BMJ Open

Reliability of Coded Data to Identify Earliest Indications of Cognitive Decline, Cognitive Evaluation, and Alzheimer's Disease Diagnosis: A Pilot Study in England

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019684
Article Type:	Research
Date Submitted by the Author:	18-Sep-2017
Complete List of Authors:	Dell'Agnello, Grazia; Eli Lilly Italia SpA Desai, Urvi; Analysis Group, Inc., Kirson, Noam; Analysis Group, Inc. Wen, Jody; Analysis Group, Inc. Meiselbach, Mark; Analysis Group, Inc. Reed, Catherine; Eli Lilly and Company (Lilly UK) Belger, Mark; Eli Lilly and Company (Lilly UK) Lenox-Smith, Alan; Eli Lilly and Company (Lilly UK) Martinez, Carlos; Institute for Epidemiology, Statistics and Informatics GmbH Rasmussen, Jill; psi-napse
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Medical management, Mental health, Health services research
Keywords:	Clinical Practice Research Datalink, medical coding, text-based data, Alzheimer's disease

Reliability of Coded Data to Identify Earliest Indications of Cognitive Decline, Cognitive Evaluation, and Alzheimer's disease Diagnosis: A Pilot Study in England

Authors:
Grazia Dell'Agnello, PhD[1]; Urvi Desai, PhD[2]; Noam Y. Kirson, PhD[2]; Jody Wen, BS[2]; Mark K Meiselbach, BS[2]; Catherine C Reed, PhD[3]; Mark Belger, BSc[3]; Alan Lenox-Smith, MBBS, FFPM, FRCP[4]; Carlos Martinez, MD[5]; Jill Rasmussen, MBChB, FRCGP, FFPM[6]

Affiliations:
[1] Eli Lilly Italia SpA, 50019 Sesto Fiorentino (FI), Italy

[2] Analysis Group, Inc., 111 Huntington Ave, 14th floor, Boston, MA 02199

[3]Eli Lilly and Company (Lilly UK), Windlesham, Surrey, UK

[4]Eli Lilly and Company (Lilly UK), Priestley Road, Basingstoke, UK

[5]Institute for Epidemiology, Statistics and Informatics GmbH, Frankfurt, Germany

[6]psi-napse, Dorking, Surrey, UK

Corresponding Author:

Urvi Desai, PhD

Analysis Group, Inc., 111 Huntington Avenue, 14th floor, Boston, MA 02199

Phone: (617)425-8315

Fax number: (617)425-8001

Email: urvi.desai@analysisgroup.com

Word Count: 3,623 (not including abstract, summary of strengths and limitations, ethics statement, data availability, competing interests, funding, author contributions, and references).

Exhibits: 5 for the main document, 2 for Appendix

Abstract (294/300 Words; excluding section headers)

Objectives: To evaluate the feasibility of using diagnosis codes and prescription data to identify timing of symptomatic onset, cognitive assessment, and diagnosis of Alzheimer's disease (AD).

Methods: This was a retrospective cohort study using the UK Clinical Practice Datalink (CPRD). The study cohort consisted of a random sample of 50 patients with a first diagnosis of AD in 2010-2013. Additionally, patients were required to have a valid text-field code and a hospital episode or a referral in the 3 years before the first AD diagnosis. The earliest indications of cognitive impairment, cognitive assessment, and AD diagnosis were identified using two approaches: 1) using an algorithm based on diagnostic codes and prescription drug information, 2) using information compiled from manual review of both text-based and coded data. The reliability of the code-based algorithm for identifying the earliest dates of the three measures described earlier was evaluated relative to the comprehensive second approach. Additionally, common cognitive assessments (with and without results) were described for both approaches.

Results: The two approaches identified the same first dates of cognitive symptoms in 33 (66%) of the 50 patients, first cognitive assessment in 29 (58%) patients, and first AD diagnosis in 43 (86%) patients. Allowing for the dates from the two approaches to be within 30 days, the code-based algorithm's success rates increased to 74%, 70%, and 94%, respectively. Mini Mental State Examination (MMSE) was the most commonly observed cognitive assessment in both approaches, however of the 53 tests performed, only 19 results were observed in the coded data.

Conclusions: The code-based algorithm shows promise for identifying the first AD diagnosis. However, the reliability of using coded data to identify earliest indications of cognitive

impairment and cognitive assessments is questionable. Additionally, CPRD is not a recommended data source to identify results of cognitive assessments.

Keywords: Clinical Practice Research Datalink, medical coding, text-based data, Alzheimer's disease

Summary of strengths and limitations

- Using enriched data elements from both structured data fields and physician notes, this study not only identified relevant medical codes and prescriptions related to timing of onset of cognitive symptoms, cognitive assessments, and AD diagnosis, but also captured an additional marker of cognitive assessment based on sequencing of clinical interactions.
- The study findings also provide important insight into the availability of results from cognitive assessments from both physician notes and coded data.
- However, the study relies on Read codes and ICD-10 codes, which do not contain information by which to confirm clinical diagnoses, severity of illness, or physician interpretation, and does not include data from memory clinics, a key setting in which cognitive assessments are conducted in England.
- Additionally, the study focuses on patients with AD who had no evidence of other dementia etiologies.
- Finally, the study utilizes data prior to 2014, so study findings may not reflect the current practices in management of patients with dementia in England.

Background

The Alzheimer's Society of the UK estimates that approximately 1% of the entire UK population currently has some form of dementia.[1] Alzheimer's disease (AD) is the most common cause of dementia and accounts for approximately 62% of all dementias in the UK. The pathophysiological changes underlying AD may develop well before a formal diagnosis, resulting in early symptoms of cognitive impairment such as memory loss, attention deficits, impaired reasoning, poor judgment, and confusion prior to the diagnosis.[2,3,4,5]

The diagnosis of AD can be challenging, and requires assessment of cognitive, functional, and/or behavioral symptoms of patients suspected of having cognitive impairment.[6,7] Recent policy efforts in England have aimed to improve diagnosis rate and management of dementia,[8] as earlier, more accurate evaluation and diagnosis is believed to be important to improving potential health outcomes for patients and their caregivers as well as reduce the burden associated with dementia.[9] Information about use of and results from various evaluation tools – including tools for initial assessment (mainly in the primary care setting) such as the General Practitioner Assessment of Cognition (GPCOG), the Abbreviated Mental Test Score (AMTS), Six-Item Cognitive Impairment Test (6CIT), and for diagnosis (mainly in the secondary care settings) such as the Addenbrooke's Cognitive Assessment- Revised (ACE-R), Mini mental state examination (MMSE) and Montreal Cognitive Assessment (MOCA)[10,11] – can provide important insight regarding practice patterns during the screening and diagnostic process as well as severity of symptoms of cognitive impairment. However, this information may often not be captured in existing, structured, real-world data sources used to conduct observational studies. In addition, early symptoms associated with cognitive decline, such as mild memory impairment, might only be noted in free text fields that summarize physicians' notes and/or

correspondence provided by specialists evaluating these patients in secondary care settings. These supplemental data elements are generally not available to researchers, which limits the ability to identify the timing of onset of symptoms and subsequent cognitive testing.

In addition, to the best of our knowledge, no study to date has evaluated whether the information captured within these supplemental text data fields provides any additional insight over the coded data (e.g., diagnosis codes) into the timing of onset of cognitive impairment symptoms and subsequent testing among patients eventually diagnosed with AD. Previous studies assessing the validity of coded data (including but not limited to dementia diagnoses) typically relied on reviews of medical records, physician surveys, and comparisons to other data sources.[12] The objective of the present exploratory study was to assess the reliability of a code-based algorithm to identify the timing of symptomatic onset, cognitive assessment (including initial screening), and formal diagnosis of AD, as compared to the combination of codes and supplemental, non-structured physicians' notes and secondary care correspondence. An additional objective was to compare the availability of results from the cognitive assessments prior to AD diagnosis between the structured data and the anonymized text data.

Methods

Data

The study was conducted using a subset of the UK Clinical Practice Research Datalink (CPRD), which includes longitudinal observational data from general practitioner (GP) electronic health record systems in primary care practices, including medical diagnoses (using Read codes), referrals to specialists and to secondary care, testing and interventional procedures conducted in primary care, lifestyle information (e.g., smoking, exercise), and drugs prescribed

in primary care.[13] The subset consisted of patients in the CPRD with a link to hospitalizations and outpatient encounters in the Hospital Episode Statistics (HES) dataset.

Until recently (May 2015), the CPRD database also included pseudo-anonymized text fields summarizing notes entered by the GP or providers during consultations, which were made available to researchers upon special request.[12] In addition, it is possible to request de-identified secondary care correspondence received by the GPs. These correspondences provide supplemental information regarding the patient's encounters in secondary care settings such as hospitals.

Sample selection

The population for this pilot study was selected in two steps. In Step 1, a cohort of patients with earliest indication of AD in 2010-2013, who were eligible for linkage to HES and were continuously enrolled in active CPRD practice for ≥12 months before the first AD diagnosis, were selected. Indication of AD was defined as the first Read code, ICD-10 code, or prescription medication for AD. Patients were required to have no records with diagnosis of other types of dementia (e.g., vascular dementia) between or after the two most recent records indicating AD.

In order to ensure that the cohort of patients with AD had at least one encounter where all data elements, including physician notes and correspondence from secondary care settings, may be available, all patients were required to have ≥1 consultation record with a non-missing, non-zero text identifier and ≥1 HES record or ≥1 referral record indicating a visit to a specialist (e.g., psychiatrist, neurologist, geriatrician) in the three years prior to the first AD diagnosis.

