


BMJ Open US veterans administration diabetes risk (VADR) national cohort: cohort profile

Sanja Avramovic ^{1,2}, Farrokh Alemi,¹ Rania Kanchi,³ Priscilla M Lopez,³ Richard B Hayes,³ Lorna E Thorpe,³ Mark D Schwartz^{2,3}

To cite: Avramovic S, Alemi F, Kanchi R, *et al.* US veterans administration diabetes risk (VADR) national cohort: cohort profile. *BMJ Open* 2020;**10**:e039489. doi:10.1136/bmjopen-2020-039489

► Prepublication history and additional materials or this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-039489>).

Received 07 May 2020

Revised 06 November 2020

Accepted 12 November 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Health Administration and Policy, George Mason University, Fairfax, Virginia, USA

²VA New York Harbor Healthcare System, New York, New York, USA

³Department of Population Health, New York University School of Medicine, New York, New York, USA

Correspondence to

Sanja Avramovic;
savramov@gmu.edu

ABSTRACT

Purpose The veterans administration diabetes risk (VADR) cohort facilitates studies on temporal and geographic patterns of pre-diabetes 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 VADR cohort includes all 6 082 018 veterans in the USA enrolled in the veteran administration (VA) for primary care who were diabetes-free as of 1 January 2008 and who had at least two diabetes-free visits to a VA primary care service at least 30 days apart within any 5-year period since 1 January 2003, or veterans subsequently enrolled and were diabetes-free at cohort entry through 31 December 2016. Cohort subjects were followed from the date of cohort entry until censor defined as date of incident diabetes, loss to follow-up of 2 years, death or until 31 December 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 (PY)) was 26 per 1000 PY. 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 neighbourhood socioeconomic status via geospatial linkage to residence address information.

INTRODUCTION

Diabetes mellitus (diabetes) is a chronic disease that affects 34.2 million (10.5%) of adults and children in the USA.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

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 veterans administration (VA) healthcare system is the nation's largest integrated healthcare system, in which veterans are followed across all VA facilities and in-system providers.
- Data are restricted to that which is collected in electronic health records 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, findings may not generalise to minorities or to women.

estimated to have pre-diabetes 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 pre-diabetes. These high rates compared with the general population may be due to the increased proportions of overweight (37%) and obesity (41%) among veterans,⁹ their older age, lower socioeconomic status¹⁰ and possible exposure to herbicides such as Agent Orange.¹¹

Behavioural 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 programmes.^{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 since 1 January 2008 who were

diabetes-free at enrollment. The cohort was developed as a part of the Diabetes Location, Environmental Attributes and Disparities 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 pre-diabetes 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 USA. 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 1 January 2008 through 31 December 2016, and followed from cohort entry through 31 December 2018. VA primary care clinics operate in 170 VA Medical Centers and in more than 1000 Community-Based Outpatient Clinics across the USA.¹⁹ As a dynamic cohort, subject follow-up is ongoing, but this paper reports on the cohort from 1 January 2008 through 31 December 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 ($\geq 6.5\%$) glycosylated haemoglobin (Hgb A1C) laboratory test results (see online supplemental appendix 1 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 pre-diabetes; including them may lead to misclassifying cases of pre-diabetes 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 two diabetes-free visits to a VA primary care service, occurring at least 30 days apart, from 1 January 2003 to 31 December 2016. Cohort entry (baseline) was defined as either 1 January 2008 or the date of the second diabetes-free primary care visit for

subjects entering after 1 January 2008. Eligible subjects were allowed to enter the cohort through 31 December 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, hospitalisations 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 8346180 patients seen for at least one 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 5 years prior to the study period start date, 1 January 2008. Patients were excluded if they had fewer than two primary care visits, at least 30 days apart during that 5-year time period and less than two primary care visits after cohort entry. After excluding patients with prevalent diabetes, the initial diabetes-free cohort included 2968763 patients. Another 3113255 diabetes-free patients met the same eligibility criteria after the start of the study period and entered the cohort between 1 January 2008 and 31 December 2016, resulting in a diabetes-free cohort of 6082018 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.²⁴ The CDW contains data integrated from VA's electronic medical record (Veterans Health Information Systems and Technology Architecture), including all administrative data (eg, all dates of encounters and diagnostic codes for outpatient and inpatient care), patient demographic characteristics, clinical data (eg, vital signs, health factors, pharmacy, laboratory, radiological, clinical notes) and healthcare utilisation 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

