6, 275.09. ICD-10: 237.4, D44.0, D44.2, D44.9, 275.01, E83.110, 275.03, E83.118, E83.119, 275.09, E83.10, E83.19, 275.1, E83.00, E83.01, E83.09, 571.42, K75.4, 571.6, K74.3, K74.4, K74.5, 576.1, K83.0

For peer review only

	At least 2 elevated ALT (≥40 U/L) at least 6 months apart within 2 years					Exclusions: - Patients with hepatitis B and C (ICD-9: 070.2x, 070.3x, 070.41, 070.44, 070.51, 070.51, 070.54, 070.7x. ICD-10: B18.x) - alcohol abuse (ICD-9: 291.x, 303.0x, 303.9x, 305.0x. ICD-10 F10.x) - patients with other rare liver disease (ICD-9: 576.1, 275.03, 275.01, 275.1, 237.4, 571.42, 571.6, 275.09. ICD-10: 237.4, D44.0, D44.2, D44.9, 275.01, E83.110, 275.03, E83.118, E83.119, 275.09, E83.10, E83.19, 275.1, E83.00, E83.01, E83.09, 571.42, K75.4, 571.6, K74.3, K74.4, K74.5, 576.1, K83.0
Fatty liver disease[34]						

* Medications are documented prescriptions, not prescriptions filled

For peer review only

References

1. Jones CD, Greenwood RH, Misra A, Bachmann MO. Incidence and progression of diabetic retinopathy during 17 years of a population-based screening program in England. *Diabetes care.* 2012;35(3):592-596.
2. Parikh SV, Saya S, Divanji P, et al. Risk of death and myocardial infarction in patients with peripheral arterial disease undergoing percutaneous coronary intervention (from the National Heart, Lung and Blood Institute Dynamic Registry). *The American journal of cardiology.* 2011;107(7):959-964.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

3. Rassen JA, Bartels DB, Schneeweiss S, Patrick AR, Murk W. Measuring prevalence and incidence of chronic conditions in claims and electronic health record databases. *Clinical epidemiology.* 2019;11:1-15.
4. Carls G, Huynh J, Tuttle E, Yee J, Edelman SV. Achievement of Glycated Hemoglobin Goals in the US Remains Unchanged Through 2014. *Diabetes Ther.* 2017;8(4):863-873.
5. Pani LN, Korenda L, Meigs JB, et al. Effect of aging on A1C levels in individuals without diabetes: evidence from the Framingham Offspring Study and the National Health and Nutrition Examination Survey 2001-2004. *Diabetes care.* 2008;31(10):1991-1996.
6. Nicolo M, Boullata JI. Serum 25OHD concentration as a predictor of haemoglobin A1c among adults living in the USA: NHANES 2003 to 2010. *BMJ Nutrition, Prevention & Health.* 2019:bmjnph-2019-000029.
7. Davis PJ, Liu M, Sherman S, et al. HbA1c, lipid profiles and risk of incident type 2 Diabetes in United States Veterans. *PloS one.* 2018;13(9):e0203484.
8. Gillett M. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes care.* 2009;32(7):1327-1334.
9. *Laboratory Procedure Manual: Glycohemoglobin*
10. Heinemann LF, G. Quality of HbA1c measurement in the practice: The german perspective. *J Diabetes Science and Technology.* 2015;9(3):687-695.
11. Fryar CK-M, D; Gu, Q; Ogden, CL. Mean body weight, height, waist circumference, and body mass index among adults: United States, 1999-2000 through 2015-2016. *National Health Statistics Reports.* 2018;122.
12. McDowell M, Hughes J, Borrud LG. Health characteristics of U.S. adults by body mass index category: results from NHANES 1999-2002. *Public Health Rep.* 2006;121(1):67-73.
13. Martinez K, Tucker L, Bailey B, LeCheminant J. Expanded Normal Weight Obesity and Insulin Resistance in US Adults of the National Health and Nutrition Examination Survey. *J Diabetes Res.* 2017;2017:9502643-9502643.
14. Tarlov E, Silva A, Wing C, et al. Neighborhood Walkability and BMI Change: A National Study of Veterans in Large Urban Areas. *Obesity (Silver Spring, Md).* 2020;28(1):46-54.
15. Armstrong C. JNC8 guidelines for the management of hypertension in adults. *American family physician.* 2014;90(7):503-504.
16. Rowan CG, Turner JR, Shah A, Spaeder JA. Antihypertensive treatment and blood pressure control relative to hypertension treatment guidelines. *Pharmacoepidemiology and drug safety.* 2014;23(12):1294-1302.
17. Freitag MH, Vasan RS. What is normal blood pressure? *Current opinion in nephrology and hypertension.* 2003;12(3):285-292.
18. O'Brien E, Parati G, Stergiou G. Ambulatory blood pressure measurement: what is the international consensus? *Hypertension (Dallas, Tex : 1979).* 2013;62(6):988-994.
19. Agarwal R, Light RP. The effect of measuring ambulatory blood pressure on nighttime sleep and daytime activity--implications for dipping. *Clin J Am Soc Nephrol.* 2010;5(2):281-285.
20. Burg MM, Brandt C, Buta E, et al. Risk for Incident Hypertension Associated With Posttraumatic Stress Disorder in Military Veterans and the Effect of Posttraumatic Stress Disorder Treatment. *Psychosomatic medicine.* 2017;79(2):181-188.
21. Fletcher RD, Jones R, Moore H, et al. INCREASED BODY MASS INDEX (BMI) PROMOTES HYPERTENSION AND WORSENS BLOOD PRESSURE CONTROL AMONG US VETERANS. *Journal of the American College of Cardiology.* 2018;71(11 Supplement):A1817.
22. Peng M, Chen G, Kaplan GG, et al. Methods of defining hypertension in electronic medical records: validation against national survey data. *Journal of public health (Oxford, England).* 2016;38(3):e392-e399.
23. Banerjee D, Chung S, Wong EC, Wang EJ, Stafford RS, Palaniappan LP. Underdiagnosis of hypertension using electronic health records. *American journal of hypertension.* 2012;25(1):97-102.
24. *Hypertension prevalence estimator tool.*
25. Soohoo M, Moradi H, Obi Y, Kovesdy CP, Kalantar-Zadeh K, Streja E. Serum triglycerides and mortality risk across stages of chronic kidney disease in 2 million U.S. veterans. *Journal of clinical lipidology.* 2019;13(5):744-753.e715.
26. Laclaustra M, Stranges S, Navas-Acien A, Ordovas JM, Guallar E. Serum selenium and serum lipids in US adults: National Health and Nutrition Examination Survey (NHANES) 2003-2004. *Atherosclerosis.* 2010;210(2):643-648.
27. Toth PP, Potter D, Ming EE. Prevalence of lipid abnormalities in the United States: the National Health and Nutrition Examination Survey 2003-2006. *Journal of clinical lipidology.* 2012;6(4):325-330.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