To facilitate detailed examination of linked free text information, a sample of 50 patients was randomly drawn from the cohort meeting the criteria in Step 1 for further analysis. A

random sampling approach was used to increase the likelihood that the sub-sample selected was representative of the overall cohort identified in Step 1.

Development of the code-based algorithm

Earliest indications of symptoms of cognitive decline (e.g., "memory loss symptom"), cognitive assessment (for either screening or diagnosis), and AD diagnosis were identified using two parallel approaches. In the first approach, the Read codes, ICD-10 codes, and prescription medications indicated to treat AD found in the structured part of CPRD from up to 3 years prior to the AD diagnosis were reviewed and categorized into an algorithm to establish first observed dates of the three key time points in the pathway of progression from onset of symptoms to AD diagnosis.

In the second approach, in addition to the diagnosis codes, a targeted search of the pseudo-anonymised text data and additional correspondence provided by the GPs was conducted to identify key phrases suggestive of the earliest markers of symptoms related to cognitive impairment (e.g., "memory loss", "cognitive impairment", "confusion", etc., and their variants), cognitive assessments (e.g., "GPCOG", "MMSE", "MOCA", "mini-mental", etc., and their variants) and AD diagnosis. The targeted search was conducted by two independent reviewers to account for any subjective interpretation of the free-text.

Based on preliminary data inspection and the combined manual review of the text and structured data for 15 of the 50 patients, the definition of cognitive assessment using the structured data was refined to include an additional marker based on referrals. Specifically, given that clinical evaluation for dementia is usually undertaken by secondary care mental health specialists (e.g., geriatricians, old age psychiatrists, neurologists)[14] several weeks after the initial referral,[8] it was determined that a combination of codes indicating referral to a specialist and a

letter from specialist within 3 months after the referral could be considered as indication of cognitive assessment. In addition, it was assumed that the earliest indication of cognitive assessment could not precede the earliest symptom of cognitive impairment.

Appendix Table 1 describes the final code-based algorithm used for quality evaluation.

Quality evaluation of the reliability of the code-based algorithm

The findings from the two approaches were compared to quantify the differences in dates for the first indicators of cognitive/functional symptoms, assessments, and AD diagnosis as identified by the code-based algorithm and manual review. Additionally, the percent of patients for whom the dates of each of the three measures (indicator for cognitive impairment symptoms, cognitive assessments, and AD diagnosis) identified by the code-based algorithm were after the dates suggested by the second approach (suggesting the code-based algorithm was less sensitive) were calculated. Similarly, the proportions of patients for whom the dates of the three measures as identified by the code-based algorithm were before the dates identified by the second approach (suggesting the code-based algorithm was more sensitive) were reported. While exact matches were preferred for all analyses, in order to account for delays between the receipt of a letter from the specialist assessing the patient and the corresponding coding of the information in CPRD, a similar metric allowing for a 30-day gap between the dates identified by the two approaches was also measured. Note that for the purpose of the analysis, if an event was not observed for both approaches, it was considered an exact match. However, if a date was identified only in the manual review and not in the code-based algorithm, then the code-based algorithm was considered less sensitive. Similarly, if a date was identified in the code-based algorithm but not in manual review, the code-based algorithm was considered more sensitive.

Additionally, the days between the dates of first symptom of cognitive impairment and first cognitive assessment, and between the first cognitive assessment and the first AD diagnosis were compared for the two approaches. Congruence between the two data sources with regards to recording the type of and results from the specific type of the cognitive assessments performed prior to AD diagnosis was described.

Results

Sample characteristics

Overall, 18,281 patients in the CPRD had an indication of AD (based on diagnosis codes or AD-related medications) in 2010-2013 (See Figure 1). Of these, 12,252 (67%) patients had their first indication of AD in 2010-2013; 11,151 had no indications of another type of dementia between or after AD diagnoses. Of these 11,151 patients, 4,515 (40%) patients had evidence of both text-field data and receipt of care in secondary settings in the 3 years prior to the first AD diagnosis. The final sample comprised 1,937 patients who met all the inclusion and exclusion criteria (mean age 82 years, 38% males). The random sample of 50 patients (selected from the 1,937 patients meeting all selection criteria) included in additional analyses had similar demographic characteristics as the 1,937 patients (mean age 82 years, 42% males). These 50 patients had a total of 2,051 records with valid pseudo-anonymized text field data and 44 correspondences from secondary care, provided by CPRD upon request.

Comparison of findings from the two approaches

Of the 50 patients included in the sample, the code-based algorithm identified 48 patients with evidence of cognitive impairment prior to AD diagnosis and 42 with evidence of cognitive assessment prior to AD diagnosis. An additional 2 and 4 patients respectively had evidence of cognitive impairment and cognitive assessment on the same date as the AD diagnosis. The

remaining 4 patients had no record of cognitive assessment prior to or on the same date as the AD diagnosis (Appendix Figure 1). For the second, comprehensive approach which utilized information from all available data elements including text-based data, the number of patients with cognitive impairment and cognitive assessments prior to AD diagnosis were 49 and 43 respectively, and the numbers of patients with the same dates for these metrics as the AD diagnosis were 1 and 4 respectively. No record of cognitive assessment was observed prior to or on the same date as the AD diagnosis for 3 patients (Appendix Figure 1).

With regards to the timing of the three key events, relative to the second approach, the code-based algorithm was able to identify exact matches for the first date of symptoms associated with cognitive impairment in 33 (66%) of the 50 patients, first cognitive assessment in 29 (58%) patients, and first AD diagnosis in 43 (86%) patients (Table 1). Allowing for matches within 30 days, the algorithm's success rates increased to 74%, 70%, and 94%, respectively, for the dates of first cognitive impairment symptom, first cognitive assessment, and first AD diagnosis. Differences in the dates detected by the code-based algorithm relative to the more comprehensive approach were mainly a result of more false negatives generated with the algorithm. There was only 1 patient (2% of the sample), for whom, the date of first symptoms of cognitive impairment identified by the algorithm was earlier than the date identified by the second, comprehensive approach, suggesting the algorithm was more sensitive. The results were similar even after allowing for a 30-day gap in the dates identified by the two approaches. With respect to identifying the dates of first cognitive assessment the code-based algorithm was found to be more sensitive than the comprehensive approach in 8 patients (16%) based on exact matches and 4 patients (8%) allowing for matches within 30 days. The differences in the

detection of the first date of AD diagnosis between the code-based algorithm and manual review based on either exact matches or matches within 30 days were very small.

Table 1: Differences in dates of earliest indications of cognitive impairment, cognitive assessment, and AD diagnosis as identified by coded-data vs. comprehensive data review (N=50)

	First symptom	First cognitive assessment	AD diagnosis
Date matches with manual review, n (%)			
Exact matches	33 (66.0%)	29 (58.0%)	43 (86.0%)
Matches ± 30 days	37 (74.0%)	35 (70.0%)	47 (94.0%)
Characteristics of mismatches, n (%)			
Code-based algorithm more sensitive than manual review	1 (2.0%)	8 (16.0%)	0 (0.0%)
Code-based algorithm more sensitive than manual review (< -30 days)	1 (2.0%)	4 (8.0%)	0 (0.0%)
Code-based algorithm less sensitive than manual review	16 (32.0%)	13 (26.0%)	7 (14.0%)
Code-based algorithm less sensitive than manual review (> + 30 days)	12 (24.0%)	11 (22.0%)	3 (6.0%)

Abbreviation: AD = Alzheimer's disease

Notes:
Manual review included the review of both structured data and text-based data; cases where dates were not observed by either approach (n=2 for cognitive assessment only) were considered exact matches; if the algorithm generated a date value that either preceded the equivalent date in the manual review or for which an equivalent date in the manual review as not observed, it was considered as having resulted in a false positive, suggesting the algorithm was more sensitive than the manual review.

Additionally, the code-based algorithm and the comprehensive review of all data elements returned qualitatively similar gaps between the dates of first symptom of cognitive impairment and first cognitive assessment, and between the first cognitive assessment and the first AD diagnosis. For both approaches, the median time between first symptom and cognitive assessment was under 6 weeks (37 days for the manual review and 14 days for the algorithm) whereas the median time between the first cognitive assessment and the first AD diagnosis was

between 6-7 months (214 days for the manual review and 181 days for the algorithm) (Figures 2 and 3).

In terms of the specific types of cognitive assessments performed prior to AD diagnosis, 34 (68%) patients had information available on the type of cognitive assessments conducted. Among these, very few patients received screening-type evaluations: 3 patients received the AMTS, 5 patients received the 6CIT, and 1 patient received GPCOG (Table 2). The more detailed evaluations captured in the data included the ACE-R (5/50 patients) and the MMSE (30/50 patients; a total of 53 assessments). A total of 9 patients received multiple tests prior to AD diagnosis, primarily in addition to ≥1 MMSE assessment (Table 2). For the most commonly administered cognitive assessment – the MMSE – the results were largely captured only in the supplemental (text-based) data. Specifically, 38 out of the 53 assessments had valid test scores available in the text-based data, only 6 of which were available and were consistent in both data sources. Additional 13 scores were observable only in the structured portion of the data, and neither data source reported scores for the remaining two assessments.

