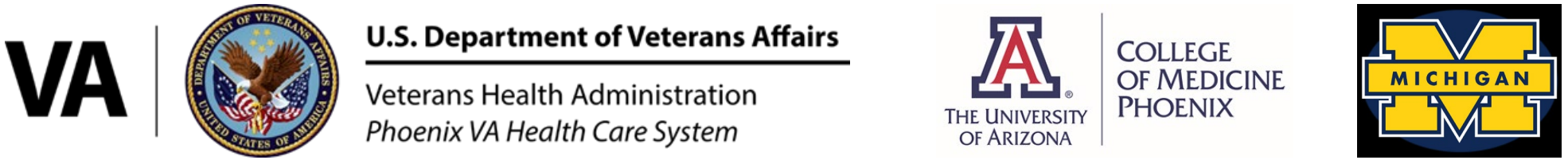
Predicting 5-Year Dementia Conversion in Veterans with Mild Cognitive Impairment Using Electronic Health Records

Chase Irwin, Donna Tjandra, Chengcheng Hu, Vinod Aggarwal, Amanda Lienau, Bruno Giordani, Jenna Wiens, Raymond Q. Migrino



Mild Cognitive Impairment

- MCI is a stage preceding Alzheimer's disease (AD) and related dementias.[1]
- Meta-analysis: 33.6% of MCI patients progressed to AD[2]
- Prior autopsy studies show that the brain pathology in MCI is intermediate in severity between cognitively normal controls and patients with AD
- Important to identify MCI patients at risk of developing dementia:
  - Target candidates for early treatment, including new, but expensive tx (e.g. mAb lecanemab)
  - Identify high-risk participants to recruit to improve clinical trial success rates of new txs

(1) Acosta-Baena N. Lancet Neurol. 2011; 10:213 – 20; (2) Mitchell AJ, Acta Psychiatr Scand. 2009; 119:252 - 65

Predicting MCI to Dementia Conversion

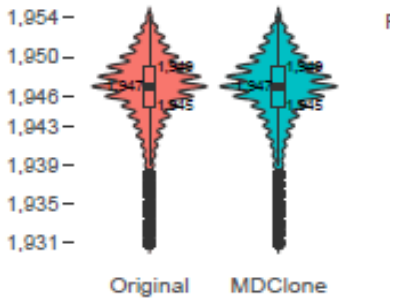
- Current methods to predict dementia rely on biomarkers (e.g. neuropsychologic tests, specialized imaging or CSF assays) that are not widely obtained, costly or invasive.[1]
- The generalizability of current models is limited due to small samples sizes.
- EHRs provide high-dimensional data collected during routine clinical encounters care that could be used to predict dementia conversion.

(1) Tjandra D. Alzheimer's Dement. 2022; 1 - 5

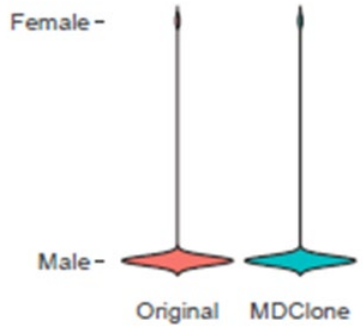
Synthetic Data

- Disadvantages of using EHR-based data include limited data access and vulnerability to privacy breaches
- Synthetic data do not pertain to real patients but closely resemble real data.
- Software (such as MDClone) could generate non-reversible, artificially created synthetic data that resemble the statistical characteristics and correlations of real data.
- The algorithm to create synthetic data is multivariate and generates all variables together, using a covariance measure.[1]
- Synthetic data is generated by random sampling from statistical distributions estimated from the original data. [1]

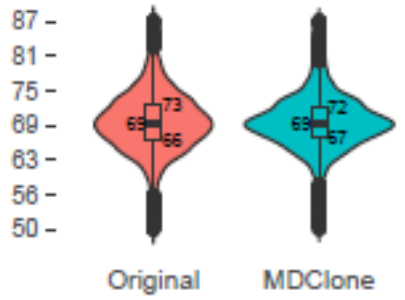
Date of Birth



Sex



ACD Age Onset



(1) Benaim RA, et al. JMIR Med Inform. 2020

Benefits of Synthetic Data

- Protection of patient privacy
- Increased access for researchers that accelerates scientific discoveries
- Increased scientific rigor through cross-validation of results
- Democratization of the statistical modelling process

Aims

- **Primary**: Develop a generalizable EHR-based model to predict MCI to all-cause dementia (ACD) conversion within 5 years

- Secondary: To compare the performance of MCI to ACD prediction models based on EHR real patient data versus synthetic data

Identifying MCI and ACD Patients

- ICD – 9/10 codes were used to identify MCI and ACD patients
- VA Centralized Interactive Phenomics Resource (CIPHER) validated algorithms aided in classification[1]
- MCI → 95% specificity; ACD → 82% specificity
- Both algorithms required a minimum of 2 visits with the Dx

(1) Logue MW, et al. Alzheimers Dementia. 2022

Identifying Comorbid Predictors of ACD Conversion

- All comorbid predictors were selected *a priori* based on previous literature testing these conditions as potential risk factors

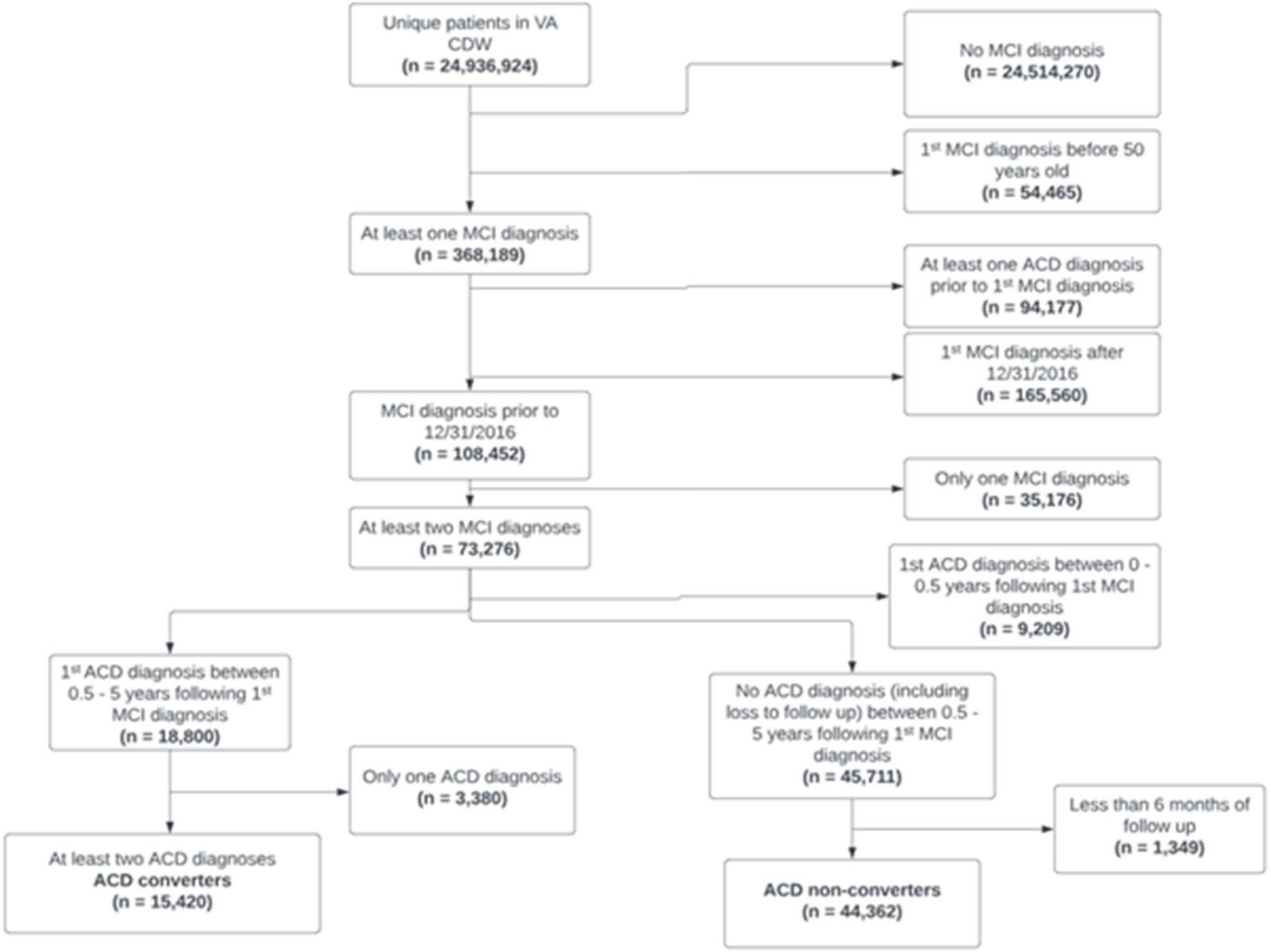
- ICD – 9/10 codes used
  - Charlson Comorbidity Index[1]
  - Elixhauser Comorbidity Index[2]
  - VA CIPHER