28. Fleet JL, Dixon SN, Shariff SZ, et al. Detecting chronic kidney disease in population-based administrative databases using an algorithm of hospital encounter and physician claim codes. *BMC Nephrology.* 2013;14(1):81.
29. Prevention CfDCa. *Prevalence of patients with an ICD-9-CM or ICD-10-CM code indicating CKD.*
30. Quan HS, V.; Halfon, P.; et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical care.* 2005;43(11):1130-1139.
31. Lix LS, M.; Piz, M.; Ahmed, R.; Quon, H.; Griffith, J., et al. *Cancer data linkage in manitoba: Expanding the instrastructure for research.* 2016.
32. Grams ME, Juraschek SP, Selvin E, et al. Trends in the prevalence of reduced GFR in the United States: a comparison of creatinine- and cystatin C-based estimates. *Am J Kidney Dis.* 2013;62(2):253-260.
33. Hemmelgarn BR, Manns BJ, Lloyd A, et al. Relation Between Kidney Function, Proteinuria, and Adverse Outcomes. *JAMA.* 2010;303(5):423-429.
34. Husain N, Blais P, Kramer J, Kowalkowski M, Richardson P, El-Serag HB, Kanwal F. Nonalcoholic fatty liver disease (NAFLD) in the Veterans Administration population: development and validation of an algorithm for NAFLD using automated data. Alimentary pharmacology & therapeutics. 2014 Oct;40(8):949-54.

For peer review only

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	2
Objectives	3	State specific objectives, including any prespecified hypotheses	2
Methods			
Study design	4	Present key elements of study design early in the paper	2,3
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	3
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	3
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	n/a
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	3,4
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	2
Bias	9	Describe any efforts to address potential sources of bias	n/a
Study size	10	Explain how the study size was arrived at	n/a
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	4
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	n/a
		(*b*) Describe any methods used to examine subgroups and interactions	n/a
		(*c*) Explain how missing data were addressed	n/a
		(*d*) If applicable, explain how loss to follow-up was addressed	3
		(*e*) Describe any sensitivity analyses	n/a
Results			

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	3
		(b) Give reasons for non-participation at each stage	n/a
		(c) Consider use of a flow diagram	Figure1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	4,5 Table 1
		(b) Indicate number of participants with missing data for each variable of interest	n/a
		(c) Summarise follow-up time (eg, average and total amount)	3
Outcome data	15*	Report numbers of outcome events or summary measures over time	n/a

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included (*b*) Report category boundaries when continuous variables were categorized (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n/a
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	4/5
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	5
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	5
Generalisability	21	Discuss the generalisability (external validity) of the study results	5
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	6

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

For peer review only