Table 2: Descriptive characteristics of cognitive assessments in the three years prior to AD diagnosis (N=50)

Cognitive testing characteristic	n (%)
Any cognitive test	34 (68.0%)
Type of cognitive test	
General Practitioner Assessment of Cognition (GPCOG)	1 (2.9%)
Abbreviated Mental Test Score (AMTS)	3 (8.8%)
Six-item cognitive impairment test (6CIT)	5 (14.7%)
Addenbrooke's Cognitive Examination - Revised (ACE-R)	5 (14.7%)
Mini-mental State Examination (MMSE)	30 (88.2%)
Multiple MMSE tests	14 (46.7%)
Multiple tests of different types	9 (26.5%)
MMSE + ACE-R	3 (33.3%)
MMSE + AMT	2 (22.2%)
MMSE + 6CIT	2 (22.2%)
6CIT + GPCOG	1 (11.1%)
MMSE + ACE-R + AMTS	1 (11.1%)

Abbreviation: AD = Alzheimer's disease

Discussion

The results of this pilot study suggest that the information captured within the supplemental text-based data fields provide increased accuracy over the structured portion of CPRD data regarding the dates of first symptom of cognitive impairment, first cognitive assessment, and first AD diagnosis. The comparison between the code-based algorithm developed in this study and a manual review of a patient's medical history (including structured data, free text, and correspondence from secondary care settings) suggests that the concordance between the two is highest for identifying the first AD diagnosis, with diminishing effectiveness of the code-based algorithm in identifying the earliest symptoms of cognitive impairment and first cognitive assessment, respectively. Additionally, nearly two-thirds of the 50 patients

included in the study had records indicative of specific types of cognitive assessments prior to or concomitantly with their AD diagnoses. For the cognitive assessment captured most commonly in the data, the MMSE, the test results were available in the text-based data for 38 of the 53 assessments, whereas the results for 13 assessments were captured only in the coded data, and the scores for the remaining 2 assessments were not available in either data source. This suggests that although the text-based data elements are more likely to capture this information, neither the coded data, nor the additional information captured in physician notes and secondary care sources provide a comprehensive view of the detailed results of cognitive assessments. This may in part be due to the fact that much of the cognitive evaluation in England is done in specialty clinics such as memory clinics and the detailed data regarding the use of and findings from cognitive assessments may not be transferred back to the GPs. Even if the information is transferred back, it may not be entered into the system. However, given the recent initiatives to increase awareness about recognizing and recording symptoms of cognitive decline within the GP setting in England (especially in populations at increased risk for dementia),[8,11] and improve care-coordination as well as documentation across different provider settings,[15,16] the quality and completeness of data recording are likely to improve in the future, which could increase the reliability of the code-based algorithm. The improved quality of the recorded data would also facilitate identification of symptoms of cognitive impairment sooner, and facilitate real-world research into implications of earlier identification of cognitive impairment on subsequent outcomes in the UK.

Study strengths and limitations

The study used data from both the structured portion of CPRD and the text fields reflecting rich, additional information from notes captured by physicians/specialists during

consultation. Using these enriched data elements, this study developed a code-based algorithm based on the findings from an intensive manual review process independently conducted by two reviewers. In doing so, we not only identified relevant medical codes and prescriptions to identify timing of onset of cognitive symptoms, cognitive assessments, and AD diagnosis, but also captured an additional marker of cognitive assessment based on sequencing of clinical interactions. In addition, the study provides important insight into the availability of results from cognitive assessments, in particular MMSE, from both physician notes and coded data.

However, this study also has a number of limitations. First, the study relies on the Read codes (Primary Care) and ICD-10 codes (secondary care) used within the CPRD and HES administrative records datasets, respectively. These codes are retrieved from electronic health records and hospital admission records and do not contain information by which to confirm clinical diagnoses, severity of illness, or physician interpretation. Accordingly, it is possible that some patients identified as having been diagnosed with AD, with no recorded diagnosis of other type, have other dementia etiologies instead.[17] In addition, for this study, though we reviewed the correspondence from secondary care, we did not have access to data from memory clinics, which is a key setting in which cognitive assessments are conducted in England. Future research should identify avenues to compare the reliability of the algorithm relative to data captured in these settings as well. This study is also limited in sample size, as the algorithm was only developed and assessed for 50 randomly selected patients who were diagnosed with AD. In addition, the algorithm may not capture all Read codes and ICD-10 codes indicative of symptoms of cognitive impairment, cognitive assessment, and AD diagnosis. As such, additional research using larger patient populations is necessary to further test the reliability and generalizability of the algorithm. Furthermore, the study was focused on patients with AD who had no evidence of

other dementia etiologies, and further research is needed to assess the reliability of the coded data for identifying the timing of cognitive impairment, cognitive assessment, and diagnosis among patients with other dementia etiologies. Finally, the study utilized data prior to 2014 and the study findings may not reflect the current practices in management of patients with dementia in England.

Conclusions

Given the limited expected future availability of free text data and secondary care correspondence in CPRD, the code-based algorithm developed using data for a small sample of AD patients shows promise as a feasible alternative for identifying the earliest indications of AD. However, the reliability of using coded data to identify earliest symptoms of cognitive impairment as well as indications of cognitive assessments prior to AD diagnosis is limited. The use of coded data, in its present form, is not recommended for identifying information regarding the specific types of cognitive assessments performed, the specialty of physicians performing the assessments or the results associated with those assessments (e.g., to assess disease severity levels).

Ethics approval and consent to participate

This study was approved by the Independent Scientific Advisory committee (ISAC): Protocol # 16_043R.

Availability of data and materials

This study used the Clinical Practice Research Datalink, provided by CPRD. Per the data use agreement, the datasets supporting the conclusions of this article cannot be made available to researchers outside of the study team. However, interested readers may request the data directly from CPRD – see https://www.cprd.com/researcher/ for more information.

Competing interests

GD, CCR, MB, and ALS are full-time employees of Eli Lilly and Company. NYK, UD, JW, and MKM are employees of Analysis Group, Inc., a company that received funding from Eli Lilly and Company for this research. CM and JR are consultants to Eli Lilly and Company.

Funding

This study was funded by Eli Lilly and Company.

Authors' contributions

NYK, UD, GD, CCR, and MB contributed to the conceptual design and reviewed and discussed the study results. JW and MKM contributed to the conceptual design and performed data analysis. ALS, JR, and CM contributed in the interpretation of study findings. All authors reviewed, edited, and approved the final manuscript.

References

1. Dementia UK: The full report. Alzheimer's Society. https://www.alzheimers.org.uk/site/scripts/download_info.php?fileID=2323 Accessed February 6, 2017.

2. Mortimer JA, Borenstein AR, Gosche KM, Snowdon DA. Very Early Detection of Alzheimer Neuropathology and the Role of Brain Reserve in Modifying Its Clinical Expression. J Geriatr Psychiatry Neurol 2005;18: 218-223.

3. Jack CR, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, Petersen RC, Trojanowski JQ. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 2010; 9:119.

4. Peterson RC. Early diagnosis of Alzheimer's disease: Is MCI too late? Curr Alzheimer Res. 2009;6(4):324-330.

5. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute of Aging and the Alzheimer's Association workgroup. Alzheimer's & Dementia 2011;1-7.

6. National Institutes of Health. National Institutes on Aging. Alzheimer's disease diagnosis. https://www.nia.nih.gov/alzheimers/topics/diagnosis#how Accessed February 6, 2017.

7. NHS Choices. Alzheimer's disease diagnosis. http://www.nhs.uk/Conditions/Alzheimers-disease/Pages/Diagnosis.aspx Accessed February 6, 2017.

8. Government of UK. Prime Minister's challenge on dementia 2020. 2015. https://www.gov.uk/government/publications/prime-ministers-challenge-on-dementia-2020/prime-ministers-challenge-on-dementia-2020 Accessed February 6, 2017.

9. Dubois B, Padovani A, Scheltens P, et al. Timely Diagnosis for Alzheimer's Disease: A Literature Review on Benefits and Challenges. J Alzheimers Dis 2015;49(3):617-631.

10. NHS Department of Health. Dementia revealed: what primary care needs to know. 2014. https://www.england.nhs.uk/wp-content/uploads/2014/09/dementia-revealed-toolkit.pdf Accessed February 6, 2017.

11. NHS England. Dementia diagnosis and management: a brief pragmatic resource for general practitioners. 2015. https://www.england.nhs.uk/wp-content/uploads/2015/01/dementia-diag-mng-ab-pt.pdf Accessed February 6, 2017.

12. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Reserach Database: a systematic review. Br J Gen Pract. 2010 Mar;60 (572):e128-e136.

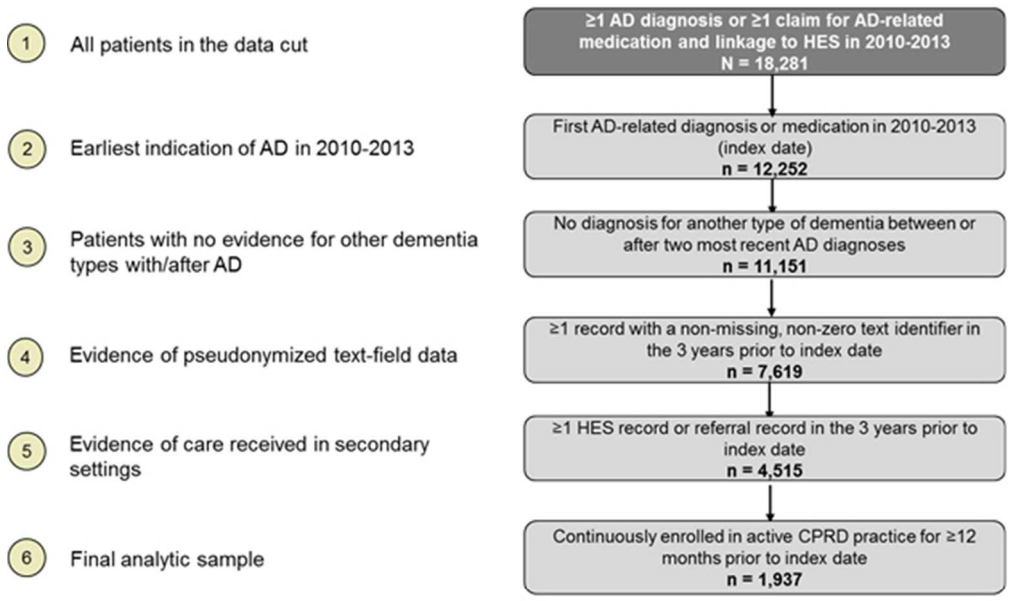
13. Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015;44(3):827–836.