Supplement Table 1. Definition of mild cognitive impairment (MCI), all cause dementia (ACD) and co-morbidities

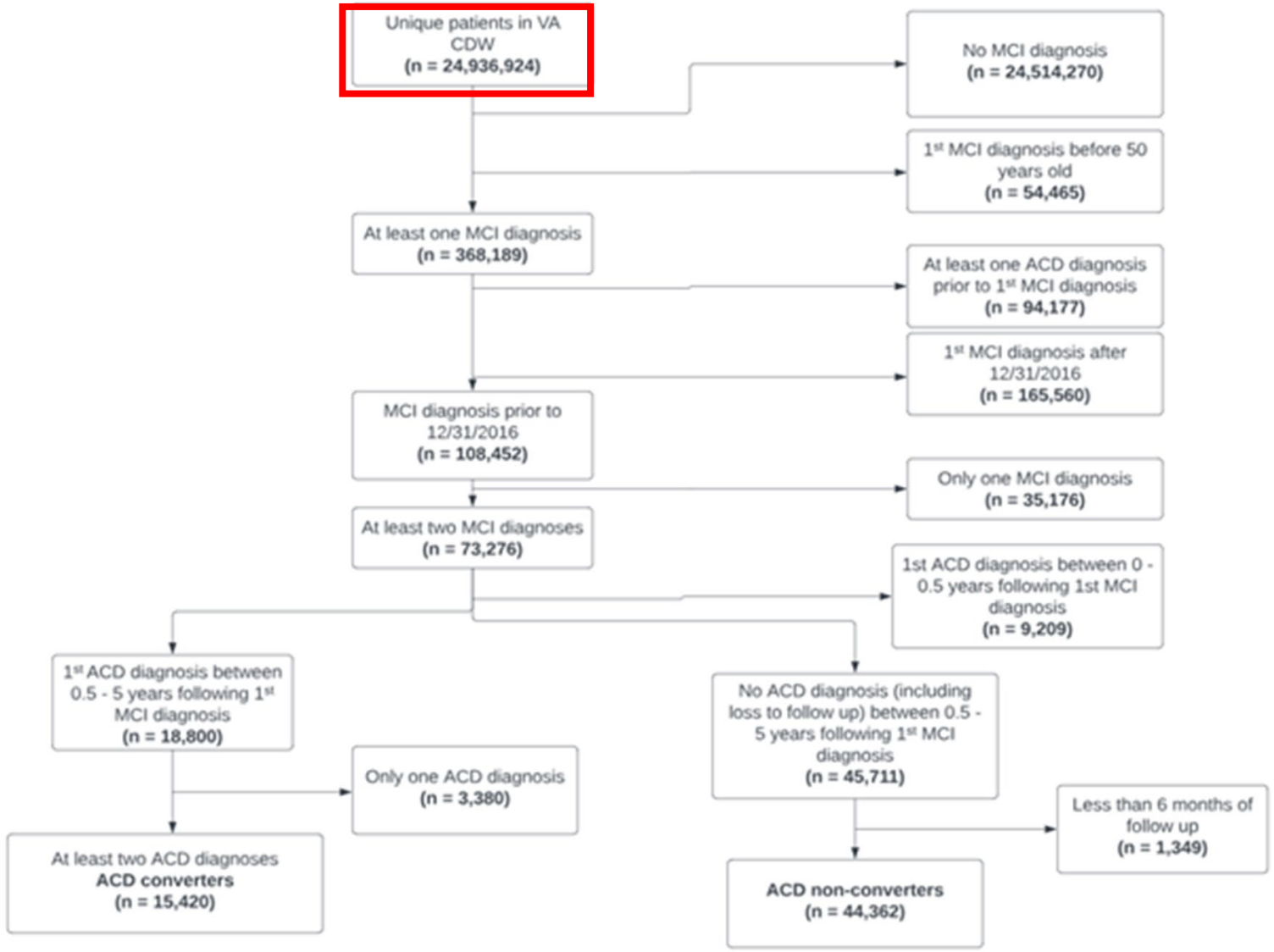
Comorbidities	ICD – 9 Codes	ICD – 10 Codes
Alcohol abuse[b]	265.2, 291.1 – 291.3, 291.5 –291.9, 303.0, 303.9, 305.0, 357.5, 425.5, 535.3, 571.0 – 571.3, 980.x, V11.3	F10, E52, G62.1, I42.6, K29.2, K70.0, K70.3, K70.9, T51.x, Z50.2, Z71.4, Z72.1
Atrial fibrillation[a]	427.31	I48.0, I48.1, I48.11, I48.19, I48.2, I48.20, I48.21, I48.91
Cerebrovascular disease[c]	362.34, 430.x – 438.x	G45.x, G46.x, H34.0, I60.x–I69.x
Depression[b]	296.2, 296.3, 296.5,300.4, 309.x, 311	F20.4, F31.3 – F31.5, F32.x, F33.x, F34.1, F41.2, F43.2
Diabetes[c]	250.0 – 250.3, 250.8, 250.9 250.4 – 250.7	E10.0, E10.1, E10.6, E10.8, E10.9, E11.0, E11.1, E11.6, E11.8, E11.9, E12.0, E12.1, E12.6, E12.8, E12.9, E13.0, E13.1, E13.6, E13.8, E13.9, E14.0, E14.1, E14.6, E14.8, E14.9, E10.2 – E10.5, E10.7, E11.2 – E11.5, E11.7, E12.2 –E12.5, E12.7, E13.2– E13.5, E13.7, E14.2–E14.5, E14.7
Hearing Loss[d]	389.x	H90.x, H91.x
Heart failure[c]	398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 425.4 – 425.9, 428.x	I09.9, I11.0, I13.0, I13.2, I25.5, I42.0, I42.5 – I42.9, I43.x, I50.x, P29.0
Hyperlipidemia[a]	272.0, 272.1, 272.2, 272.3, 272.4	E78.0, E78.00, E78.01, E78.1, E78.2, E78.3, E78.4, E78.41, E78.49, E78.5
Hypertension[b]	401.x, 402.x – 405.x	I10.x, I11.x – I13.x, I15.x
Liver disease[c]	070.22, 070.23, 070.32, 070.33, 070.44, 070.54, 070.6, 070.9, 570.x, 571.x, 573.3, 573.4, 573.8, 573.9, V42.7 456.0 – 456.2, 572.2 – 572.8	B18.x, K70.0 – K70.3, K70.9, K71.3 – K71.5, K71.7, K73.x, K74.x, K76.0, K76.2 – K76.4, K76.8, K76.9, Z94.4 I85.0, I85.9, I86.4, I98.2, K70.4, K71.1, K72.1, K72.9, K76.5, K76.6, K76.7

(1) Charlson ME, et al. J Chronic Dis. 1987 (2) Elixhauser S, et al. Med Care. 1998

