

Supplementary Material

Rule-based identification of individuals with mild cognitive impairment or Alzheimer's disease using clinical notes from the United States Veterans Affairs Health System

APPENDIX: Iterative Process for Clinical Note Identification

Removal of clinical notes with pharmacy-related titles

All clinical notes with pharmacy-related titles were excluded due to the nature of standardized forms/notes in Veteran's Affairs Health System (VAHS) pharmacies and the determination that drug indications were not reliable surrogates for an MCI or AD diagnosis (especially given off-label usage for other dementia/conditions).

MCI

For the MCI searches, using the search "mild cognitive impairment" with pharmacy-related notes excluded resulted in a high yield of relevant notes (PPV, 97%; **Table S1**). In contrast, using the search term "MCI" it was found that almost half the notes yielded were actually referring to the term "millicurie" (mCi). Filtering the "MCI" notes using a case-sensitive search eliminated all notes matching "mCi". In order to eliminate any remaining notes that were referring to millicurie rather than mild cognitive impairment, but were abbreviating the term in all caps, additional exclusions were applied, such as removing notes with titles associated with the term millicurie such as "radiology" and "nuclear medicine" (**full list of exclusions in Table S2**); the final PPV for the "MCI" search term, after application of all exclusions was 86% (**Table S1**).

AD

For the AD searches, in addition to the removal of the pharmacy-related notes, we determined which of the notes identified by the “Alz*” should be excluded as follows: (1) AD notes were separated into statements with ± 50 characters near “Alz*”; (2) review of these statements identified 5 categories for exclusions: family history, standardized risk assessment language, drug indication in drug lists, facilities/services, and rule out; (3) notes containing the “Alz*” term only within these categories were excluded. The “Alz*” keyword also identified Veterans and clinicians with a first or last name beginning with A-L-Z; while this was uncommon, these notes were also excluded. The PPVs for “Alz*” ranged from 83% to 99% depending on the exclusions applied (**Table S1**).

Removal of redundant or discrepant data

Potential redundancies and discrepancies in the data were handled as follows: (1) Veterans with notes meeting iterative criteria for MCI and AD were included in both cohorts (typically this was because notes appeared in different years over the entire data extraction period) (2) Veterans who had notes identified via both of the MCI-related key word searches were only counted once.

Table S1. Positive Predictive Value (PPV) for MCI and AD Keywords/Inclusion Criteria

Description of FY10-19 clinical notes from Veterans at least 50 years old at the time of the note (before denominator limitation)	Notes (n)	Veterans (n)	PPV ^a (%)
“Mild cognitive impairment” key word search^b	1,760,080	273,577	97%
Contains key word and diagnostic code (331.83; G31.84) within 8 words before/after	221,300	29,699	98%
Contains key word and on “problem list” within 300 words before/after	56,465	13,976	93%
“MCI” key word search^c	470,488	140,984	86%
“Alz*” key word search^d	4,224,086	570,557	83%
Contains key word and diagnostic code (331.0; G30.*) within 8 words before/after	236,531	35,771	99%
Contains key word and on “problem list” within 300 words before/after	136,895	38,074	93%

AD, Alzheimer’s disease. MCI, mild cognitive impairment; PPV, positive predictive value.

^aFinal PPV calculations were based on 500 notes/Veterans; exception Alz* key word search PPV calculations were based on 100 notes/Veterans.

^b “Mild cognitive impairment” key word search with pharmacy titles excluded; since PPV was well above the minimum threshold, additional exclusions were unnecessary.

^c “MCI” key word search with pharmacy title excluded as well as MCI-specific exclusions (**Appendix; Table S2**).

^d “Alz*” key word search with pharmacy titles excluded as well as AD-specific exclusions (**Appendix**).

