SUPPLEMENTAL MATERIALS

1. **Supplemental Methods**

2. **Supplemental Table 1**

3. **Supplemental References**

SUPPLEMENTAL METHODS

Identification of published classifiers for initial application

Atrial fibrillation (AFib)

A systematic review of 16 studies found that the ICD-9-CM code 427.3 was a commonly used definition of positive matches that performed relatively well by itself (sensitivity ranging between 71% and 95% for prevalent AFib).[1] This conclusion was supported by studies on the VHA population at large[2] and on a VA Boston cohort specifically.[3] In contrast, classifiers which also used cardioversion procedures or rhythm control medication were highly nonspecific. During chart review at design time, as well as during trial recruitment, we defined AFib using the 2014 American Heart Association definition.[4]

Breast cancer (BrCa)

For BrCa, Abraha systematically reviewed 21 studies, a majority of whom identified incident cases.[5] They described sensitivities of 53% to 99% for ICD-9-CM codes 174, 175, 233.0, V10.3, and ICD-10-CM codes C50 - C50.9, D05, Z85.3. This combination was persistently superior to more complex classifiers, which included surgical procedures, radiation therapy, or chemotherapy. During chart review and trial recruitment, we defined the three cancers studied here as the presence of a pathology report documenting the malignant histological appearance of a tumor.

Coronary artery disease (CAD)

A VA study by Floyd reported a CAD classifier that achieved 90% sensitivity and 89% specificity, compared to chart review during a 24-month window, using ICD-9-CM codes 410.x, 411.x, 412.x, 413.9, 414.x (except 414.1x), ICD-9-PCS codes 36.01-36.06, 36.09, 36.1x, 36.2x, and CPT codes 33510-33545, 92975, 92980-92984, 92995-92996.[6] For better harmonization with the CAD phenotype used in the development and validation of the CAD polygenic risk score implemented in the GenoVA Study,[7] we

additionally excluded stable angina-specific ICD-9-CM code 413.9 and added CPT codes 92973 (percutaneous transluminal coronary thrombectomy) and 92977 (coronary thrombolysis by intravenous infusion). During chart review and trial recruitment, we defined ST-elevation myocardial infarction[8] and unstable angina[9] as either meeting the definitions from the American Heart Association respective guidelines, or having a history of coronary syndrome-related interventions.

Colorectal cancer (CRCa)

A high-quality VHA-based comparative study assessed multiple CRCa classifiers within the same cohort.[10] We defined baseline history of CRCa first by querying the Oncology Domain, using ICD-O-3-equivalent site codes (67180, 67182-67189, 67199, 67209), reported to have 100% specificity and sensitivity. To account for potential lag in data entry, we also employed ICD-9-CM codes 153.0-153.4, 153.6-153.9, 154.0, and 154.1, found, in the same study, to have 100% sensitivity and 98% specificity for CRCa.

Prostate cancer (PrCa)

Radomski identified prevalent PrCa using one instance of ICD-9-CM codes 222.2x, 233.4x, 236.5x, 185.x, V10.46 in the inpatient diagnoses or outpatient diagnoses table, or one instance of prostatectomy-specific CPT codes 55810, 55812, 55815, 55801, 55821, 55831, 55842, 55845.[11]

Type 2 diabetes (T2D)

Raghavan identified incident T2D as one prescription for diabetes medication from VHA and non-VHA tables, or two ICD-9-CM code 250.x from the inpatient or outpatient diagnostic codes.[12]

During chart review and trial recruitment, we defined diabetes mellitus according to American Diabetes Association 2014 guidelines.[13]

SUPPLEMENTAL TABLE 1: VHA data tables used for disease classifiers[10,14,15]

Table	Notes
Demographics	
SPatient.SPatient	Provides name, sex at birth, Veteran status*, date of birth*, and date of death, if applicable
SPatient.SPatientAddress, SPatient.SPatientPhone	Provides contact information, whose absence, in rare cases, prevents enrollment in GenoVA
PatSub.PatientEthnicity, PatSub.PatientRace	Lists zero, one or more ethnicity and race qualifiers, as entered in VistA
*In certain situations, VHA providers may occasionally care for non-Veterans, such as non-Veteran employees requiring urgent care while on premises. The flag IsVeteran was checked to exclude records of such encounters.	
Clinical encounters and relationships	
PCMM.PatientProviders	Primary Care Management Model (PCMM) was the standard for storing patient-provider relationships in CDW. For the GenoVA Study trial, we extracted relationships where the field TeamPurpose was entered as 'PRIMARY CARE'. Additionally, we tested the fields RelationshipStartDate and RelationshipEndDate, to exclude outdated relationships.
RPCMM.CurrentPatientProviderRelationship	Reengineered Primary Care Management Model (RPCMM) superseded PCMM in 2018. For the GenoVA Study trial, we extracted relationships where the field RPCMMTeamCareType coded for either 'PRIMARY CARE' or 'PRIMARY CARE - HBPC', after conversion of the field RPCMMStaffRole to plain text using NDim.RPCMMStaffRole. Additionally, we tested the fields TeamPatientAssignStatus, RelationshipStartDateTime and RelationshipEndDateTime, to exclude outdated relationships.
Outpat.Visit**	Provides stop codes associated with an outpatient visit, after conversion of the field StopCodeSID to plain text, through the use of Dim.StopCode. We extracted visits which had primary care purposes per stop code. From these, we excluded encounters where the field AppointmentType had values such as 'RESEARCH', after converting appointment types to plain text through

	the use of the Dim.AppointmentType table.
Outpat.VProvider**	Augments Outpat.Visit by listing anonymized numeric codes for each provider involved in an outpatient encounter. We extracted visits which involved providers whose main role at VHA, per SStaff.SStaff, is that of a primary care provider.