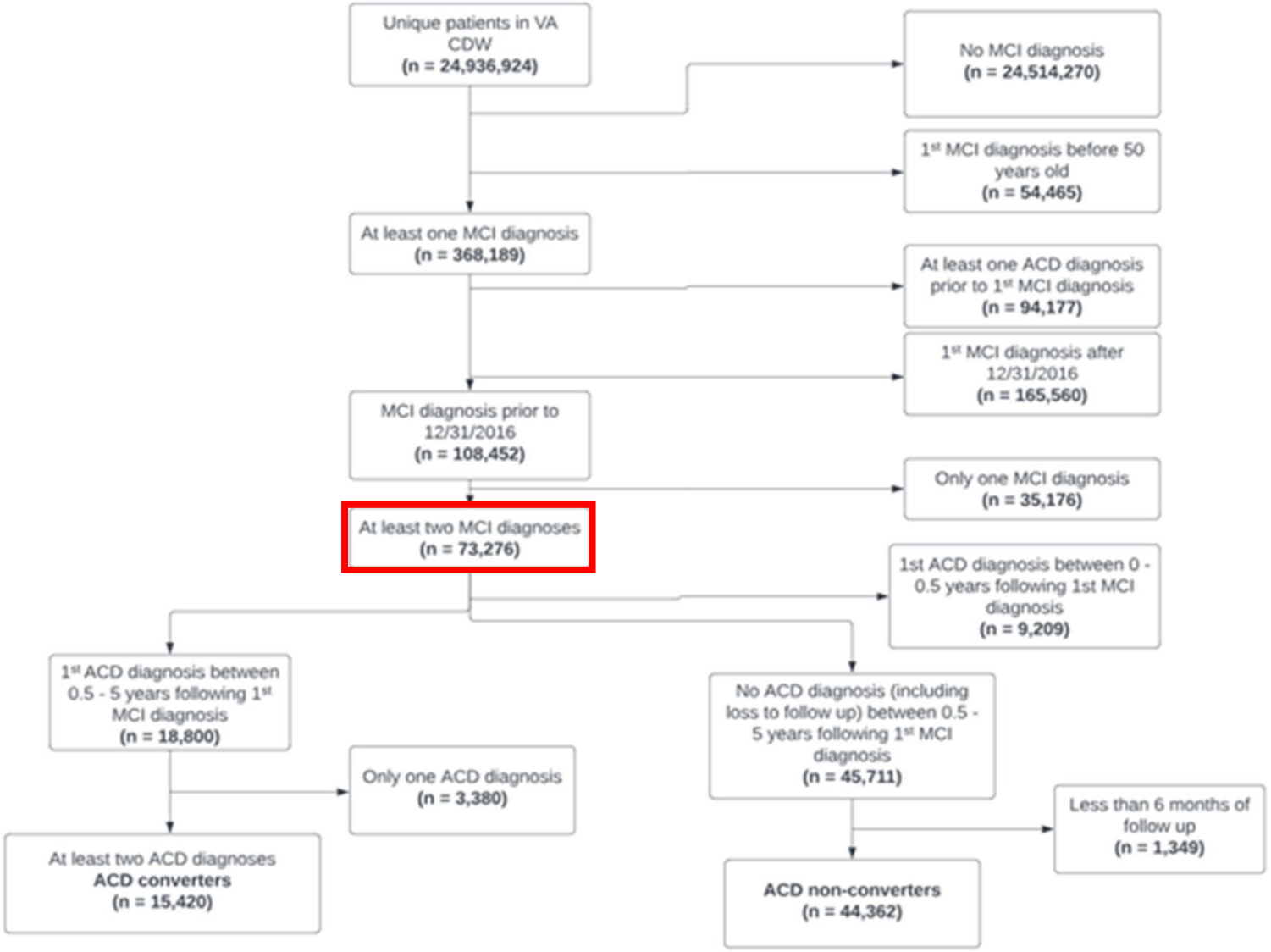
Cohort Flowchart



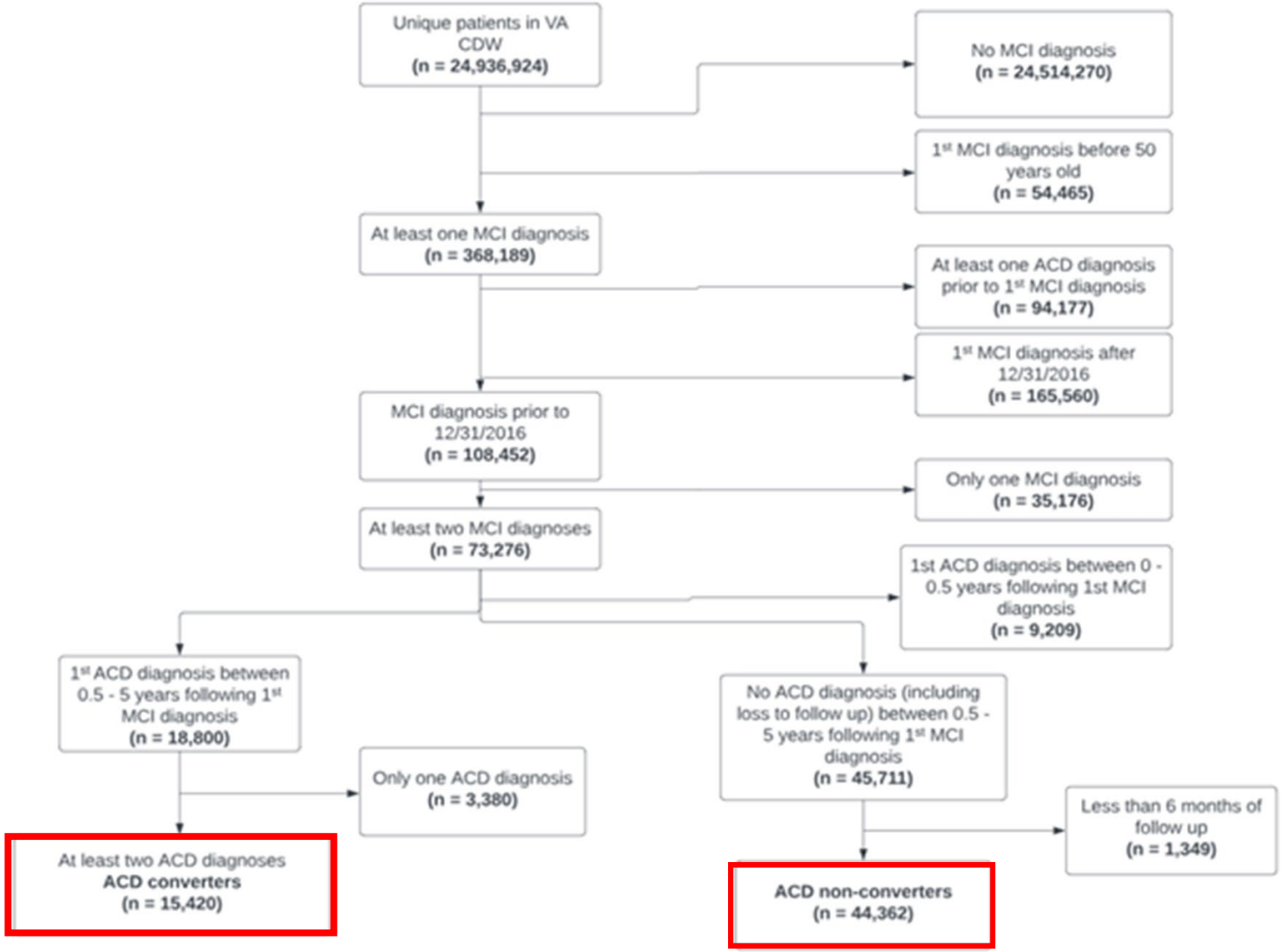
Cohort Flowchart



Cohort Flowchart

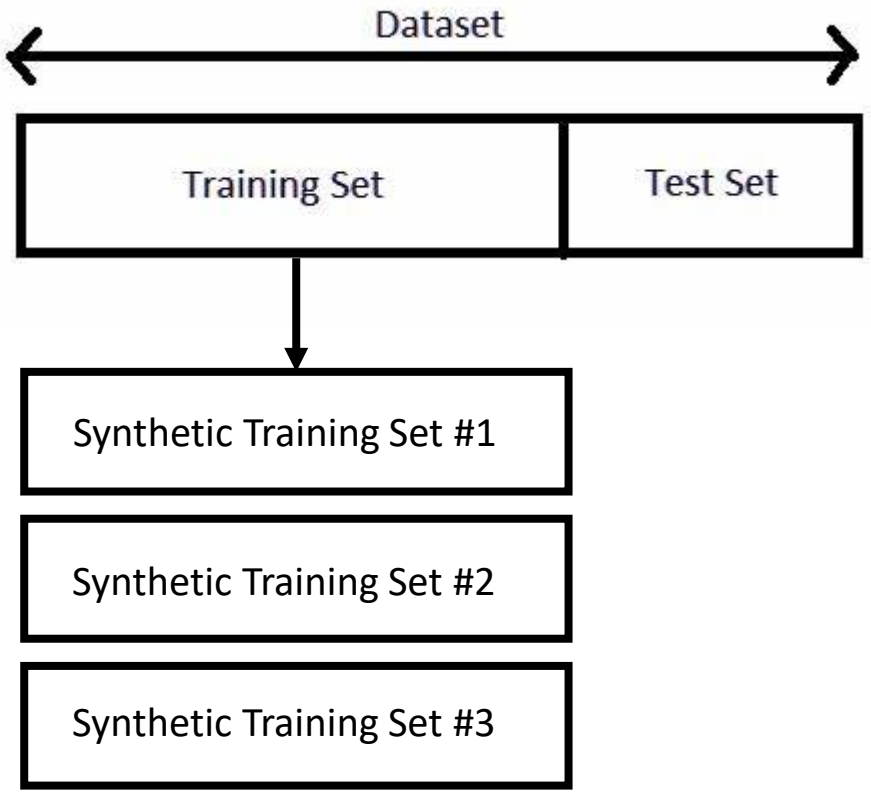


Cohort Flowchart



Statistical Methods

- We randomly partitioned our full cohort into a training set (70%) and test set (30%)
- Created three synthetic training sets using MDClone
- Fit Cox proportional hazard models for each training set
- Applied each training set to our held-out test set and estimated the time-dependent AUC and Brier score for ACD conversion at 5 years



Results: Demographic Profiles

- 15,420 out of 59,782 MCI patients (26%) converted to ACD within 5 years
- Kaplan Meier estimate of 5-year conversion to ACD was 28.4% (95% CI: 28.0% - 28.8%)
- Median time to conversion was 1.94 (IQR: 1.09 – 3.10) years
- Median age of MCI Dx was 70.9 (IQR: 63.7 – 80.7) years