14. Alzheimer's Society UK. https://www.alzheimers.org.uk/info/20071/diagnosis/95/assessment_process_and_tests/2 Accessed February 6, 2017.

15. The King's Fund. Transforming our healthcare system: ten priorities for commissioners. 2015. https://www.kingsfund.org.uk/sites/files/kf/field/field_publication_file/10PrioritiesFinal2.pdf Accessed February 13, 2017.

16. NHS National Information Board. Personalized health and care 2020: using data and technology to transform outcomes for patients and citizens. 2014. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/384650/NIB_Report.pdf Accessed February 13, 2017.

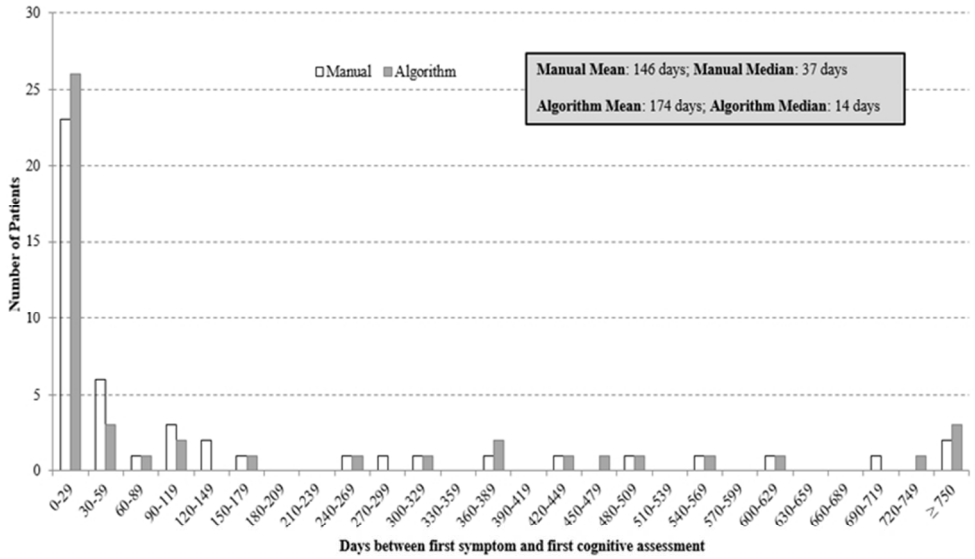
17. Happich M, Kirson NY, Desai U, et al. Excess costs associated with possible misdiagnosis of Alzheimer's disease among patients with vascular dementia in a UK CPRD population. J Alzheimers Dis 2016;53:171-183.



Abbreviations: AD = Alzheimer's disease, HES = Hospital Episode Statistics, CPRD = Clinical Practice Research Datalink

Figure 1: Sample Selection

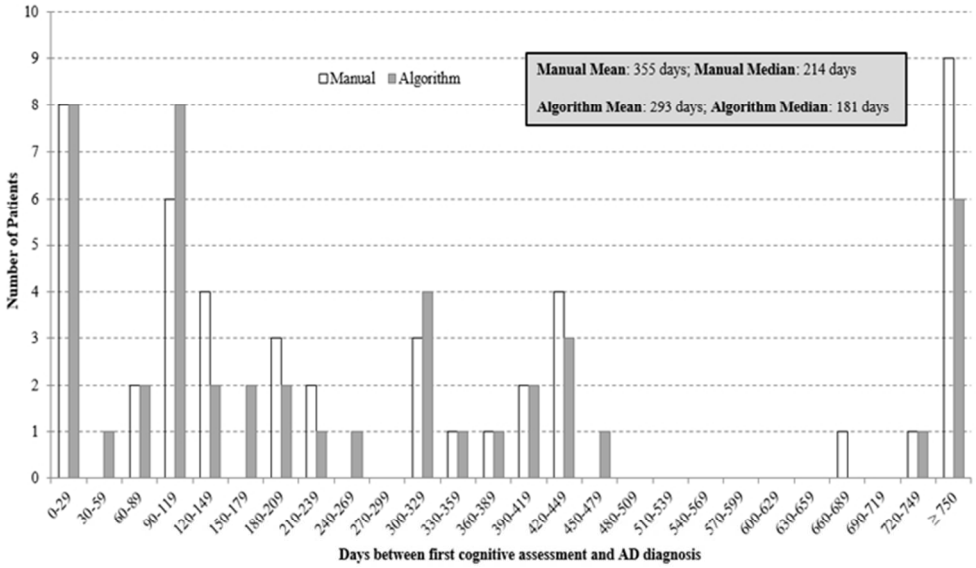
165x112mm (96 x 96 DPI)



Note:
Manual review included the review of both structured data and text-based data

Figure 2: Distribution of days between first cognitive symptom to cognitive assessment: code-based algorithm vs. comprehensive data review (N=50)

214x137mm (96 x 96 DPI)



Abbreviation: AD = Alzheimer's disease

Note:
Manual review included the review of both structured data and text-based data

Figure 3: Distribution of days between first cognitive assessment to AD diagnosis: code-based algorithm vs. comprehensive data review (N=50)

205x148mm (96 x 96 DPI)

Appendix Table 1: Final code-based algorithm to identify early indications of cognitive symptoms, cognitive assessment, and AD diagnosis

Category	Diagnosis code	Description
Read codes		
Symptom	1B1A.12	memory loss symptom
	F110.00	alzheimer's disease
	Eu00.00	[x]dementia in alzheimer's disease
	Eu02z00	[x] unspecified dementia
	28G..00	forgetful
	Eu00100	[x]dementia in alzheimer's disease with late onset
	E2A1000	mild memory disturbance
	E00z.00	senile or presenile psychoses nos
	1B1A.13	memory disturbance
	Z7CF800	poor short-term memory
	Z7C1.00	impaired cognition
	R009.00	[d]confusion
	Eu00z11	[x]alzheimer's dementia unspec
	Eu05700	[X]Mild cognitive disorder
	2841.00	Confused
	2841.11	Confusion
	1461.00	H/O: dementia
	168..14	C/O 'Muzzy head'
	1JA2.00	Suspected dementia
	28E..00	Cognitive decline
	28H..00	Mentally vague
	E00..11	Senile dementia
	E00..12	Senile/presenile dementia
	Eu01y00	[X]Other vascular dementia
	Eu02500	[X]Lewy body dementia
	F116.00	Lewy body disease
	R00z011	[D]Memory deficit
	Z7CEH14	Memory problem
Cognitive assessment	9N1T.00	seen in psychiatry clinic
	388m.00	mini-mental state examination
	388V.00	mini mental state score
	6AB..00	dementia annual review
	9N1M.00	seen in psychology clinic
	ZL9D.00	seen by psychiatrist
	9Nk1.00	seen in memory clinic