**For the GenoVA Study trial, a primary care relationship in the VA Boston Healthcare System was defined as a current relationship listed in the PCMM or RPCMM tables, or a patient-provider relationship which led to an Outpat.Visit-listed encounter occurring during the previous 12 months and meeting at least one criterion for primary care visits described above.

Disease identification

Outpat_VDiagnosis, Inpat_InpatientDiagnosis	Lists ICD-9-CM or ICD-10-CM-coded procedures in inpatient and outpatient settings, respectively. Literal ICD-9-CM / ICD-10-CM codes were obtained through conversion, using the tables Dim.ICD9 or Dim.ICD10.
Inpat.InpatientICDProcedure	Lists ICD-9-PCS or ICD-10-PCS-coded procedures in inpatient settings, as used for the coronary artery disease classifier. Literal ICD-9-PCS / ICD-10-PCS codes were obtained through conversion, using the tables Dim.ICD9Procedure or Dim.ICD10Procedure.
Outpat.VProcedure, Inpat.InpatientCPTProcedure	Lists CPT-coded procedures in outpatient and inpatient settings, respectively, as used for the coronary artery disease and prostate cancer classifiers. Literal CPT codes were obtained through conversion using the table Dim.CPT.
Oncology.Oncology_Primary_165_5 and Oncology.Oncology_Patient_160	Combined, describe diagnoses abstracted by local cancer registrars for the purposes of the VA Central Cancer Registry (VACCR), with their respective ICD-O-3.
RxOut.RxOutpat, NonVAMed.NonVAMed	Lists prescriptions from VA and non-VHA providers, respectively. Also used in conjunction with Dim.LocalDrug, for the type 2 diabetes classifier.

Auxiliary

SStaff.SStaff, StaffSub.ProviderTypeAssignment	Combined, describe the care role of a VHA staff member. This information was used in extracting visits to primary care providers at VA Boston, as well as to pharmacy providers for the refined atrial fibrillation classifier (see main text).

SUPPLEMENTAL REFERENCES

1. Jensen, P. N. *et al.* A systematic review of validated methods for identifying atrial fibrillation using administrative data. *Pharmacoepidemiol. Drug Saf.* **21 Suppl 1**, 141–147 (2012).

2. Borzecki, A. M., Wong, A. T., Hickey, E. C., Ash, A. S. & Berlowitz, D. R. Identifying hypertension-related comorbidities from administrative data: what's the optimal approach? *Am. J. Med. Qual. Off. J. Am. Coll. Med. Qual.* **19**, 201–206 (2004).

3. Brophy, M. T. *et al.* Anticoagulant use for atrial fibrillation in the elderly. *J. Am. Geriatr. Soc.* **52**, 1151–1156 (2004).

4. January, C. T. *et al.* 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society. *J. Am. Coll. Cardiol.* **64**, e1-76 (2014).

5. Abraha, I. *et al.* Accuracy of administrative databases in detecting primary breast cancer diagnoses: a systematic review. *BMJ Open* **8**, e019264 (2018).

6. Floyd, J. S., Blondon, M., Moore, K. P., Boyko, E. J. & Smith, N. L. Validation of methods for assessing cardiovascular disease using electronic health data in a cohort of Veterans with diabetes. *Pharmacoepidemiol. Drug Saf.* **25**, 467–471 (2016).

7. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* **50**, 1219–1224 (2018).

8. O'Gara, P. T. *et al.* 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation* **127**, e362-425 (2013).

9. Amsterdam, E. A. *et al.* 2014 AHA/ACC Guideline for the Management of Patients with Non-ST-Elevation Acute Coronary Syndromes: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J. Am. Coll. Cardiol.* **64**, e139–e228 (2014).

10. Earles, A. *et al.* Structured Approach for Evaluating Strategies for Cancer Ascertainment Using

Large-Scale Electronic Health Record Data. *JCO Clin. Cancer Inform.* **2**, 1–12 (2018).

11.   Radomski, T. R. *et al.* Low-Value Prostate Cancer Screening Among Older Men Within the Veterans Health Administration. *J. Am. Geriatr. Soc.* **67**, 1922–1927 (2019).

12.   Raghavan, S. *et al.* Diabetes Mellitus–Related All-Cause and Cardiovascular Mortality in a National Cohort of Adults. *J. Am. Heart Assoc. Cardiovasc. Cerebrovasc. Dis.* **8**, (2019).

13.   American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* **37 Suppl 1**, S81-90 (2014).

14.   Culbreath, C. & Gonsoulin, M. VIReC Corporate Data Warehouse (CDW) Domain Descriptions. (2019).

15.   VA Information Resource Center. VIReC VHA Corporate Data Warehouse (CDW) Domain Layout. (2017).