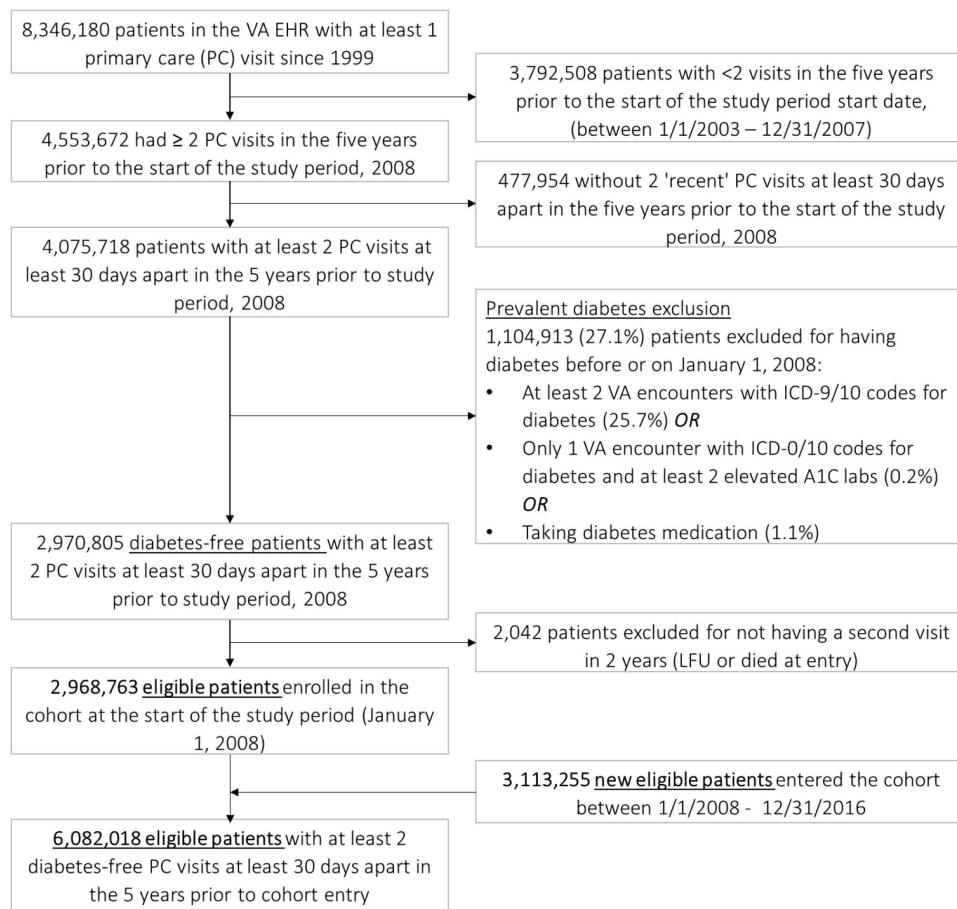


Figure 1 Cohort flow diagram of diabetes-free cohort of US veterans, 2008–2016. EHR, electronic health record; LFU, lost to follow-up; VA, veterans administration.

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 and race/ethnicity. First address on file per patient in cohort was exported out of the VINCI environment, geocoded using ArcGIS²⁶ and Python,²⁷ and mapped to show number of patients in the cohort per census tract using QGIS.²⁸

Glycaemia and body weight are important predictors of diabetes. We measured Hgb A1c as a continuous value, and classified as normal (<5.7%), pre-diabetes (5.7% to 6.4%) or diabetes (≥6.5%). We measured weight in pounds and body mass index (BMI), defined as (weight in kilograms)/(height in metres).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, hyperuricaemia, 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. Hyperlipidaemia was defined as at least two encounters with ICD codes for hyperlipidaemia, 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 two consecutive elevated blood pressure (BP) within the last 2 years prior to cohort entry.^{31 32} Elevated BP was included 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 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 Besides this select list, all documented diagnoses and treatments are available for the cohort.