Table S2. Exclusions for “MCI” Keyword

Key word-based exclusions	Title-based exclusions
-“MCI of Tc” (6,372 notes) -“Tc99m” near “MCI” within 4 words (5,087 notes) -“Tc 99m” near “MCI” within 4 words (13,898 notes) -“99m-tc” near “MCI” within 4 words (889 notes) -“Dosage” near “MCI” within 4 words (11,417 notes) -“Dose” near “MCI” within 4 words (11,160 notes) -“IV” near “MCI” within 4 words (3,326 notes) -“intravenously” near “MCI” within 4 words (3,809 notes) -“sestamibi” near “MCI” within 4 words (21,776 notes) -“technetium” near “MCI” within 4 words (16,883 notes) -“myoview” near “MCI” within 4 words (3,893 notes) -“dobutamine” near “MCI” within 4 words (27 notes) -“regadenoson” near “MCI” within 4 words (383 notes) -“radiopharmaceutical” near “MCI” within 4 words (4,634 notes) -“thallous chloride” near “MCI” within 4 words (1,420 notes) -“RAI” near “MCI” within 4 words (1,011 notes) -“RAIA” near “MCI” within 4 words (58 notes) -“131” near “MCI” within 4 words (2,001 notes) -“131-I” near “MCI” within 4 words (2,014 notes) -“I131” near “MCI” within 4 words (419 notes) -“I-131” near “MCI” within 4 words (1,395 notes) -“RAI-131” near “MCI” within 4 words (20 notes) -“Iodine” near “MCI” within 4 words (335 notes) -“F18” near “MCI” within 4 words (453 notes) -“18-F-FDG” near “MCI” within 4 words (1,650 notes)	- Pharmacy - Nuclear medicine - Dermatology - Radiology - Endocrinology - Cardiology

MCI, mild cognitive impairment.

Table S3. Distribution of notes and Veterans by Fiscal Year (FY) with clinical notes containing the keyword (A) “mild cognitive impairment”; or (B) “MCI”; or (C) “Alz*”.

A.

FY	Raw Clinical Notes Sample^a	Raw Veteran Sample^a	Final Clinical Notes Sample^b	Final Veteran Sample^b
2010	93,157	32,344	85,104	29,268
2011	116,041	37,550	106,052	34,291
2012	143,392	45,108	131,925	41,253
2013	169,094	48,932	156,305	44,789
2014	176,571	50,038	163,488	45,781
2015	186,773	53,271	173,208	48,884
2016	213,347	62,260	198,805	57,398
2017	232,732	65,903	218,209	61,150
2018	258,382	72,239	243,355	67,294
2019	294,992	79,437	279,911	74,609
Total	3,104,409	472,862	1,756,362	272,754

B.

FY	Raw Clinical Notes Sample^a	Raw Veteran Sample^a	Final Clinical Notes Sample^b	Final Veteran Sample^b
2010	94,995	62,998	24,966	13,639
2011	103,675	67,485	30,244	15,636
2012	104,463	66,074	65,172	14,533
2013	110,003	67,437	41,193	20,108
2014	118,904	72,857	43,478	21,337
2015	128,446	77,639	47,570	22,632
2016	137,408	81,565	51,819	24,131
2017	145,533	85,049	56,636	25,850
2018	158,925	89,549	64,972	28,967
2019	173,301	96,519	73,636	32,312
Total	1,275,653	532,035	469,686	140,640