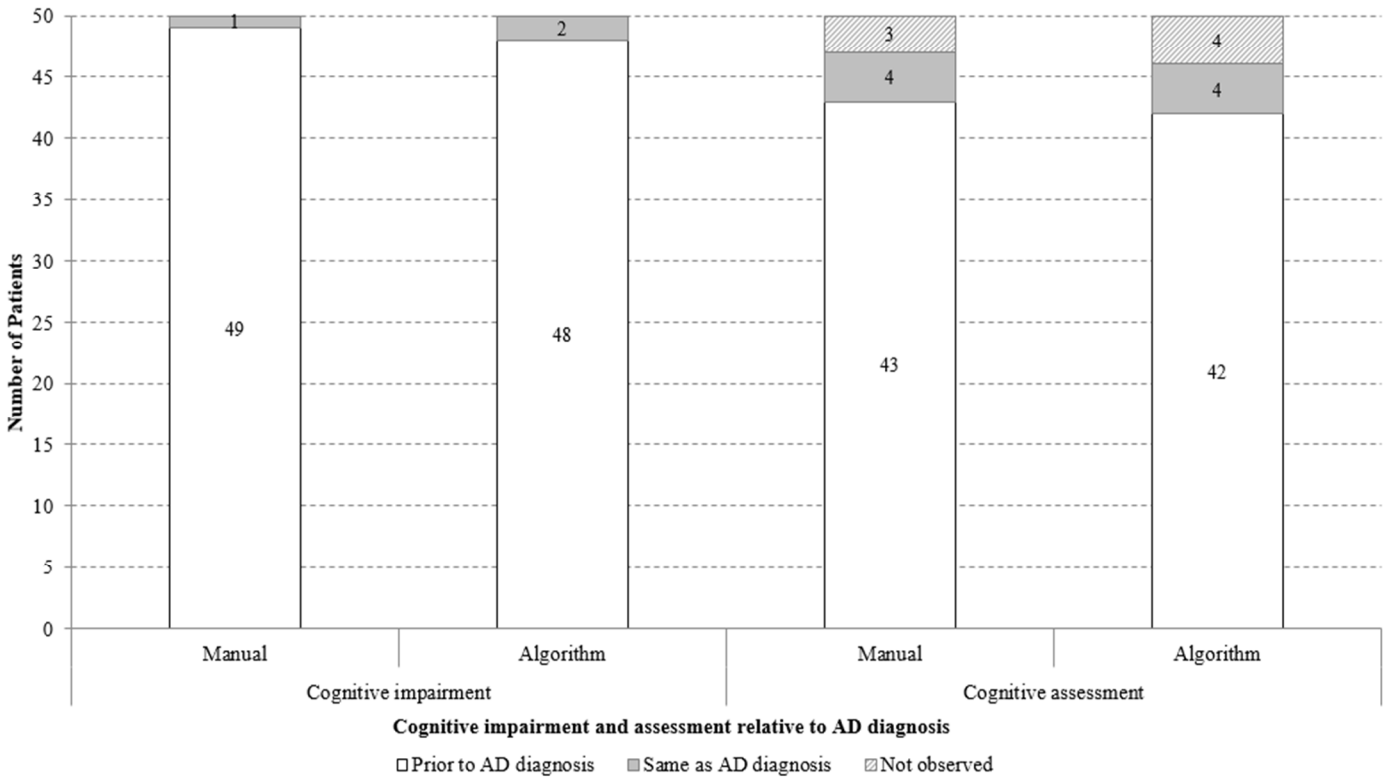
Category	Diagnosis code	Description
	3AD3.00	six item cognitive impairment test
	9Nk6.00	seen in mental health clinic
	6A6..00	mental health review
	388m.11	mmse score
	9N1R.00	seen in neurology clinic
	ZRaA.00	mini-mental state examination
	9N2a.11	Seen by CPN
	ZL9D412	Seen by old age psychiatrist
	ZQ3E.00	Mental health assessment
	3A...11	Memory assessment
	8CM2.00	Psychiatry care plan
	ZL9D400	Seen by psychogeriatrician
	38C1000	Assessment for dementia
	38Dv.00	GPCOG - general practitioner assessment of cognition
	3A...12	Dementia assessment
	3AF..00	Addenbrooke's cognitive examination revised
	66h..00	Dementia monitoring
	8A2..00	Psychiatric monitoring
	8CMZ.00	Dementia care plan
	8HLC.00	Psychogeriatric D.V. done
	9N1yA00	Seen in psychogeriatric clinic
	9NN7.00	Under care of mental health team
	ZLA2E00	Seen by psychiatric nurse
	ZLA3111	Seen by CPN
	ZLB5.00	Seen by mental health counsellor
Relevant referral	8H4D.00	Referral to psychogeriatrician
	8H47.00	Geriatric referral
	8HKC.00	Psychogeriatrics D.V. requestd
	8HTY.00	Referral to memory clinic
	8Hc..00	Referral to mental health team
	8H49.00	Psychiatric referral
	8HHo.00	Referral to older age community mental health team
	ZL5B.00	Referral to psychiatrist
Encounter	9N1C.11	Home visit
	9N33.11	Letter encounter
	9N33.00	Letter encounter from patient
	9N35.00	Letter encounter to patient
	9N36.11	Letter from consultant
	9N36.00	Letter from specialist
	8H87.00	Follow-up 1 month

Category	Diagnosis code	Description
	9NV..00	Follow-up encounter
	9N32.00	Third party encounter
	6A...00	Patient reviewed
	9N3D.00	Letter received
	2....11	Examination of patient
	9H...00	Mental health administration
	ZL9AL00	Seen by care of the elderly physician
	68Q..00	Geriatric screening
	69D1.00	Geriatric health exam.
	9N1U.00	Seen in elderly assessment clinic
	9Nk5.00	Seen in elderly care clinic
	3876.00	Multidisciplinary assessment
	3891.00	Initial patient assessment
	3Z...00	Diagnostic procedure NOS
	67...11	Counselling
	671C.00	Discussed with doctor
	68P..00	Adult screening
	68Q3.00	Geriatric 75 year screen
	9N02.00	Seen in geriatric clinic
	9N0c.00	Seen in private clinic
	9N11.00	Seen in GP's surgery
	9N1C.00	Seen in own home
	9N22.00	Seen by practice nurse
	9N2G.00	Seen by consultant
	9N2N.00	Seen by Rota Doctor
	9N2R.00	Seen by co-operative doctor
	9N2o.00	Seen by health support worker
	9N7..11	Follow-up consultation
	9NFA.00	District nurse visit
	9NY..00	Appointment
	9Na..00	Consultation
	ZL23300	Under care of district nurse
	ZV67.00	[V]Follow-up examination
Other referral	8HR1.00	Refer for ECG recording
	8H7Y.00	Refer to acupuncture
	8H77.00	Refer to physiotherapist
	8H...00	Referral for further care
	8H68.00	Referral to haematologist
	8HTb.00	Referral to male urology clinic
	8H7..12	Referral to nurse
	8H4J.00	Referred to anaesthetist

Category	Diagnosis code	Description
	8H4K.00	Referred to endocrinologist
	8H52.00	Ophthalmological referral
	8H53.00	ENT referral
	8H54.00	Orthopaedic referral
	8H43.00	Dermatological referral
	8H7R.00	Refer to chiropodist
	8H48.00	Gastroenterological referral
	8H4L.00	Referred to nephrologist
	8H58.00	Gynaecological referral
	8H59.00	Referred to plastic surgeon
	8H5B.00	Referred to urologist
	8H5D.00	Referred to vascular surgeon
	8H5J.00	Referral to colorectal surgeon
	8H72.00	Refer to district nurse
	8H7G.00	Refer to speech therapist
	8H7Q.00	Refer to surgical fitter
	8H7V.00	Refer to audiologist
	8H7X.00	Refer to podiatry
	8HBJ.00	Stroke / transient ischaemic attack referral
	8HD..00	Refer to hospital OPD
	8HH5.00	Refer to domiciliary physiotherapy
	8HHk.00	Referral to hospital phlebotomist
	8HHl.00	Referral to practice phlebotomist
	8HQ..00	Refer for imaging
	8HQ2.00	Refer for ultrasound investign
	8HQ8.00	Referral for dual energy X-ray photon absorptiometry scan
	8HR8.00	Referral for 24 hour blood pressure recording
	8HTX.00	Referral to incontinence clinic
	8HVQ.00	Private referral to rheumatologist
	8He..00	Referral to intermediate care
	8He0.00	Referral to intermediate care - hospital at home
	8Hj0.00	Referral to diabetes structured education programme
	ZL85111	Referral to community physiotherapist
Diagnosis	F110.00	alzheimer's disease
	Eu00.00	[x]dementia in alzheimer's disease
	Eu00100	[x]dementia in alzheimer's disease with late onset
	Eu00z11	[x]alzheimer's dementia unspec
ICD-10 codes		
Symptom	F03	unspecified dementia
	R418	other and unspecified symptoms and signs involving cognitive functions and awareness

Category	Diagnosis code	Description
	R54	senility
	G309	Alzheimer's disease, unspecified
	G309D	Alzheimer's disease, unspecified
	R410	Disorientation, unspecified
	F051	Delirium superimposed on dementia
	F028	Dementia in other specified diseases classified elsewhere
	F067	Mild cognitive disorder
	F99	Mental disorder, not otherwise specified
Cognitive assessment - Encounter	Z139	Special screening examination, unspecified
Diagnosis	G309	Alzheimer's disease, unspecified
	G309D	Alzheimer's disease, unspecified

Appendix Figure 1: Cognitive impairment and cognitive assessment relative to AD diagnosis



Abbreviation: AD = Alzheimer's disease

Note:
Manual review included the review of both structured data and text-based data

BMJ Open

Reliability of Coded Data to Identify Earliest Indications of Cognitive Decline, Cognitive Evaluation, and Alzheimer's Disease Diagnosis: A Pilot Study in England

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019684.R1
Article Type:	Research
Date Submitted by the Author:	22-Jan-2018
Complete List of Authors:	Dell'Agnello, Grazia; Eli Lilly Italia SpA Desai, Urvi; Analysis Group, Inc., Kirson, Noam; Analysis Group, Inc. Wen, Jody; Analysis Group, Inc. Meiselbach, Mark; Analysis Group, Inc. Reed, Catherine; Eli Lilly and Company (Lilly UK) Belger, Mark; Eli Lilly and Company (Lilly UK) Lenox-Smith, Alan; Eli Lilly and Company (Lilly UK) Martinez, Carlos; Institute for Epidemiology, Statistics and Informatics GmbH Rasmussen, Jill; psi-napse
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Medical management, Mental health, Health services research
Keywords:	Clinical Practice Research Datalink, medical coding, text-based data, Alzheimer's disease

Reliability of Coded Data to Identify Earliest Indications of Cognitive Decline, Cognitive Evaluation, and Alzheimer's disease Diagnosis: A Pilot Study in England

Authors:
Grazia Dell'Agnello, PhD[1]; Urvi Desai, PhD[2]; Noam Y. Kirson, PhD[2]; Jody Wen, BS[2]; Mark K Meiselbach, BS[2]; Catherine C Reed, PhD[3]; Mark Belger, BSc[3]; Alan Lenox-Smith, MBBS, FFPM, FRCP[4]; Carlos Martinez, MD[5]; Jill Rasmussen, MBChB, FRCGP, FFPM[6]

Affiliations:
[1] Eli Lilly Italia SpA, 50019 Sesto Fiorentino (FI), Italy

[2] Analysis Group, Inc., 111 Huntington Ave, 14th floor, Boston, MA 02199

[3]Eli Lilly and Company (Lilly UK), Windlesham, Surrey, UK

[4]Eli Lilly and Company (Lilly UK), Priestley Road, Basingstoke, UK

[5]Institute for Epidemiology, Statistics and Informatics GmbH, Frankfurt, Germany

[6]psi-napse, Dorking, Surrey, UK

Corresponding Author:

Urvi Desai, PhD

Analysis Group, Inc., 111 Huntington Avenue, 14th floor, Boston, MA 02199

Phone: (617)425-8315

Fax number: (617)425-8001

Email: urvi.desai@analysisgroup.com

Word Count: 3,753 (not including abstract, summary of strengths and limitations, ethics statement, data availability, competing interests, funding, author contributions, and references).