Access to cohort data

Access to VA EHRs data is limited to researchers with active, VA appointments and an IRB-approved protocol. Once a researcher has a VA appointment and IRB

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

	All veterans*	Veterans without incident diabetes†	Veterans with incident diabetes	Not measured or missing
	N (%) or mean (SD)	N (%) or mean (SD)	N (%) or mean (SD)	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	745 511 (12.3)	592 256 (16.7)	21 788 (2.3)	
35–49	1 009 677 (16.6)	725 427 (20.4)	130 345 (13.9)	
50–64	2 114 275 (34.8)	1 280 132 (36.1)	451 964 (48.3)	
65–79	1 499 787 (24.7)	786 213 (22.2)	261 592 (27.9)	
80+	712 681 (11.7)	165 264 (4.7)	70 900 (7.6)	
Gender				124 (0)
Male	5 577 892 (91.7)	3 167 502 (89.2)	886 763 (94.7)	
Female	504 002 (8.3)	381 768 (10.8)	49 818 (5.3)	
Race ethnicity				612 210 (10.07)
Non-Hispanic white	4 092 942 (74.8)	2 412 225 (73.8)	614 511 (70.8)	
Non-Hispanic black	889 465 (16.3)	544 434 (16.7)	173 057 (19.9)	
Hispanic	331 817 (6.1)	211 459 (6.5)	55 301 (6.4)	
Non-Hispanic Asian	55 564 (1)	37 545 (1.2)	7 705 (0.9)	
NH-Native Hawaiian or other Pacific Islander	50 426 (0.9)	30 228 (0.9)	9 020 (1)	
NH-American Indian or Alaska Native	49 594 (0.9)	31 107 (1)	8 474 (1)	
Marital status				453 512 (7.46)
Married or living with a partner	3 104 312 (55.2)	1 832 844 (55.2)	477 337 (56.2)	
Single	2 524 194 (44.9)	1 488 661 (44.8)	372 630 (43.8)	
Clinical characteristics				
HbA1c	5.7 (0.6)	5.5 (0.4)	6.1 (0.9)	3 608 600 (59.33)
HbA1c categories				
Normal (<5.7%)	1 311 768 (53)	875 677 (62.3)	128 681 (25.8)	
Pre-diabetes (5.7%–6.49%)	1 027 336 (41.5)	514 758 (36.6)	268 105 (53.8)	
Diabetes range (≥6.5%)	134 314 (5.4)	15 302 (1.1)	101 746 (20.4)	
Weight in pounds	196.9 (40.7)	196.6 (38.9)	214.7 (45.1)	260 313 (4.28)
BMI	28.8 (5.4)	28.7 (5.1)	31.3 (6)	260 313 (4.28)
Underweight (<18.5)	48 954 (0.8)	18 042 (0.5)	4 417 (0.5)	
Normal weight (18.5–<25)	1 308 427 (22.5)	732 429 (21.6)	105 224 (11.7)	
Overweight (25–<30)	2 362 867 (40.6)	1 441 730 (42.5)	294 224 (32.6)	
Obese (≥30)	2 101 457 (36.1)	1 197 716 (35.3)	499 082 (55.3)	
Measured blood pressure‡				
Systolic blood pressure	130.4 (14.7)	129.3 (14.2)	133.2 (14.9)	290 882 (4.78)
Diastolic blood pressure	76.8 (10)	77.6 (9.6)	78 (10.1)	290 882 (4.78)
Elevated blood pressure (≥130/80)	3 516 559 (60.7)	2 011 621 (59.4)	609 072 (67.9)	290 882 (4.78)
Elevated blood pressure (≥140/90)	1 499 481 (25.9)	796 617 (23.5)	287 665 (32.1)	290 882 (4.78)
Hypertension				
≥1 International classification of disease (ICD) code or two consecutive elevated BP‡ (≥130/80)	2 032 490 (33.4)	2 137 040 (60.2)	760 631 (81.2)	
≥1 ICD code or two consecutive elevated BP‡ (≥140/90)	2 805 063 (46.1)	1 617 389 (45.6)	667 469 (71.3)	
Lipids				