Results: Demographic Profiles

	A. Training Set Real (n = 41,817)			B. Test Set Real (n = 17,965)		
Demographics	**ACD (n = 10,784)**	**No ACD (n = 31,033)**	**p-value**	**ACD (n = 4,636)**	**No ACD (n = 13,329)**	**p-value**
Age at MCI DX, median (IQR), years	77.01 (69.17 – 83.34)	69.01 (62.11 – 79.18)	<0.001*	76.72 (69.08 – 83.16)	68.91 (62.09 – 79.07)	<0.001*
Age yrs			<0.001*			<0.001*
S 50 – 55	162 (1.50)	2,611 (8.41)		67 (1.44)	1,105 (8.29)	
55 – 60	311 (2.88)	3,220 (10.38)		136 (2.93)	1,405 (10.54)	
60 – 65	890 (8.25)	4,916 (15.84)		398 (8.59)	2,103 (15.78)	
65 – 70	1,673 (15.51)	5,809 (18.72)		710 (15.32)	2,560 (19.21)	
70 – 75	1,662 (15.41)	3,835 (12.36)		743 (16.03)	1,655 (12.42)	
75 – 80	1,924 (17.84)	3,445 (11.10)		831 (17.93)	1,469 (11.02)	
80 – 85	2,135 (19.80)	3,542 (11.41)		928 (20.02)	1,531 (11.49)	
>85	2,027 (18.80)	3,655 (11.78)		823 (17.75)	1,501 (11.26)	
Race, No. (%)			<0.001*			0.002*
Asian/Pacific	146 (1.35)	426 (1.37)		60 (1.29)	178 (1.34)	
Black	1,349 (12.51)	4,431 (14.28)		590 (12.73)	1,998 (14.99)	
Native American	56 (0.52)	210 (0.68)		20 (0.43)	78 (0.59)	
Other†	1,006 (9.33)	2,886 (9.30)		423 (9.12)	1,234 (9.26)	
White	8,227 (76.29)	23,080 (74.37)		3,543 (76.42)	9,841 (73.83)	
Ethnicity, No. (%)			0.03*			0.03*
Not Hispanic or Latino	9,590 (88.93)	27,436 (88.41)		4,069 (87.77)	11,802 (88.54)	
Hispanic or Latino	650 (6.03)	1,828 (5.89)		322 (6.95)	783 (5.87)	
Other	544 (5.05)	1,769 (5.70)		245 (5.29)	744 (5.58)	
Sex, No. (%)			<0.001*			<0.001*
Female	354 (3.28)	1,472 (4.74)		156 (3.37)	667 (5.00)	
Male	10,430 (96.72)	29,561 (95.26)		4,480 (96.64)	12,662 (95.00)	
BMI, No. (%)			<0.001*			<0.001*
Underweight	105 (0.97)	372 (1.20)		60 (1.29)	166 (1.25)	
Normal	2,952 (27.37)	6,717 (21.65)		1,273 (27.46)	2,909 (21.83)	
Overweight	4,725 (43.82)	12,584 (40.55)		1,958 (42.24)	5,312 (39.85)	
Obese	3,002 (27.84)	11,360 (36.61)		1,345 (29.01)	4,942 (37.08)	

Results: Demographic Profiles

	A. Training Set Real (n = 41,817)			B. Test Set Real (n = 17,965)		
Demographics	**ACD (n = 10,784)**	**No ACD (n = 31,033)**	**p-value**	**ACD (n = 4,636)**	**No ACD (n = 13,329)**	**p-value**
Age at MCI DX, median (IQR), years	77.01 (69.17 – 83.34)	69.01 (62.11 – 79.18)	<0.001*	76.72 (69.08 – 83.16)	68.91 (62.09 – 79.07)	<0.001*
Age yrs			<0.001*			<0.001*
S 50 – 55	162 (1.50)	2,611 (8.41)		67 (1.44)	1,105 (8.29)	
55 – 60	311 (2.88)	3,220 (10.38)		136 (2.93)	1,405 (10.54)	
60 – 65	890 (8.25)	4,916 (15.84)		398 (8.59)	2,103 (15.78)	
65 – 70	1,673 (15.51)	5,809 (18.72)		710 (15.32)	2,560 (19.21)	
70 – 75	1,662 (15.41)	3,835 (12.36)		743 (16.03)	1,655 (12.42)	
75 – 80	1,924 (17.84)	3,445 (11.10)		831 (17.93)	1,469 (11.02)	
80 – 85	2,135 (19.80)	3,542 (11.41)		928 (20.02)	1,531 (11.49)	
>85	2,027 (18.80)	3,655 (11.78)		823 (17.75)	1,501 (11.26)	
Race, No. (%)			<0.001*			0.002*
Asian/Pacific	146 (1.35)	426 (1.37)		60 (1.29)	178 (1.34)	
Black	1,349 (12.51)	4,431 (14.28)		590 (12.73)	1,998 (14.99)	
Native American	56 (0.52)	210 (0.68)		20 (0.43)	78 (0.59)	
Other†	1,006 (9.33)	2,886 (9.30)		423 (9.12)	1,234 (9.26)	
White	8,227 (76.29)	23,080 (74.37)		3,543 (76.42)	9,841 (73.83)	
Ethnicity, No. (%)			0.03*			0.03*
Not Hispanic or Latino	9,590 (88.93)	27,436 (88.41)		4,069 (87.77)	11,802 (88.54)	
Hispanic or Latino	650 (6.03)	1,828 (5.89)		322 (6.95)	783 (5.87)	
Other	544 (5.05)	1,769 (5.70)		245 (5.29)	744 (5.58)	
Sex, No. (%)			<0.001*			<0.001*
Female	354 (3.28)	1,472 (4.74)		156 (3.37)	667 (5.00)	
Male	10,430 (96.72)	29,561 (95.26)		4,480 (96.64)	12,662 (95.00)	
BMI, No. (%)			<0.001*			<0.001*
Underweight	105 (0.97)	372 (1.20)		60 (1.29)	166 (1.25)	
Normal	2,952 (27.37)	6,717 (21.65)		1,273 (27.46)	2,909 (21.83)	
Overweight	4,725 (43.82)	12,584 (40.55)		1,958 (42.24)	5,312 (39.85)	
Obese	3,002 (27.84)	11,360 (36.61)		1,345 (29.01)	4,942 (37.08)	

- MCI patients who converted were older than those who did not

Results: Demographic Profiles

	A. Training Set Real (n = 41,817)			B. Test Set Real (n = 17,965)		
Demographics	**ACD (n = 10,784)**	**No ACD (n = 31,033)**	**p-value**	**ACD (n = 4,636)**	**No ACD (n = 13,329)**	**p-value**
Age at MCI DX, median (IQR), years	77.01 (69.17 – 83.34)	69.01 (62.11 – 79.18)	<0.001*	76.72 (69.08 – 83.16)	68.91 (62.09 – 79.07)	<0.001*
Age yrs			<0.001*			<0.001*
S 50 – 55	162 (1.50)	2,611 (8.41)		67 (1.44)	1,105 (8.29)	
55 – 60	311 (2.88)	3,220 (10.38)		136 (2.93)	1,405 (10.54)	
60 – 65	890 (8.25)	4,916 (15.84)		398 (8.59)	2,103 (15.78)	
65 – 70	1,673 (15.51)	5,809 (18.72)		710 (15.32)	2,560 (19.21)	
70 – 75	1,662 (15.41)	3,835 (12.36)		743 (16.03)	1,655 (12.42)	
75 – 80	1,924 (17.84)	3,445 (11.10)		831 (17.93)	1,469 (11.02)	
80 – 85	2,135 (19.80)	3,542 (11.41)		928 (20.02)	1,531 (11.49)	
>85	2,027 (18.80)	3,655 (11.78)		823 (17.75)	1,501 (11.26)	
Race, No. (%)			<0.001*			0.002*
Asian/Pacific	146 (1.35)	426 (1.37)		60 (1.29)	178 (1.34)	
Black	1,349 (12.51)	4,431 (14.28)		590 (12.73)	1,998 (14.99)	
Native American	56 (0.52)	210 (0.68)		20 (0.43)	78 (0.59)	
Other†	1,006 (9.33)	2,886 (9.30)		423 (9.12)	1,234 (9.26)	
White	8,227 (76.29)	23,080 (74.37)		3,543 (76.42)	9,841 (73.83)	
Ethnicity, No. (%)			0.03*			0.03*
Not Hispanic or Latino	9,590 (88.93)	27,436 (88.41)		4,069 (87.77)	11,802 (88.54)	
Hispanic or Latino	650 (6.03)	1,828 (5.89)		322 (6.95)	783 (5.87)	
Other	544 (5.05)	1,769 (5.70)		245 (5.29)	744 (5.58)	
Sex, No. (%)			<0.001*			<0.001*
Female	354 (3.28)	1,472 (4.74)		156 (3.37)	667 (5.00)	
Male	10,430 (96.72)	29,561 (95.26)		4,480 (96.64)	12,662 (95.00)	
BMI, No. (%)			<0.001*			<0.001*
Underweight	105 (0.97)	372 (1.20)		60 (1.29)	166 (1.25)	
Normal	2,952 (27.37)	6,717 (21.65)		1,273 (27.46)	2,909 (21.83)	
Overweight	4,725 (43.82)	12,584 (40.55)		1,958 (42.24)	5,312 (39.85)	
Obese	3,002 (27.84)	11,360 (36.61)		1,345 (29.01)	4,942 (37.08)	