Exhibits: 5 for the main document, 2 for Appendix

Abstract (300/300 Words; including section headers)

Objectives: Evaluate the reliability of using diagnosis codes and prescription data to identify timing of symptomatic onset, cognitive assessment, and diagnosis of Alzheimer's disease (AD) among patients diagnosed with AD.

Methods: This was a retrospective cohort study using the UK Clinical Practice Datalink (CPRD). The study cohort consisted of a random sample of 50 patients with first AD diagnosis in 2010-2013. Additionally, patients were required to have a valid text-field code and a hospital episode or a referral in the 3 years before the first AD diagnosis. The earliest indications of cognitive impairment, cognitive assessment, and AD diagnosis were identified using two approaches: 1) using an algorithm based on diagnostic codes and prescription drug information, 2) using information compiled from manual review of both text-based and coded data. The reliability of the code-based algorithm for identifying the earliest dates of the three measures described earlier was evaluated relative to the comprehensive second approach. Additionally, common cognitive assessments (with and without results) were described for both approaches.

Results: The two approaches identified the same first dates of cognitive symptoms in 33 (66%) of the 50 patients, first cognitive assessment in 29 (58%) patients, and first AD diagnosis in 43 (86%) patients. Allowing for the dates from the two approaches to be within 30 days, the code-based algorithm's success rates increased to 74%, 70%, and 94%, respectively. Mini Mental State Examination (MMSE) was the most commonly observed cognitive assessment in both approaches, however of the 53 tests performed, only 19 results were observed in the coded data.

Conclusions: The code-based algorithm shows promise for identifying the first AD diagnosis. However, the reliability of using coded data to identify earliest indications of cognitive

impairment and cognitive assessments is questionable. Additionally, CPRD is not a recommended data source to identify results of cognitive assessments.

Keywords: Clinical Practice Research Datalink, medical coding, text-based data, Alzheimer's disease

Summary of strengths and limitations

- Using enriched data elements from both structured data fields and physician notes, this study not only identified relevant medical codes and prescriptions related to timing of onset of cognitive symptoms, cognitive assessments, and AD diagnosis, but also captured an additional marker of cognitive assessment based on sequencing of clinical interactions.
- The study findings also provide important insight into the availability of results from cognitive assessments from both physician notes and coded data.
- However, the study relies on Read codes and ICD-10 codes, which do not contain information by which to confirm clinical diagnoses, severity of illness, or physician interpretation, and does not include data from memory clinics, a key setting in which cognitive assessments are conducted in England.
- Additionally, the study focuses on patients with AD who had no evidence of other dementia etiologies.
- Finally, the study utilizes data prior to 2014, so study findings may not reflect the current practices in management of patients with dementia in England.

Background

The Alzheimer's Society of the UK estimates that approximately 1% of the entire UK population currently has some form of dementia.[1] Alzheimer's disease (AD) is the most common cause of dementia and accounts for approximately 62% of all dementias in the UK. The pathophysiological changes underlying AD may develop well before a formal diagnosis, resulting in early symptoms of cognitive impairment such as memory loss, attention deficits, impaired reasoning, poor judgment, and confusion prior to the diagnosis.[2,3,4,5]

The diagnosis of AD can be challenging, and requires assessment of multiple domains related to patients' cognition and function.[6] Some guidelines suggest evaluation of behavioral symptoms as well.[7] Recent policy efforts in England have aimed to improve diagnosis rate and management of dementia,[8] as earlier, more accurate evaluation and diagnosis is believed to be important to improving potential health outcomes for patients and their caregivers as well as reduce the burden associated with dementia.[9] Information about use of and results from various evaluation tools – including tools for initial assessment (mainly in the primary care setting) such as the General Practitioner Assessment of Cognition (GPCOG), the Abbreviated Mental Test Score (AMTS), Six-Item Cognitive Impairment Test (6CIT), and those used to inform a diagnosis (mainly in the secondary care settings) such as the Addenbrooke's Cognitive Assessment-Revised (ACE-R), Mini mental state examination (MMSE) and Montreal Cognitive Assessment (MOCA)[10,11] – can provide important insight regarding practice patterns during the screening and diagnostic process as well as severity of symptoms of cognitive impairment. However, this information may often not be captured in existing, structured, real-world data sources used to conduct observational studies.[12] In addition, early symptoms associated with cognitive decline, such as mild memory impairment, might only be noted in free text fields that summarize

physicians' notes and/or correspondence provided by specialists evaluating these patients in secondary care settings. These supplemental data elements are generally not available to researchers,[12] which limits the ability to identify the timing of onset of symptoms and subsequent cognitive testing.

In addition, to the best of our knowledge, no study to date has evaluated whether the information captured within these supplemental text data fields provides any additional insight over the coded data (e.g., diagnosis codes) into the timing of onset of cognitive impairment symptoms and subsequent testing among patients eventually diagnosed with AD. Previous studies assessing the reliability of coded data (including but not limited to dementia diagnoses) typically relied on reviews of medical records, physician surveys, and comparisons to other data sources.[13] The objective of the present exploratory study was to assess the reliability of using a code-based algorithm to identify the timing of symptomatic onset, cognitive assessment (including initial screening), and formal diagnosis of AD, as compared to the combination of codes and supplemental, non-structured physicians' notes and secondary care correspondence, among patients diagnosed with AD. An additional objective was to compare the availability of results from the cognitive assessments prior to AD diagnosis between the structured data and the anonymized text data.

Methods

Data

The study was conducted using a subset of the UK Clinical Practice Research Datalink (CPRD), which includes longitudinal observational data from general practitioner (GP) electronic health record systems in primary care practices, including medical diagnoses (using Read codes), referrals to specialists and to secondary care, testing and interventional procedures

conducted in primary care, lifestyle information (e.g., smoking, exercise), and drugs prescribed in primary care.[12] The subset consisted of patients in the CPRD with a link to hospitalizations and outpatient encounters in the Hospital Episode Statistics (HES) dataset.

Until recently (May 2015), the CPRD database also included pseudo-anonymized text fields summarizing notes entered by the GP or providers during consultations, which were made available to researchers upon special request.[13] In addition, it is possible to request de-identified secondary care correspondence received by the GPs. These correspondences provide supplemental information regarding the patient's encounters in secondary care settings such as hospitals.

Sample selection

The population for this pilot study was selected in two steps. In Step 1, a cohort of patients with earliest indication of AD in 2010-2013, who were eligible for linkage to HES and were continuously enrolled in active CPRD practice for ≥12 months before the first AD diagnosis, were selected. Indication of AD was defined as the first Read code or ICD-10 code for AD (see Appendix Table 1 for details). Patients were required to have no records with diagnosis of other types of dementia (e.g., vascular dementia) between or after the two most recent records indicating AD.

In order to ensure that the cohort of patients with AD had at least one encounter where all data elements, including physician notes and correspondence from secondary care settings, may be available, all patients were required to have ≥1 consultation record with a non-missing, non-zero text identifier and ≥1 HES record or ≥1 referral record indicating a visit to a specialist (e.g., psychiatrist, neurologist, geriatrician) in the three years prior to the first AD diagnosis.

To facilitate detailed examination of linked free text information, a sample of 50 patients was randomly drawn (using a computer-generated randomization algorithm) from the cohort meeting the criteria in Step 1 for further analysis. In particular, using the SAS software (SAS Institute, Cary, NC), all patients were assigned a random number. Following this, the first 50 patients with the smallest values for the randomly assigned numbers were selected from the dataset. A random sampling approach was used to increase the likelihood that the sub-sample selected was representative of the overall cohort identified in Step 1.

Development of the code-based algorithm

Earliest indications of symptoms of cognitive decline (e.g., "memory loss symptom"), cognitive assessment (for either screening or diagnosis), and AD diagnosis were identified using two parallel approaches. In the first approach, the Read codes, ICD-10 codes, and prescription medications indicated to treat AD (i.e., cholinesterase inhibitors and memantine) found in the structured part of CPRD from up to 3 years prior to the AD diagnosis were reviewed and categorized into an algorithm to establish first observed dates of the three key time points in the pathway of progression from onset of symptoms to AD diagnosis.

In the second approach, in addition to the diagnosis codes, a targeted search of the pseudo-anonymised text data and additional correspondence provided by the GPs was conducted to identify key phrases suggestive of the earliest markers of symptoms related to cognitive impairment (e.g., "memory loss", "cognitive impairment", "confusion", etc., and their variants), cognitive assessments (e.g., "GPCOG", "MMSE", "MOCA", "mini-mental", etc., and their variants) and AD diagnosis (see Appendix Table 2 for a list of all phrases identified from this process). The targeted search was conducted by two independent reviewers to account for any subjective interpretation of the free-text.

Based on preliminary data inspection and the combined manual review of the text and structured data for 15 of the 50 patients, the definition of cognitive assessment using the structured data was refined to include an additional marker based on referrals. Specifically, given that clinical evaluation for dementia is usually undertaken by secondary care mental health specialists (e.g., geriatricians, old age psychiatrists, neurologists)[14] several weeks after the initial referral,[8] it was determined that a combination of codes indicating referral to a specialist and a letter from specialist within 3 months after the referral could be considered as indication of cognitive assessment. In addition, it was assumed that the earliest indication of cognitive assessment could not precede the earliest symptom of cognitive impairment.