Continued

Table 1 Continued

	All veterans*	Veterans without incident diabetes†	Veterans with incident diabetes	Not measured or missing
	N (%) or mean (SD)	N (%) or mean (SD)	N (%) or mean (SD)	N (%)
Total cholesterol	185.5 (38.4)	188.3 (37.7)	185.4 (40.2)	1 101 271 (18.11)
Triglyceride	140.8 (88.9)	137.6 (86.9)	168.1 (103.2)	1 118 227 (18.39)
Low-density lipoproteins (LDL)	112.6 (33.1)	115.3 (32.8)	111.4 (34.1)	1 151 308 (18.93)
High-density lipoproteins (HDL)	46.3 (14.2)	47.3 (14.3)	42.4 (12.5)	1 126 777 (18.53)
Hyperlipidaemia§	2 681 683 (44.1)	1 371 518 (38.6)	408 288 (43.6)	
Smoking status				3 853 822 (63.36)
Current smoker	948 272 (42.6)	561 968 (41.9)	140 333 (44.3)	
Not a smoker	1 279 924 (57.4)	780 134 (58.1)	176 777 (55.8)	
Estimated glomerular filtration rate (eGFR)	80.1 (18.3)	82.5 (16.7)	78.6 (18.4)	2 718 664 (44.7)
eGFR categories				
eGFR ≥90 (stage 1)	1 036 900 (30.8)	691 968 (33.9)	145 390 (27.9)	
eGFR ≥60 to <90 (stage 2)	1 899 314 (56.5)	1 180 774 (57.8)	300 183 (57.7)	
eGFR <60 (stages 3, 4 or 5)	427 140 (12.7)	169 856 (8.3)	75 168 (14.4)	
Chronic kidney disease (ICD codes)	150 823 (2.5)	47 353 (1.3)	29 236 (3.1)	
Ischaemic heart disease (ICD codes)	999 927 (16.4)	378 832 (10.7)	209 448 (22.4)	
Heart failure (ICD codes)	181 375 (3)	40 411 (1.1)	41 272 (4.4)	
Peripheral vascular disease (ICD codes)	256 054 (4.2)	77 339 (2.2)	55 242 (5.9)	
Stroke (ICD codes)	30 423 (0.5)	10 724 (0.3)	6 275 (0.7)	
Agent orange exposure (exposed)				
Chronic hepatitis C	102 534 (1.7)	51 789 (1.5)	21 382 (2.3)	
Hyperuricaemia	180 941 (3)	74 543 (2.1)	54 981 (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				
Aspartate aminotransferase (AST)	25.7 (14.3)	25.7 (14.2)	26.7 (14.7)	1 731 271 (28.47)
Alanine aminotransferase (ALT)	29.5 (19.8)	30.3 (19.9)	32.4 (21.1)	1 634 812 (26.88)
Elevated AST (>40 U/L)	277 603 (6.4)	156 503 (6.2)	56 466 (8.3)	1 731 271 (28.47)
Elevated ALT (>40 U/L)	778 931 (17.5)	482 342 (18.6)	158 747 (22.8)	1 634 812 (26.88)
Fatty liver disease	253 134 (10.4)	136 847 (10.8)	67 367 (15.7)	3 654 047 (60.08)

*Including those lost to follow-up and those died during the study period.

†Only patients who completed the follow-up and were diabetes free at the end of the study.

‡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.

§Hyperlipidaemia was defined as at least 2 ICD-9/10 codes for hyperlipidaemia, total cholesterol >240 mg/dL or lipid lowering medication use. BMI, body mass index; BP, blood pressure.

approval, the VA has a comprehensive data infrastructure to support secure and remote access to data via the VINCI platform. Additionally, deidentified data sets 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 PY for this national cohort with 6082018 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 pre-diabetes 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%

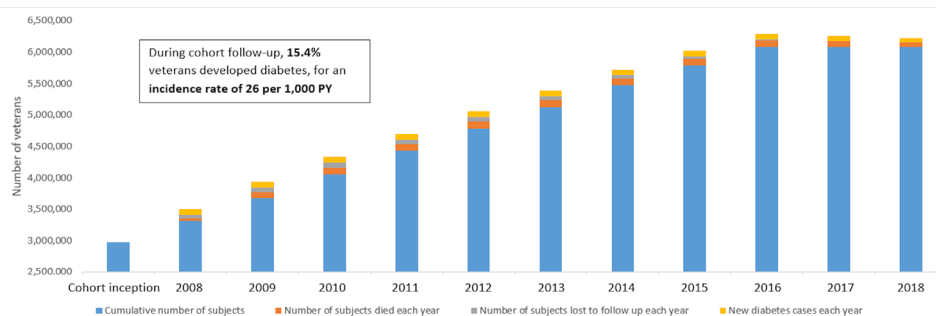


Figure 2 Cohort trends, with cumulative numbers and percentage of patients, 2008 through 2018. PY, person-years.