- MCI patients who converted were older than those who did not
- Fewer obese MCI patients converted to ACD

Results: Demographic Profiles

	A. Training Set Real (n = 41,817)			B. Test Set Real (n = 17,965)		
Demographics	**ACD (n = 10,784)**	**No ACD (n = 31,033)**	**p-value**	**ACD (n = 4,636)**	**No ACD (n = 13,329)**	**p-value**
Age at MCI DX, median (IQR), years	77.01 (69.17 – 83.34)	69.01 (62.11 – 79.18)	<0.001*	76.72 (69.08 – 83.16)	68.91 (62.09 – 79.07)	<0.001*
Age yrs			<0.001*			<0.001*
S 50 – 55	162 (1.50)	2,611 (8.41)		67 (1.44)	1,105 (8.29)	
55 – 60	311 (2.88)	3,220 (10.38)		136 (2.93)	1,405 (10.54)	
60 – 65	890 (8.25)	4,916 (15.84)		398 (8.59)	2,103 (15.78)	
65 – 70	1,673 (15.51)	5,809 (18.72)		710 (15.32)	2,560 (19.21)	
70 – 75	1,662 (15.41)	3,835 (12.36)		743 (16.03)	1,655 (12.42)	
75 – 80	1,924 (17.84)	3,445 (11.10)		831 (17.93)	1,469 (11.02)	
80 – 85	2,135 (19.80)	3,542 (11.41)		928 (20.02)	1,531 (11.49)	
>85	2,027 (18.80)	3,655 (11.78)		823 (17.75)	1,501 (11.26)	
Race, No. (%)			<0.001*			0.002*
Asian/Pacific	146 (1.35)	426 (1.37)		60 (1.29)	178 (1.34)	
Black	1,349 (12.51)	4,431 (14.28)		590 (12.73)	1,998 (14.99)	
Native American	56 (0.52)	210 (0.68)		20 (0.43)	78 (0.59)	
Other†	1,006 (9.33)	2,886 (9.30)		423 (9.12)	1,234 (9.26)	
White	8,227 (76.29)	23,080 (74.37)		3,543 (76.42)	9,841 (73.83)	
Ethnicity, No. (%)			0.03*			0.03*
Not Hispanic or Latino	9,590 (88.93)	27,436 (88.41)		4,069 (87.77)	11,802 (88.54)	
Hispanic or Latino	650 (6.03)	1,828 (5.89)		322 (6.95)	783 (5.87)	
Other	544 (5.05)	1,769 (5.70)		245 (5.29)	744 (5.58)	
Sex, No. (%)			<0.001*			<0.001*
Female	354 (3.28)	1,472 (4.74)		156 (3.37)	667 (5.00)	
Male	10,430 (96.72)	29,561 (95.26)		4,480 (96.64)	12,662 (95.00)	
BMI, No. (%)			<0.001*			<0.001*
Underweight	105 (0.97)	372 (1.20)		60 (1.29)	166 (1.25)	
Normal	2,952 (27.37)	6,717 (21.65)		1,273 (27.46)	2,909 (21.83)	
Overweight	4,725 (43.82)	12,584 (40.55)		1,958 (42.24)	5,312 (39.85)	
Obese	3,002 (27.84)	11,360 (36.61)		1,345 (29.01)	4,942 (37.08)	

- MCI patients who converted were older than those who did not
- Fewer obese MCI patients converted to ACD
- The overall cohort was predominately white and male

Results: Comorbid Profiles

	A. Training Set Real (n = 41,497)			B. Test Set Real (n = 17,848)		
	ACD (n = 10,794)	No ACD (n = 30,703)	p-value	ACD (n = 4,638)	No ACD (n = 13,210)	p-value
Comorbidities, No. (%)						
Heart Failure	1,693 (15.69)	4,518 (14.72)	0.02*	750 (16.17)	1,928 (14.59)	0.01*
Renal Disease	1,864 (17.27)	4,776 (15.56)	<0.001*	865 (18.65)	1,997 (15.12)	<0.001*
Rheumatic Disease	468 (4.34)	1,179 (3.84)	0.03*	219 (4.72)	491 (3.72)	0.003*
Hyperlipidemia	8,712 (80.71)	23,755 (77.37)	<0.001*	3,771 (81.31)	10,246 (77.56)	<0.001*
Sleep Apnea	2,304 (21.35)	8,209 (26.74)	<0.001*	1,002 (21.60)	3,625 (27.44)	<0.001*
Peripheral Vascular Disease	2,665 (24.69)	6,384 (20.79)	<0.001*	1,140 (24.58)	2,698 (20.42)	<0.001*
Peptic Ulcer Disease	733 (6.79)	1,822 (5.93)	0.002*	346 (7.46)	780 (5.91)	<0.001*
Atrial Fibrillation	1,657 (15.35)	4,018 (13.09)	<0.001*	783 (16.88)	1,687 (12.77)	<0.001*
Myocardial Infarction	1,286 (11.91)	3,113 (10.14)	<0.001*	559 (12.05)	1,330 (10.07)	<0.001*
Hypertension	9,080 (84.12)	24,493 (79.77)	<0.001*	3,940 (84.95)	10,497 (79.46)	<0.001*
Cerebrovascular Disease	3,192 (29.57)	7,884 (25.68)	<0.001*	1,423 (30.68)	3,320 (25.13)	<0.001*
Depression	5,708 (52.88)	19,103 (62.22)	<0.001*	2,472 (53.30)	8,269 (62.60)	<0.001*
Alcohol Abuse	1,676 (15.53)	6,968 (22.70)	<0.001*	689 (14.86)	3,045 (23.05)	<0.001*
Liver Disease	934 (8.65)	3,387 (11.03)	<0.001*	393 (8.47)	1,527 (11.56)	<0.001*
Diabetes	4,304 (39.87)	11,953 (38.93)	0.09	1,967 (42.41)	5,081 (38.46)	<0.001*
Hearing Loss	6,168 (57.14)	15,897 (51.78)	<0.001*	2,641 (56.94)	6,812 (51.57)	<0.001*

- ACD converters had significantly higher proportions for all comorbid predictors besides
  - Sleep Apnea
  - Depression
  - Alcohol Abuse

Results: Cox Regression

	Hazard Ratio (95% CI)	p-value		Hazard Ratio (95% CI)	p-value
Age MCI DX, years			**Comorbidities**		
50 – 55	Ref	<0.001*	Cerebrovascular Disease	1.10 (1.06 – 1.15)*	<0.001*
55 – 60	1.53 (1.26 – 1.85)*		Myocardial Infarction	1.09 (1.03 – 1.16)*	0.004*
60 – 65	2.76 (2.33 – 3.26)*		Hypertension	1.07 (1.02 – 1.13)*	0.01*
65 – 70	4.17 (3.54 – 4.90)*		Diabetes	1.06 (1.02 – 1.10)	0.007*
70 – 75	6.00 (5.09 – 7.05)*		Liver Disease	1.06 (0.99 – 1.14)	0.09
75 – 80	7.53 (6.40 – 8.86)*		Sleep Apnea	0.95 (0.91 – 1.00)*	0.04*
80 – 85	8.35 (7.10 – 9.82)*		Alcohol Abuse	0.93 (0.88 – 0.98)*	0.007*
>85	8.94 (7.59 – 10.52)*		Peripheral Vascular Disease	NA	NA
Race			Heart Failure	NA	NA
Black	1.02 (0.96 – 1.08)	0.63	Renal Disease	NA	NA
Other[a]	0.98 (0.92 – 1.04)		Rheumatic Disease	NA	NA
White	Ref		Hyperlipidemia	NA	NA
Sex			Peptic Ulcer Disease	NA	NA
Female	0.99 (0.89 – 1.10)	0.85	Atrial Fibrillation	NA	NA
Male	Ref		Depression	NA	NA
BMI			Hearing Loss	NA	NA
Underweight	0.89 (0.73 – 1.09)	<0.001*			
Normal	Ref				
Overweight	0.88 (0.84 – 0.92)*				
Obese	0.75 (0.71 – 0.80)*				

NA: Not applicable due to variable being removed from final Cox proportional hazards model by selection procedure