Appendix Table 3 describes the final code-based algorithm used for quality evaluation.

Quality evaluation of the reliability of the code-based algorithm

The findings from the two approaches were compared to quantify the differences in dates for the first indicators of cognitive/functional symptoms, assessments, and AD diagnosis as identified by the code-based algorithm and manual review. Additionally, the percent of patients for whom the dates of each of the three measures (indicator for cognitive impairment symptoms, cognitive assessments, and AD diagnosis) identified by the code-based algorithm were after the dates suggested by the second approach (suggesting the code-based algorithm was less sensitive) were calculated. Similarly, the proportions of patients for whom the dates of the three measures as identified by the code-based algorithm were before the dates identified by the second approach (suggesting the code-based algorithm was more sensitive) were reported. While exact matches were preferred for all analyses, in order to account for delays between the receipt of a letter from the specialist assessing the patient and the corresponding coding of the information in CPRD, a similar metric allowing for a 30-day gap between the dates identified by the two

approaches was also measured. Note that for the purpose of the analysis, if an event was not observed for both approaches, it was considered an exact match. However, if a date was identified only in the manual review and not in the code-based algorithm, then the code-based algorithm was considered less sensitive. Similarly, if a date was identified in the code-based algorithm but not in manual review, the code-based algorithm was considered more sensitive.

Additionally, the days between the dates of first symptom of cognitive impairment and first cognitive assessment, and between the first cognitive assessment and the first AD diagnosis were compared for the two approaches. Congruence between the two data sources with regards to recording the type of and results from the specific type of the cognitive assessments performed prior to AD diagnosis was described.

The study approach is illustrated in Appendix Figure 1.

Results

Sample characteristics

Overall, 18,281 patients in the CPRD had an indication of AD (based on diagnosis codes or AD-related medications) in 2010-2013 (See Figure 1). Of these, 12,252 (67%) patients had their first indication of AD in 2010-2013; 11,151 had no indications of another type of dementia between or after AD diagnoses. Of these 11,151 patients, 4,515 (40%) patients had evidence of both text-field data and receipt of care in secondary settings in the 3 years prior to the first AD diagnosis. The final sample comprised 1,937 patients who met all the inclusion and exclusion criteria (mean age 82 years, 38% males). The random sample of 50 patients (selected from the 1,937 patients meeting all selection criteria) included in additional analyses had similar demographic characteristics as the 1,937 patients (mean age 82 years, 42% males). These 50

patients had a total of 2,051 records with valid pseudo-anonymized text field data and 44 correspondences from secondary care, provided by CPRD upon request.

Comparison of findings from the two approaches

Of the 50 patients included in the sample, the code-based algorithm identified 48 patients with evidence of cognitive impairment prior to AD diagnosis and 42 with evidence of cognitive assessment prior to AD diagnosis. An additional 2 and 4 patients respectively had evidence of cognitive impairment and cognitive assessment on the same date as the AD diagnosis. The remaining 4 patients had no record of cognitive assessment prior to or on the same date as the AD diagnosis (Appendix Figure 2). For the second, comprehensive approach which utilized information from all available data elements including text-based data, the number of patients with cognitive impairment and cognitive assessments prior to AD diagnosis were 49 and 43 respectively, and the numbers of patients with the same dates for these metrics as the AD diagnosis were 1 and 4 respectively. No record of cognitive assessment was observed prior to or on the same date as the AD diagnosis for 3 patients (Appendix Figure 2).

With regards to the timing of the three key events, relative to the second approach, the code-based algorithm was able to identify exact matches for the first date of symptoms associated with cognitive impairment in 33 (66%) of the 50 patients, first cognitive assessment in 29 (58%) patients, and first AD diagnosis in 43 (86%) patients (Table 1). Allowing for matches within 30 days, the algorithm's success rates increased to 74%, 70%, and 94%, respectively, for the dates of first cognitive impairment symptom, first cognitive assessment, and first AD diagnosis. For most of the remaining patients, the dates detected by the code-based algorithm were several days after the dates detected by the more comprehensive approach. There was only 1 patient (2% of the sample), for whom, the date of first symptoms of cognitive impairment

identified by the algorithm was earlier than the date identified by the second, comprehensive approach, suggesting the algorithm was more sensitive. The results were similar even after allowing for a 30-day gap in the dates identified by the two approaches. With respect to identifying the dates of first cognitive assessment the code-based algorithm was found to be more sensitive than the comprehensive approach in 8 patients (16%) based on exact matches and 4 patients (8%) allowing for matches within 30 days. The differences in the detection of the first date of AD diagnosis between the code-based algorithm and manual review based on either exact matches or matches within 30 days were very small.

Table 1: Differences in dates of earliest indications of cognitive impairment, cognitive assessment, and AD diagnosis as identified by coded-data vs. comprehensive data review (N=50)

	First symptom	First cognitive assessment	AD diagnosis
Date matches with manual review, n (%)			
Exact matches	33 (66.0%)	29 (58.0%)	43 (86.0%)
Matches ± 30 days	37 (74.0%)	35 (70.0%)	47 (94.0%)
Characteristics of mismatches, n (%)			
Code-based algorithm more sensitive than manual review	1 (2.0%)	8 (16.0%)	0 (0.0%)
Code-based algorithm more sensitive than manual review (< -30 days)	1 (2.0%)	4 (8.0%)	0 (0.0%)
Code-based algorithm less sensitive than manual review	16 (32.0%)	13 (26.0%)	7 (14.0%)
Code-based algorithm less sensitive than manual review (> + 30 days)	12 (24.0%)	11 (22.0%)	3 (6.0%)

Abbreviation: AD = Alzheimer's disease

Notes:
Manual review included the review of both structured data and text-based data; cases where dates were not observed by either approach (n=2 for cognitive assessment only) were considered exact matches; if the algorithm generated a date value that either preceded the equivalent date in the manual review or for which an equivalent date in the manual review as not observed, it was considered as being more sensitive than the manual review.

Additionally, the code-based algorithm and the comprehensive review of all data elements returned qualitatively similar gaps between the dates of first symptom of cognitive impairment and first cognitive assessment, and between the first cognitive assessment and the first AD diagnosis. For both approaches, the median time between first symptom and cognitive assessment was under 6 weeks (37 days for the manual review and 14 days for the algorithm) whereas the median time between the first cognitive assessment and the first AD diagnosis was between 6-7 months (214 days for the manual review and 181 days for the algorithm) (Figures 2 and 3).

In terms of the specific types of cognitive assessments performed prior to AD diagnosis, 34 (68%) patients had information available on the type of cognitive assessments conducted. Among these, very few patients received screening-type evaluations: 3 patients received the AMTS, 5 patients received the 6CIT, and 1 patient received GPCOG (Table 2). The more detailed evaluations captured in the data included the ACE-R (5/50 patients) and the MMSE (30/50 patients; a total of 53 assessments). A total of 9 patients received multiple tests prior to AD diagnosis, primarily in addition to ≥1 MMSE assessment (Table 2). For the most commonly administered cognitive assessment – the MMSE – the results were largely captured only in the supplemental (text-based) data. Specifically, 38 out of the 53 assessments had valid test scores available in the text-based data, only 6 of which were available and were consistent in both data sources. Additional 13 scores were observable only in the structured portion of the data, and neither data source reported scores for the remaining two assessments.

Table 2: Descriptive characteristics of cognitive assessments in the three years prior to AD diagnosis (N=50)

Cognitive testing characteristic	n (%)
Any cognitive test	34 (68.0%)

Type of cognitive test	
General Practitioner Assessment of Cognition (GPCOG)	1 (2.9%)
Abbreviated Mental Test Score (AMTS)	3 (8.8%)
Six-item cognitive impairment test (6CIT)	5 (14.7%)
Addenbrooke's Cognitive Examination - Revised (ACE-R)	5 (14.7%)
Mini-mental State Examination (MMSE)	30 (88.2%)
Multiple MMSE tests	14 (46.7%)
Multiple tests of different types	9 (26.5%)
MMSE + ACE-R	3 (33.3%)
MMSE + AMT	2 (22.2%)
MMSE + 6CIT	2 (22.2%)
6CIT + GPCOG	1 (11.1%)
MMSE + ACE-R + AMTS	1 (11.1%)

Abbreviation: AD = Alzheimer's disease

Discussion

The results of this pilot study suggest that the information captured within the supplemental text-based data fields provide increased accuracy over the structured portion of CPRD data regarding the dates of first symptom of cognitive impairment, first cognitive assessment, and first AD diagnosis, among patients diagnosed with AD. The comparison between the code-based algorithm developed in this study and a manual review of a patient's medical history (including structured data, free text, and correspondence from secondary care settings) suggests that the concordance between the two is highest for identifying the timing of the first recorded AD diagnosis, with diminishing effectiveness of the code-based algorithm in identifying the earliest records for symptoms of cognitive impairment and first cognitive assessment, respectively. Additionally, nearly two-thirds of the 50 patients included in the study had records indicative of specific types of cognitive assessments prior to or concomitantly with their AD diagnoses. For the cognitive assessment captured most commonly in the data, the

MMSE, the test results were available in the text-based data for 38 of the 53 assessments, whereas the results for 13 assessments were captured only in the coded data, and the scores for the remaining 2 assessments were not available in either data source. This suggests that although the text-based data elements are more likely to capture this information, neither the coded data, nor the additional information captured in physician notes and secondary care sources provide a comprehensive view of the detailed results of cognitive assessments. This may in part be due to the fact that much of the cognitive evaluation in England is done in specialty clinics such as memory clinics and the detailed data regarding the use of and findings from cognitive assessments may not be transferred back to the GPs. Even if the information is transferred back, it may not be entered into the system. However, given the recent initiatives to increase awareness about recognizing and recording symptoms of cognitive decline within the GP setting in England (especially in populations at increased risk for dementia),[8,11] and improve care-coordination as well as documentation across different provider settings,[15,16] the quality and completeness of data recording are likely to improve in the future, which could increase the reliability of the code-based algorithm. The improved quality of the recorded data would also facilitate identification of symptoms of cognitive impairment sooner, and facilitate real-world research into implications of earlier identification of cognitive impairment on subsequent outcomes in the UK.