had hyperlipidaemia and 42.6% were smokers. Other clinical risk factors for diabetes included ischaemic 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 1 January 2008 through 31 December 2018. Almost half (48.8%) of the cohort entered at cohort inception in 1 January 2008, with the remainder entering during the study period through 31 December 2016. During cohort follow-up, 936 596 (15.4%) veterans developed diabetes, for an incidence rate of 26 per 1000 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 and 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 or

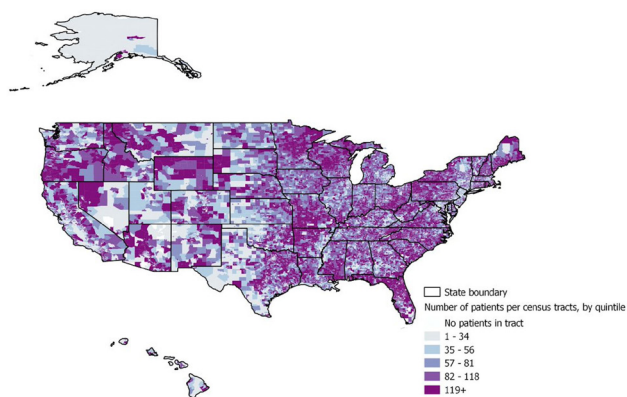


Figure 3 Geographic distribution of VADR cohort. VADR, veterans administration diabetes risk.

unmeasured 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 multimorbidity 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 healthcare outside of the VA will be included by merging the study cohort with data from the Centers for Medicare and Medicaid. Finally, home addresses are available and were geocoded in order to study the effect of community level characteristics 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 healthcare utilisers are more likely to be lost to follow-up or excluded, and higher utilisers 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 generalise 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 generalisability of our findings to non-veteran populations.

CONCLUSION

The VADR cohort is an important example of how large retrospective cohorts can be developed using EHRs, designed with methodologic and statistical approaches to increase generalisability 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 31 December 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 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 non-ignorable missing data.

Patient and public involvement

This cohort study was conducted without engagement or coproduction by patients or the public.

Twitter Farrokh Alemi @DrAlemi

Contributors SA 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: MS, LET, RH, SA and FA. Acquisition, analysis or interpretation of data: SA, PL, RK and MS. Drafting of the manuscript: SA, MS, RK, PL and LET. Critical revision of the manuscript for important intellectual content: all authors, SA, FA, RK, PL, RH, LET and MS. Statistical analysis: RK. All authors, SA, FA, RK, PL, RH, LET and MS, attest that they meet the full authorship criteria.

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

Map disclaimer The depiction of boundaries on this map does not imply the expression of any opinion whatsoever on the part of BMJ (or any member of its group) concerning the legal status of any country, territory, jurisdiction or area or of its authorities. This map is provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. 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 data sets 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.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines,

terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Sanja Avramovic <http://orcid.org/0000-0002-6155-9241>