Results: Cox Regression

	Hazard Ratio (95% CI)	p-value		Hazard Ratio (95% CI)	p-value
Age MCI DX, years			**Comorbidities**		
50 – 55	Ref	<0.001*	Cerebrovascular Disease	1.10 (1.06 – 1.15)*	<0.001*
55 – 60	1.53 (1.26 – 1.85)*		Myocardial Infarction	1.09 (1.03 – 1.16)*	0.004*
60 – 65	2.76 (2.33 – 3.26)*		Hypertension	1.07 (1.02 – 1.13)*	0.01*
65 – 70	4.17 (3.54 – 4.90)*		Diabetes	1.06 (1.02 – 1.10)	0.007*
70 – 75	6.00 (5.09 – 7.05)*		Liver Disease	1.06 (0.99 – 1.14)	0.09
75 – 80	7.53 (6.40 – 8.86)*		Sleep Apnea	0.95 (0.91 – 1.00)*	0.04*
80 – 85	8.35 (7.10 – 9.82)*		Alcohol Abuse	0.93 (0.88 – 0.98)*	0.007*
>85	8.94 (7.59 – 10.52)*		Peripheral Vascular Disease	NA	NA
Race			Heart Failure	NA	NA
Black	1.02 (0.96 – 1.08)	0.63	Renal Disease	NA	NA
Other[a]	0.98 (0.92 – 1.04)		Rheumatic Disease	NA	NA
White	Ref		Hyperlipidemia	NA	NA
Sex			Peptic Ulcer Disease	NA	NA
Female	0.99 (0.89 – 1.10)	0.85	Atrial Fibrillation	NA	NA
Male	Ref		Depression	NA	NA
BMI			Hearing Loss	NA	NA
Underweight	0.89 (0.73 – 1.09)	<0.001*			
Normal	Ref				
Overweight	0.88 (0.84 – 0.92)*				
Obese	0.75 (0.71 – 0.80)*				

NA: Not applicable due to variable being removed from final Cox proportional hazards model by selection procedure

- Patient age is the overwhelming risk factor

Results: Cox Regression

	Hazard Ratio (95% CI)	p-value		Hazard Ratio (95% CI)	p-value
Age MCI DX, years			**Comorbidities**		
50 – 55	Ref	<0.001*	Cerebrovascular Disease	1.10 (1.06 – 1.15)*	<0.001*
55 – 60	1.53 (1.26 – 1.85)*		Myocardial Infarction	1.09 (1.03 – 1.16)*	0.004*
60 – 65	2.76 (2.33 – 3.26)*		Hypertension	1.07 (1.02 – 1.13)*	0.01*
65 – 70	4.17 (3.54 – 4.90)*		Diabetes	1.06 (1.02 – 1.10)	0.007*
70 – 75	6.00 (5.09 – 7.05)*		Liver Disease	1.06 (0.99 – 1.14)	0.09
75 – 80	7.53 (6.40 – 8.86)*		Sleep Apnea	0.95 (0.91 – 1.00)*	0.04*
80 – 85	8.35 (7.10 – 9.82)*		Alcohol Abuse	0.93 (0.88 – 0.98)*	0.007*
>85	8.94 (7.59 – 10.52)*		Peripheral Vascular Disease	NA	NA
Race			Heart Failure	NA	NA
Black	1.02 (0.96 – 1.08)	0.63	Renal Disease	NA	NA
Other[a]	0.98 (0.92 – 1.04)		Rheumatic Disease	NA	NA
White	Ref		Hyperlipidemia	NA	NA
Sex			Peptic Ulcer Disease	NA	NA
Female	0.99 (0.89 – 1.10)	0.85	Atrial Fibrillation	NA	NA
Male	Ref		Depression	NA	NA
BMI			Hearing Loss	NA	NA
Underweight	0.89 (0.73 – 1.09)	<0.001*			
Normal	Ref				
Overweight	0.88 (0.84 – 0.92)*				
Obese	0.75 (0.71 – 0.80)*				

NA: Not applicable due to variable being removed from final Cox proportional hazards model by selection procedure

- Patient age is the overwhelming risk factor
- High BMI was protective of ACD conversion

Results: Cox Regression

	Hazard Ratio (95% CI)	p-value		Hazard Ratio (95% CI)	p-value
Age MCI DX, years			**Comorbidities**		
50 – 55	Ref	<0.001*	Cerebrovascular Disease	1.10 (1.06 – 1.15)*	<0.001*
55 – 60	1.53 (1.26 – 1.85)*		Myocardial Infarction	1.09 (1.03 – 1.16)*	0.004*
60 – 65	2.76 (2.33 – 3.26)*		Hypertension	1.07 (1.02 – 1.13)*	0.01*
65 – 70	4.17 (3.54 – 4.90)*		Diabetes	1.06 (1.02 – 1.10)	0.007*
70 – 75	6.00 (5.09 – 7.05)*		Liver Disease	1.06 (0.99 – 1.14)	0.09
75 – 80	7.53 (6.40 – 8.86)*		Sleep Apnea	0.95 (0.91 – 1.00)*	0.04*
80 – 85	8.35 (7.10 – 9.82)*		Alcohol Abuse	0.93 (0.88 – 0.98)*	0.007*
>85	8.94 (7.59 – 10.52)*		Peripheral Vascular Disease	NA	NA
Race			Heart Failure	NA	NA
Black	1.02 (0.96 – 1.08)	0.63	Renal Disease	NA	NA
Other[a]	0.98 (0.92 – 1.04)		Rheumatic Disease	NA	NA
White	Ref		Hyperlipidemia	NA	NA
Sex			Peptic Ulcer Disease	NA	NA
Female	0.99 (0.89 – 1.10)	0.85	Atrial Fibrillation	NA	NA
Male	Ref		Depression	NA	NA
BMI			Hearing Loss	NA	NA
Underweight	0.89 (0.73 – 1.09)	<0.001*			
Normal	Ref				
Overweight	0.88 (0.84 – 0.92)*				
Obese	0.75 (0.71 – 0.80)*				

NA: Not applicable due to variable being removed from final Cox proportional hazards model by selection procedure

- Patient age is the overwhelming risk factor
- High BMI was protective of ACD conversion
- Vascular disease-related comorbidities were the strongest comorbid predictors

Results: Cox Regression

	Hazard Ratio (95% CI)	p-value		Hazard Ratio (95% CI)	p-value
Age MCI DX, years			**Comorbidities**		
50 – 55	Ref	<0.001*	Cerebrovascular Disease	1.10 (1.06 – 1.15)*	<0.001*
55 – 60	1.53 (1.26 – 1.85)*		Myocardial Infarction	1.09 (1.03 – 1.16)*	0.004*
60 – 65	2.76 (2.33 – 3.26)*		Hypertension	1.07 (1.02 – 1.13)*	0.01*
65 – 70	4.17 (3.54 – 4.90)*		Diabetes	1.06 (1.02 – 1.10)	0.007*
70 – 75	6.00 (5.09 – 7.05)*		Liver Disease	1.06 (0.99 – 1.14)	0.09
75 – 80	7.53 (6.40 – 8.86)*		Sleep Apnea	0.95 (0.91 – 1.00)*	0.04*
80 – 85	8.35 (7.10 – 9.82)*		Alcohol Abuse	0.93 (0.88 – 0.98)*	0.007*
>85	8.94 (7.59 – 10.52)*		Peripheral Vascular Disease	NA	NA
Race			Heart Failure	NA	NA
Black	1.02 (0.96 – 1.08)	0.63	Renal Disease	NA	NA
Other[a]	0.98 (0.92 – 1.04)		Rheumatic Disease	NA	NA
White	Ref		Hyperlipidemia	NA	NA
Sex			Peptic Ulcer Disease	NA	NA
Female	0.99 (0.89 – 1.10)	0.85	Atrial Fibrillation	NA	NA
Male	Ref		Depression	NA	NA
BMI			Hearing Loss	NA	NA
Underweight	0.89 (0.73 – 1.09)	<0.001*			
Normal	Ref				
Overweight	0.88 (0.84 – 0.92)*				
Obese	0.75 (0.71 – 0.80)*				

NA: Not applicable due to variable being removed from final Cox proportional hazards model by selection procedure

- Patient age is the overwhelming risk factor
- High BMI was protective of ACD conversion
- Vascular disease-related comorbidities were the strongest comorbid predictors
- Alcohol abuse and sleep apnea were found to be protective