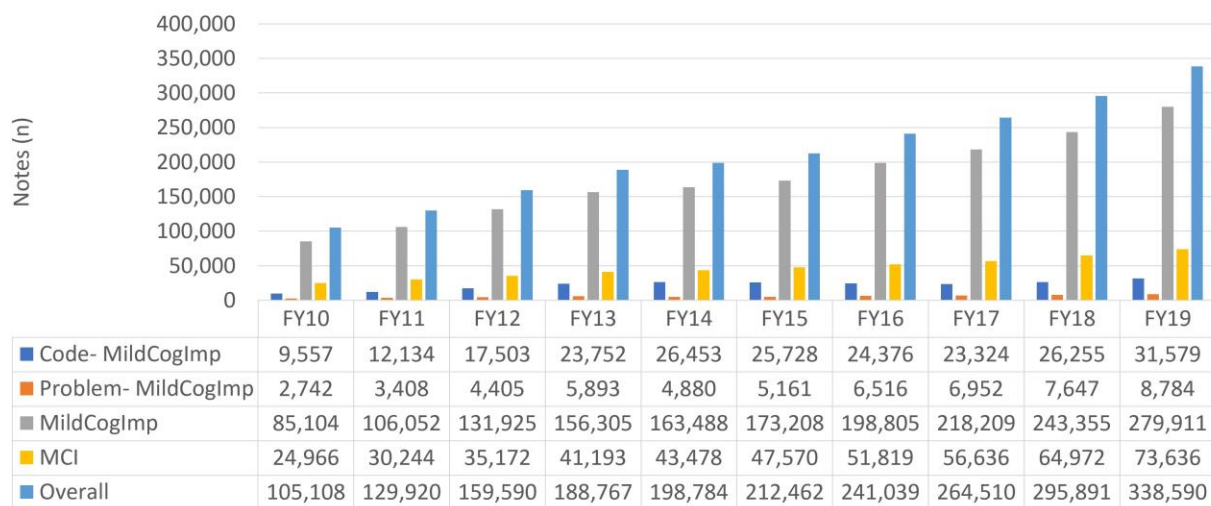
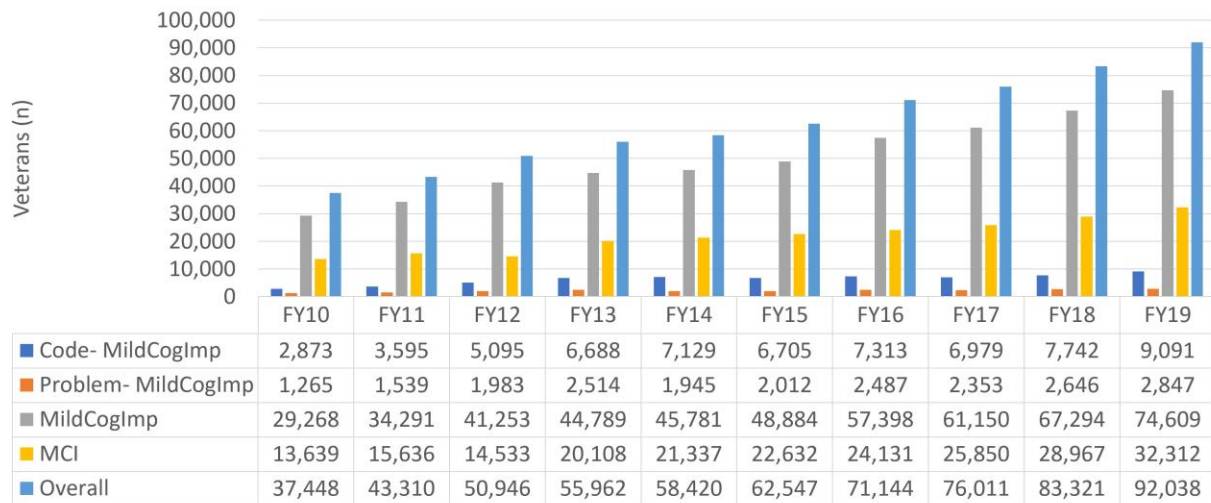
C.

FY	Raw Clinical Notes Sample^a	Raw Veteran Sample^a	Final Clinical Notes Sample^b	Final Veteran Sample^b
2010	655,148	225,409	388,673	103,877
2011	678,447	233,414	395,675	104,896
2012	809,976	264,526	395,751	100,174
2013	809,037	264,247	391,357	100,691
2014	792,473	247,734	396,049	101,097
2015	793,004	252,917	413,914	104,181
2016	872,638	270,210	451,134	109,668
2017	851,493	273,936	462,658	108,560
2018	869,124	276,772	467,744	108,036
2019	888,390	280,156	468,978	107,511
Total	8,019,730	1,301,690	4,231,933	572,063

FY, fiscal year; MCI, mild cognitive impairment.

^aThe raw clinical notes sample contained the keyword and were from Veterans at least 50 years old at the time of the note within the FY; the raw Veterans sample was comprised of Veterans whose notes were in the raw clinical notes sample.

^bThe final clinical notes sample met iterative criteria and were from Veterans with an inpatient/outpatient encounter in the FY; the final Veterans sample was comprised of Veterans whose notes were in the final clinical notes sample.



Min # of overall notes = 1, Max # of overall notes = 725, Avg # of overall notes = 6

Supplemental Figure 1. Fiscal Year distribution of Veterans/notes containing MCI keywords