REFERENCES

- 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.
- Centers for Disease Control and Prevention. *Underlying cause of death 1999-2008*. National Center for Health Statistics, 2018.
- American Diabetes Association. Economic costs of diabetes in the U.S. in 2012. *Diabetes Care* 2013;36:1033-46.
- Prevention CfDCA. Diabetes quick facts, 2019. Available: <https://www.cdc.gov/diabetes/basics/quick-facts.html> [Accessed 27 Feb 2020].
- Veterans Health Administration. *VA research on diabetes*, 2019.
- Liu Y, Sayam S, Shao X, et al. Prevalence of and trends in diabetes among Veterans, United States, 2005-2014. *Prev Chronic Dis* 2017;14:E135.
- Miller DR, Safford MM, Pogach LM. Who has diabetes? best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data. *Diabetes Care* 2004;27:B10-21.
- Veterans Health Administration. Close to 25 percent of VA patients have diabetes, 2011. Available: <http://www.va.gov/health/NewsFeatures/20111115a.asp> [Accessed 31 Dec 2015].
- Breland JY, Phibbs CS, Hoggatt KJ, et al. The obesity epidemic in the Veterans health administration: prevalence among key populations of women and men veterans. *J Gen Intern Med* 2017;32:11-17.
- Kondo K, Low A, Everson T, et al. Health disparities in veterans: a map of the evidence. *Med Care* 2017;55 Suppl 9 Suppl 2:S9-15.
- National Academies of Sciences E, and Medicine. *Veterans and agent orange*. National Academies Press, 2018.
- Prevention CfDCA. *National diabetes prevention program*, 2019.
- Diabetes Prevention Program Research Group, Knowler WC, Fowler SE, et al. 10-Year follow-up of diabetes incidence and weight loss in the diabetes prevention program outcomes study. *Lancet* 2009;374:1677-86.
- Ackermann RT, Finch EA, Brizendine E, et al. Translating the diabetes prevention program into the community. The DEPLOY pilot study. *Am J Prev Med* 2008;35:357-63.
- Ackermann RT, Marrero DG. Adapting the diabetes prevention program lifestyle intervention for delivery in the community. *Diabetes Educ* 2007;33:69-78.
- Baker MK, Simpson K, Lloyd B, et al. Behavioral strategies in diabetes prevention programs: a systematic review of randomized controlled trials. *Diabetes Res Clin Pract* 2011;91:1-12.
- Schellenberg ES, Dryden DM, Vandermeer B, et al. Lifestyle interventions for patients with and at risk for type 2 diabetes: a systematic review and meta-analysis. *Ann Intern Med* 2013;159:543-51.
- Diabetes lead (location, environmental attributes, and disparities) network. Available: <http://diabetesleadnetwork.org/> [Accessed 27 Feb 2020].
- Administration VH, About VHA. Available: <https://www.va.gov/health/aboutvha.asp> [Accessed 27 Feb 2020].
- Bowe B, Xie Y, Li T, et al. The 2016 global and national burden of diabetes mellitus attributable to PM_{2.5} air pollution. *Lancet Planet Health* 2018;2:e301-12.
- 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:e0203484.
- Tseng E, Yeh H-C, Maruthur NM. Metformin use in prediabetes among U.S. adults, 2005-2012. *Diabetes Care* 2017;40:887-93.
- Hu R, Li Y, Lv Q, et al. Acarbose monotherapy and type 2 diabetes prevention in eastern and Western prediabetes: an Ethnicity-specific meta-analysis. *Clin Ther* 2015;37:1798-812.



- 24 Affairs DoV. Va informatics and computing infrastructure (Vinci), 2015. Available: http://www.hsrd.research.va.gov/for_researchers/vinci/ [Accessed 27 Feb 2020].
- 25 Casey JA, Schwartz BS, Stewart WF, *et al.* Using electronic health records for population health research: a review of methods and applications. *Annu Rev Public Health* 2016;37:61–81.
- 26 ArcGIS: the mapping and analytics platform. Available: <https://www.esri.com/en-us/arcgis/about-arcgis/overview>
- 27 Python. Available: <https://www.python.org/>
- 28 Discover QGIS. Available: <https://qgis.org/en/site/about/index.html> [Accessed 30 Dec 2020].
- 29 NHLBI N. Classification of overweight and obesity by BMI, waist circumference, and associated disease risks. Available: https://www.nhlbi.nih.gov/health/educational/lose_wt/BMI/bmi_dis.htm [Accessed 27 Feb 2020].
- 30 Soohoo M, Moradi H, Obi Y, *et al.* Serum triglycerides and mortality risk across stages of chronic kidney disease in 2 million U.S. veterans. *J Clin Lipidol* 2019;13:744–53.
- 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. *Psychosom Med* 2017;79:181–8.
- 32 Peng M, Chen G, Kaplan GG, *et al.* Methods of defining hypertension in electronic medical records: validation against national survey data. *J Public Health* 2016;38:e392–9.
- 33 Rowan CG, Turner JR, Shah A, *et al.* Antihypertensive treatment and blood pressure control relative to hypertension treatment guidelines. *Pharmacoepidemiol Drug Saf* 2014;23:1294–302.
- 34 Armstrong C, Joint National Committee. JNC8 guidelines for the management of hypertension in adults. *Am Fam Physician* 2014;90:503–4.
- 35 Hernandez SE, Sylling PW, Mor MK, *et al.* Developing an algorithm for combining race and ethnicity data sources in the Veterans health administration. *Mil Med* 2020;185:e495–500.
- 36 Gillett M. International expert Committee report on the role of the A1c assay in the diagnosis of diabetes. *Diabetes Care* 2009;32:1327–34.
- 37 Alemi F, Levy CR, Kheirbek RE. The multimorbidity index: a tool for assessing the prognosis of patients from their history of illness. *EGEMS* 2016;4:1235