Results: Cox Regression

	Hazard Ratio (95% CI)	p-value		Hazard Ratio (95% CI)	p-value
Age MCI DX, years			**Comorbidities**		
50 – 55	Ref	<0.001*	Cerebrovascular Disease	1.10 (1.06 – 1.15)*	<0.001*
55 – 60	1.53 (1.26 – 1.85)*		Myocardial Infarction	1.09 (1.03 – 1.16)*	0.004*
60 – 65	2.76 (2.33 – 3.26)*		Hypertension	1.07 (1.02 – 1.13)*	0.01*
65 – 70	4.17 (3.54 – 4.90)*		Diabetes	1.06 (1.02 – 1.10)	0.007*
70 – 75	6.00 (5.09 – 7.05)*		Liver Disease	1.06 (0.99 – 1.14)	0.09
75 – 80	7.53 (6.40 – 8.86)*		Sleep Apnea	0.95 (0.91 – 1.00)*	0.04*
80 – 85	8.35 (7.10 – 9.82)*		Alcohol Abuse	0.93 (0.88 – 0.98)*	0.007*
>85	8.94 (7.59 – 10.52)*		Peripheral Vascular Disease	NA	NA
Race			Heart Failure	NA	NA
Black	1.02 (0.96 – 1.08)	0.63	Renal Disease	NA	NA
Other[a]	0.98 (0.92 – 1.04)		Rheumatic Disease	NA	NA
White	Ref		Hyperlipidemia	NA	NA
Sex			Peptic Ulcer Disease	NA	NA
Female	0.99 (0.89 – 1.10)	0.85	Atrial Fibrillation	NA	NA
Male	Ref		Depression	NA	NA
BMI			Hearing Loss	NA	NA
Underweight	0.89 (0.73 – 1.09)	<0.001*			
Normal	Ref				
Overweight	0.88 (0.84 – 0.92)*				
Obese	0.75 (0.71 – 0.80)*				

NA: Not applicable due to variable being removed from final Cox proportional hazards model by selection procedure

- Predicting ACD conversion at five years
  - Time-dependent AUC = 0.73 (95% CI: 0.72 – 0.74)
  - Time-dependent Brier score = 0.18 (95% CI: 0.17 – 0.18)

Synthetic vs Real Data: Multivariate Analysis

	A. Training Set Real		B. Training Set Synthetic #1		C. Training Set Synthetic #2		D. Training Set Synthetic #3	
	Hazard Ratio (95% CI)	p-value	Hazard Ratio (95% CI)	p-value	Hazard Ratio (95% CI)	p-value	Hazard Ratio (95% CI)	p-value
Age MCI DX, years		<0.001*		<0.001*		<0.001*		<0.001*
50 – 55	Ref		Ref		Ref		Ref	
55 – 60	1.53 (1.26 – 1.85)*		1.59 (1.31 – 1.93)*		1.55 (1.28 – 1.89)*		1.55 (1.28 – 1.89)*	
60 – 65	2.76 (2.33 – 3.26)*		2.96 (2.50 – 3.53)*		2.88 (2.43 – 3.42)*		2.90 (2.44 – 3.45)*	
65 – 70	4.17 (3.54 – 4.90)*		4.43 (3.74 – 5.24)*		4.33 (3.66 – 5.12)*		4.37 (3.70 – 5.17)*	
70 – 75	6.00 (5.09 – 7.05)*		6.37 (5.38 – 7.55)*		6.23 (5.27 – 7.37)*		6.26 (5.29 – 7.41)*	
75 – 80	7.53 (6.40 – 8.86)*		8.03 (6.78 – 9.50)*		7.84 (6.64 – 9.27)*		7.89 (6.67 – 9.33)*	
80 – 85	8.35 (7.10 – 9.82)*		8.88 (7.51 – 10.51)*		8.72 (7.38 – 10.30)*		8.74 (7.39 – 10.33)*	
>85	8.94 (7.59 – 10.52)*		9.50 (8.02 – 11.25)*		9.27 (7.84 – 10.96)*		9.32 (7.88 – 11.03)*	
Race		0.63		0.70		0.65		0.63
Black	1.02 (0.96 – 1.08)		1.02 (0.96 – 1.08)		1.02 (0.96 – 1.08)		1.02 (0.96 – 1.08)	
Other[a]	0.98 (0.92 – 1.04)		0.98 (0.92 – 1.04)		0.98 (0.92 – 1.04)		0.98 (0.92 – 1.04)	
White	Ref		Ref		Ref		Ref	
Sex		0.85		0.97		0.92		0.97
Female	0.99 (0.89 – 1.10)		1.00 (0.90 – 1.11)		0.99 (0.89 – 1.11)		1.00 (0.90 – 1.11)	
Male	Ref		Ref		Ref		Ref	
BMI								
Underweight	0.89 (0.73 – 1.09)	<0.001*	0.88 (0.71 – 1.08)	<0.001*	0.88 (0.71 – 1.08)	<0.001*	0.87 (0.71 – 1.08)	<0.001*
Normal	Ref		Ref		Ref		Ref	
Overweight	0.88 (0.84 – 0.92)*		0.89 (0.85 – 0.93)*		0.89 (0.85 – 0.93)*		0.88 (0.84 – 0.93)*	
Obese	0.75 (0.71 – 0.80)*		0.76 (0.72 – 0.80)*		0.76 (0.72 – 0.80)*		0.75 (0.71 – 0.80)*	

Synthetic vs Real Data: Multivariate Analysis

	A. Training Set Real		B. Training Set Synthetic #1		C. Training Set Synthetic #2		D. Training Set Synthetic #3	
	Hazard Ratio (95% CI)	p-value	Hazard Ratio (95% CI)	p-value	Hazard Ratio (95% CI)	p-value	Hazard Ratio (95% CI)	p-value
Comorbidities								
Cerebrovascular Disease	1.10 (1.06 – 1.15)*	<0.001*	1.10 (1.05 – 1.15)*	<0.001*	1.10 (1.05 – 1.15)*	<0.001*	1.10 (1.05 – 1.15)*	<0.001*
Myocardial Infarction	1.09 (1.03 – 1.16)*	0.004*	1.09 (1.02 – 1.15)*	0.008*	1.09 (1.02 – 1.15)*	0.008*	1.08 (1.02 – 1.15)*	0.01*
Hypertension	1.07 (1.02 – 1.13)*	0.01*	1.07 (1.01 – 1.13)*	0.02*	1.07 (1.01 – 1.13)*	0.01*	1.08 (1.02 – 1.14)*	0.008*
Diabetes	1.06 (1.02 – 1.10)	0.007*	1.05 (1.01 – 1.10)	0.02*	1.05 (1.01 – 1.10)	0.02*	1.05 (1.01 – 1.10)	0.02*
Liver Disease	1.06 (0.99 – 1.14)	0.09	1.06 (0.99 – 1.14)	0.11	1.07 (0.99 – 1.14)	0.07	1.06 (0.99 – 1.14)	0.09
Peripheral Vascular Disease	NA	NA	1.03 (0.99 – 1.08)	0.15	1.04 (0.99 – 1.09)	0.13	1.04 (0.99 – 1.08)	0.14
Sleep Apnea	0.95 (0.91 – 1.00)*	0.04*	0.95 (0.90 – 1.00)*	0.04*	0.95 (0.90 – 1.00)*	0.04*	0.95 (0.90 – 1.00)*	0.04*
Alcohol Abuse	0.93 (0.88 – 0.98)*	0.007*	0.92 (0.87 – 0.98)*	0.004*	0.92 (0.87 – 0.97)*	0.003*	0.92 (0.87 – 0.97)*	0.003*
Heart Failure	NA	NA	NA	NA	NA	NA	NA	NA
Renal Disease	NA	NA	NA	NA	NA	NA	NA	NA
Rheumatic Disease	NA	NA	NA	NA	NA	NA	NA	NA
Hyperlipidemia	NA	NA	NA	NA	NA	NA	NA	NA
Peptic Ulcer Disease	NA	NA	NA	NA	NA	NA	NA	NA
Atrial Fibrillation	NA	NA	NA	NA	NA	NA	NA	NA
Depression	NA	NA	NA	NA	NA	NA	NA	NA
Hearing Loss	NA	NA	NA	NA	NA	NA	NA	NA

NA: Not applicable due to variable being removed from final Cox proportional hazards model by selection procedure

Synthetic vs Real Data: Prediction

	A. Training Set Real	B. Training Set Synthetic #1	C. Training Set Synthetic #2	D. Training Set Synthetic #3
Time – Dependent AUC (95% CI)	0.73 (0.72 – 0.74)	0.73 (0.72 – 0.74)	0.73 (0.72 – 0.74)	0.73 (0.72 – 0.74)
Time – Dependent AUC Comparison[a], (Difference) [p-value]	Ref	(<0.001) [p=0.72]	(<0.001) [p=0.95]	(<0.001) [p=0.60]
Time – Dependent Brier (95% CI)	0.18 (0.17 – 0.18)	0.18 (0.17 – 0.18)	0.18 (0.17 – 0.18)	0.18 (0.17 – 0.18)
Brier Score Comparisons[a], (Difference) [p-value]	Ref	(<0.001) [p=0.61]	(<0.001) [p=0.48]	(<0.001) [p=0.91]
Prediction Expected Conversion Probability, Median (IQR)	22.55% (27.67)	22.37% (27.79)	22.33% (27.78)	22.43% (27.78)
Correlation of Expected Conversion Probability	Ref	0.99	0.99	0.99