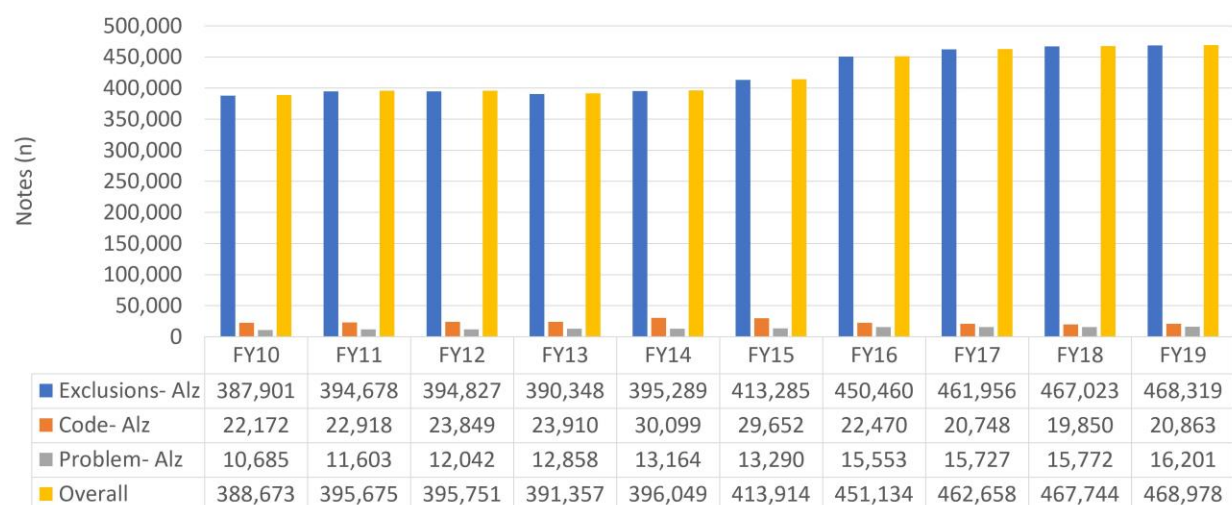
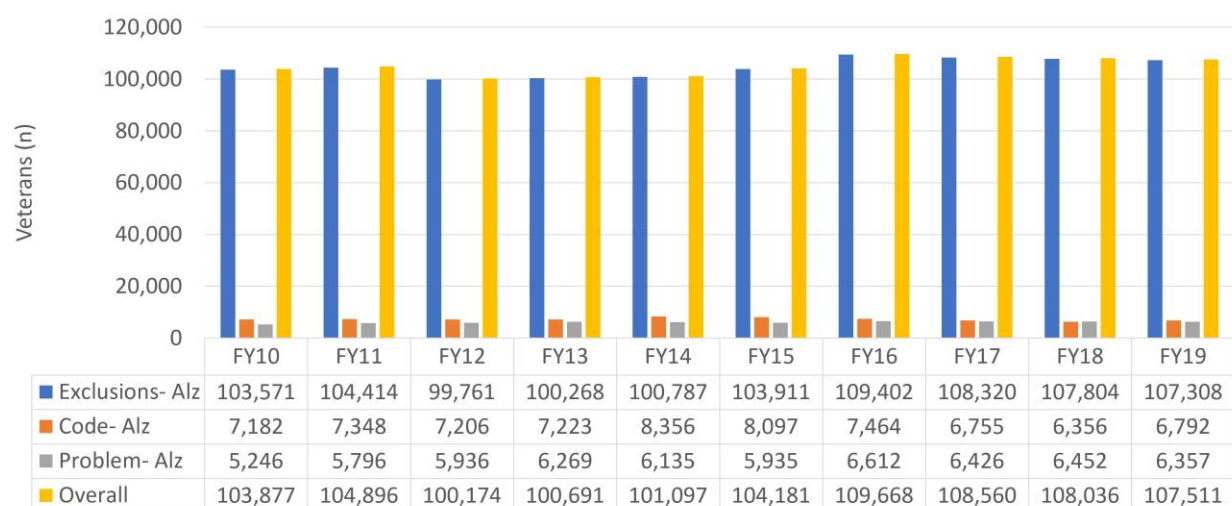
Code- MildCogImp = ICD9/10 code and “mild cognitive impairment” keyword

Problem- MildCogImp = “Problem List” and “mild cognitive impairment” keyword

MildCogImp = “mild cognitive impairment” keyword

MCI = “MCI” keyword

Overall = any keyword



Min # of overall notes = 1, Max # of overall notes = 2,105, Avg # of overall notes = 7

Supplemental Figure 2: FY distribution of Veterans/notes containing AD keywords

Exclusions = "Alz*" keyword

Code = ICD9/10 code and "Alz*" keyword

Problem = "Problem List" and "Alz*" keyword

Overall = any keyword