[a]Absolute value of Real minus Synthetic

Synthetic vs Real Data: Prediction

	A. Training Set Real	B. Training Set Synthetic #1	C. Training Set Synthetic #2	D. Training Set Synthetic #3
Time – Dependent AUC (95% CI)	0.73 (0.72 – 0.74)	0.73 (0.72 – 0.74)	0.73 (0.72 – 0.74)	0.73 (0.72 – 0.74)
Time – Dependent AUC Comparison[a], (Difference) [p-value]	Ref	(<0.001) [p=0.72]	(<0.001) [p=0.95]	(<0.001) [p=0.60]
Time – Dependent Brier (95% CI)	0.18 (0.17 – 0.18)	0.18 (0.17 – 0.18)	0.18 (0.17 – 0.18)	0.18 (0.17 – 0.18)
Brier Score Comparisons[a], (Difference) [p-value]	Ref	(<0.001) [p=0.61]	(<0.001) [p=0.48]	(<0.001) [p=0.91]
Prediction Expected Conversion Probability, Median (IQR)	22.55% (27.67)	22.37% (27.79)	22.33% (27.78)	22.43% (27.78)
Correlation of Expected Conversion Probability	Ref	0.99	0.99	0.99

[a]Absolute value of Real minus Synthetic

Synthetic vs Real Data: Prediction

	A. Training Set Real	B. Training Set Synthetic #1	C. Training Set Synthetic #2	D. Training Set Synthetic #3
Time – Dependent AUC (95% CI)	0.73 (0.72 – 0.74)	0.73 (0.72 – 0.74)	0.73 (0.72 – 0.74)	0.73 (0.72 – 0.74)
Time – Dependent AUC Comparison[a], (Difference) [p-value]	Ref	(<0.001) [p=0.72]	(<0.001) [p=0.95]	(<0.001) [p=0.60]
Time – Dependent Brier (95% CI)	0.18 (0.17 – 0.18)	0.18 (0.17 – 0.18)	0.18 (0.17 – 0.18)	0.18 (0.17 – 0.18)
Brier Score Comparisons[a], (Difference) [p-value]	Ref	(<0.001) [p=0.61]	(<0.001) [p=0.48]	(<0.001) [p=0.91]
Prediction Expected Conversion Probability, Median (IQR)	22.55% (27.67)	22.37% (27.79)	22.33% (27.78)	22.43% (27.78)
Correlation of Expected Conversion Probability	Ref	0.99	0.99	0.99

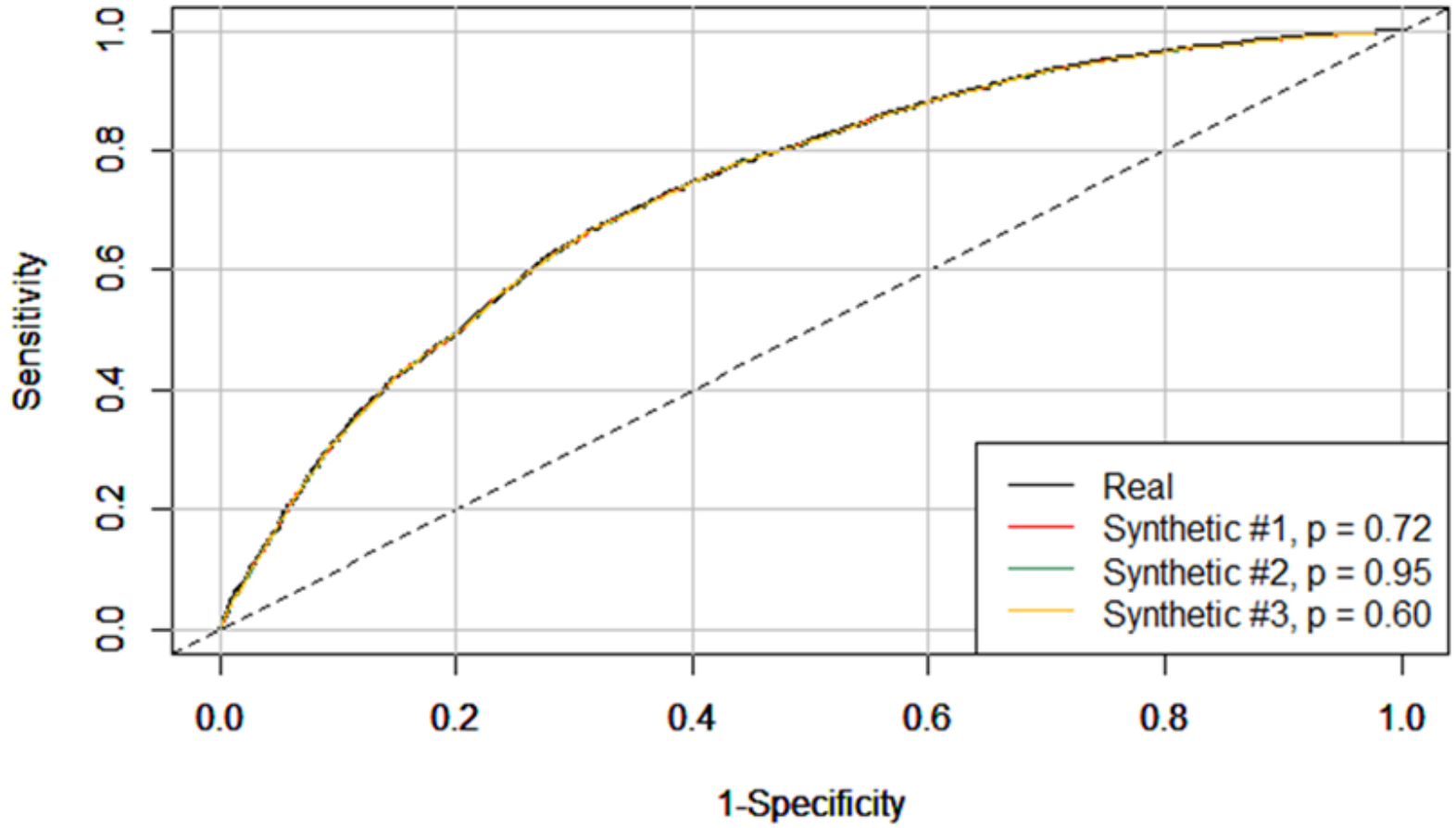
[a]Absolute value of Real minus Synthetic

Synthetic vs Real Data: Prediction

	A. Training Set Real	B. Training Set Synthetic #1	C. Training Set Synthetic #2	D. Training Set Synthetic #3
Time – Dependent AUC (95% CI)	0.73 (0.72 – 0.74)	0.73 (0.72 – 0.74)	0.73 (0.72 – 0.74)	0.73 (0.72 – 0.74)
Time – Dependent AUC Comparison[a], (Difference) [p-value]	Ref	(<0.001) [p=0.72]	(<0.001) [p=0.95]	(<0.001) [p=0.60]
Time – Dependent Brier (95% CI)	0.18 (0.17 – 0.18)	0.18 (0.17 – 0.18)	0.18 (0.17 – 0.18)	0.18 (0.17 – 0.18)
Brier Score Comparisons[a], (Difference) [p-value]	Ref	(<0.001) [p=0.61]	(<0.001) [p=0.48]	(<0.001) [p=0.91]
Prediction Expected Conversion Probability, Median (IQR)	22.55% (27.67)	22.37% (27.79)	22.33% (27.78)	22.43% (27.78)
Correlation of Expected Conversion Probability	Ref	0.99	0.99	0.99

[a]Absolute value of Real minus Synthetic

Synthetic vs Real Data: Prediction



Conclusions

- Primary Aim: Develop an EHR-based model to predict MCI to ACD conversion
  - Age, cerebrovascular disease, myocardial infarction, hypertension and diabetes are risk factors, with age being the overwhelming risk factor
  - High BMI, sleep apnea, and alcohol abuse are protective factors
  - EHR-based model showed good discriminative performance (AUC = 0.73) and good calibration (Brier score 0.18)

- Secondary Aim: Compare model performance using real vs. synthetic data
  - Point estimates and 95% CIs in synthetic data closely mirrored real data
  - Prediction metrics of synthetic data were comparable to real data

Implications For Veterans and the VA

- VA EHR could be used to identify MCI patients at highest risk of developing dementia for early treatment or clinical trial recruitment

  -> cost-effective care

  -> improve clinical trial success rate, reduce number of participants and cost

- Wide access to synthetic data, with minimized privacy risk, to create, test and verify models that could then be validated internally with real data

- Future Directions
  - Improve model by expanding with available EHR data elements
  - Need to discover unidentified non-traditional factors behind the overwhelming role of aging in dementia conversion
  - Further validation of synthetic data in non-linear or ML models

Acknowledgments

- Funding and Support: Phoenix VA Office of Research (Dr. Aguayo, Dr. Permana, Dr. Thompson), NSF award no. IIS 2124127
- VA Office of Healthcare Innovation and Learning (ARCHES platform)
- VA CIPHER (phenomic classification)