Study strengths and limitations

The study used data from both the structured portion of CPRD and the text fields reflecting rich, additional information from notes captured by physicians/specialists during consultation. Using these enriched data elements, this study developed a code-based algorithm based on the findings from an intensive manual review process independently conducted by two reviewers. In doing so, we not only identified relevant medical codes and prescriptions to

identify timing of onset of cognitive symptoms, cognitive assessments, and AD diagnosis, but also captured an additional marker of cognitive assessment based on sequencing of clinical interactions. In addition, the study provides important insight into the availability of results from cognitive assessments, in particular MMSE, from both physician notes and coded data.

However, this study also has a number of limitations. First, the study relies on the Read codes (Primary Care) and ICD-10 codes (secondary care) used within the CPRD and HES administrative records datasets, respectively. These codes are retrieved from electronic health records and hospital admission records and do not contain information by which to confirm clinical diagnoses, severity of illness, or physician interpretation. Accordingly, it is possible that some patients identified as having been diagnosed with AD, with no recorded diagnosis of other type, have other dementia etiologies instead.[17] Relatedly, the earliest marker of onset of cognitive symptoms is based on the information captured in the data, and the precise timing of perceived onset of cognitive impairment is not known. In addition, for this study, though we reviewed the correspondence from some secondary care interactions, we did not have access to data from memory clinics, which is a key setting in which cognitive assessments are conducted in England. Future research should identify avenues to compare the reliability of the algorithm relative to data captured in these settings as well. This study is also limited in sample size, as the algorithm was only developed and assessed for 50 randomly selected patients who were diagnosed with AD. In addition, the algorithm may not capture all Read codes and ICD-10 codes indicative of symptoms of cognitive impairment, cognitive assessment, and AD diagnosis. As such, additional research using larger patient populations is necessary to further test the reliability and generalizability of the algorithm. Furthermore, the study was focused on patients with AD who had no evidence of other dementia etiologies, and further research is needed to assess the

reliability of the coded data for identifying the timing of cognitive impairment, cognitive assessment, and diagnosis among patients with other dementia etiologies. Finally, the study utilized data prior to 2014 and the study findings may not reflect the current practices in management of patients with dementia in England.

Conclusions

Given the limited expected future availability of free text data and secondary care correspondence in CPRD, the code-based algorithm developed using data for a small sample of AD patients shows promise as a reliable alternative for identifying the earliest indications of AD. However, the reliability of using coded data to identify earliest symptoms of cognitive impairment as well as indications of cognitive assessments prior to AD diagnosis is limited. The use of coded data, in its present form, is not recommended for identifying information regarding the specific types of cognitive assessments performed, the specialty of physicians performing the assessments or the results associated with those assessments (e.g., to assess disease severity levels).

Ethics approval and consent to participate

This study was approved by the Independent Scientific Advisory committee (ISAC): Protocol # 16_043R.

Availability of data and materials

This study used the Clinical Practice Research Datalink, provided by CPRD. Per the data use agreement, the datasets supporting the conclusions of this article cannot be made available to researchers outside of the study team. However, interested readers may request the data directly from CPRD – see https://www.cprd.com/researcher/ for more information.

Competing interests

GD, CCR, MB, and ALS are full-time employees of Eli Lilly and Company. NYK, UD, JW, and MKM are employees of Analysis Group, Inc., a company that received funding from Eli Lilly and Company for this research. CM and JR are consultants to Eli Lilly and Company.

Funding

This study was funded by Eli Lilly and Company.

Authors' contributions

NYK, UD, GD, CCR, and MB contributed to the conceptual design and reviewed and discussed the study results. JW and MKM contributed to the conceptual design and performed data analysis. ALS, JR, and CM contributed in the interpretation of study findings. All authors reviewed, edited, and approved the final manuscript.

References

1. Dementia UK: The full report. Alzheimer's Society. https://www.alzheimers.org.uk/site/scripts/download_info.php?fileID=2323 Accessed February 6, 2017.

2. Mortimer JA, Borenstein AR, Gosche KM, Snowdon DA. Very Early Detection of Alzheimer Neuropathology and the Role of Brain Reserve in Modifying Its Clinical Expression. J Geriatr Psychiatry Neurol 2005;18: 218-223.

3. Jack CR, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, Petersen RC, Trojanowski JQ. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 2010; 9:119.

4. Peterson RC. Early diagnosis of Alzheimer's disease: Is MCI too late? Curr Alzheimer Res. 2009;6(4):324-330.

5. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute of Aging and the Alzheimer's Association workgroup. Alzheimer's & Dementia 2011;1-7.

6. NHS Choices. Alzheimer's disease diagnosis. http://www.nhs.uk/Conditions/Alzheimers-disease/Pages/Diagnosis.aspx Accessed February 6, 2017.

7. National Institutes of Health. National Institutes on Aging. Alzheimer's disease diagnosis. https://www.nia.nih.gov/alzheimers/topics/diagnosis#how Accessed February 6, 2017.

8. Government of UK. Prime Minister's challenge on dementia 2020. 2015. https://www.gov.uk/government/publications/prime-ministers-challenge-on-dementia-2020/prime-ministers-challenge-on-dementia-2020 Accessed February 6, 2017.

9. Dubois B, Padovani A, Scheltens P, et al. Timely Diagnosis for Alzheimer's Disease: A Literature Review on Benefits and Challenges. J Alzheimers Dis 2015;49(3):617-631.

10. NHS Department of Health. Dementia revealed: what primary care needs to know. 2014. https://www.england.nhs.uk/wp-content/uploads/2014/09/dementia-revealed-toolkit.pdf Accessed February 6, 2017.

11. NHS England. Dementia diagnosis and management: a brief pragmatic resource for general practitioners. 2015. https://www.england.nhs.uk/wp-content/uploads/2015/01/dementia-diag-mng-ab-pt.pdf Accessed February 6, 2017.

12. Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015;44(3):827–836.

13. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Reserach Database: a systematic review. Br J Gen Pract. 2010 Mar;60 (572):e128-e136.

14. Alzheimer's Society UK. https://www.alzheimers.org.uk/info/20071/diagnosis/95/assessment_process_and_tests/2 Accessed February 6, 2017.

15. The King's Fund. Transforming our healthcare system: ten priorities for commissioners. 2015. https://www.kingsfund.org.uk/sites/files/kf/field/field_publication_file/10PrioritiesFinal2.pdf Accessed February 13, 2017.

16. NHS National Information Board. Personalized health and care 2020: using data and technology to transform outcomes for patients and citizens. 2014. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/384650/NIB_Report.pdf Accessed February 13, 2017.

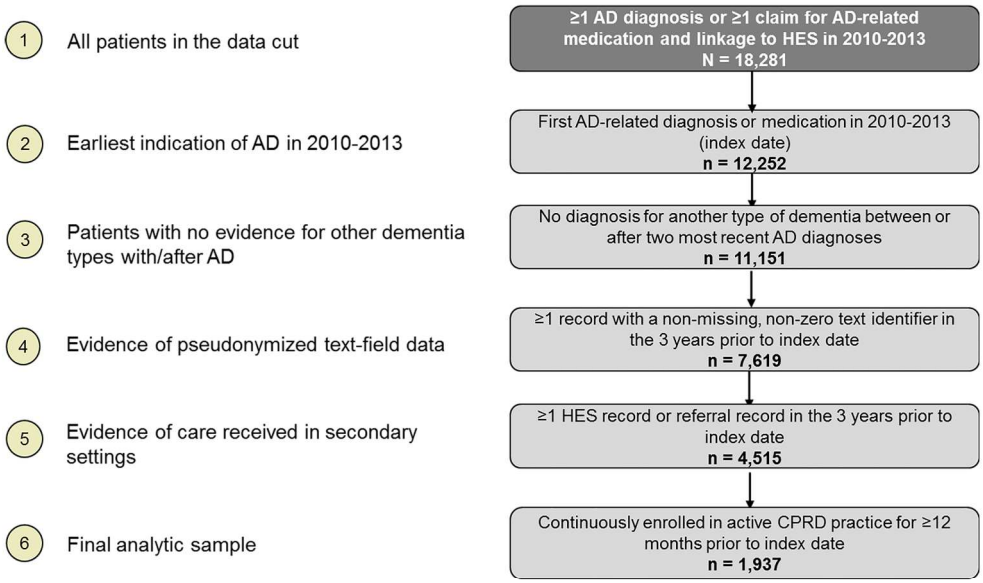
17. Happich M, Kirson NY, Desai U, et al. Excess costs associated with possible misdiagnosis of Alzheimer's disease among patients with vascular dementia in a UK CPRD population. J Alzheimers Dis 2016;53:171-183.

Figure Captions/Legends

Figure 1: Sample selection

Figure 2: Distribution of days between first cognitive symptom to cognitive assessment: code-based algorithm vs. comprehensive data review (N=50)

Figure 3: Distribution of days between first cognitive assessment to AD diagnosis: code-based algorithm vs. comprehensive data review (N=50)

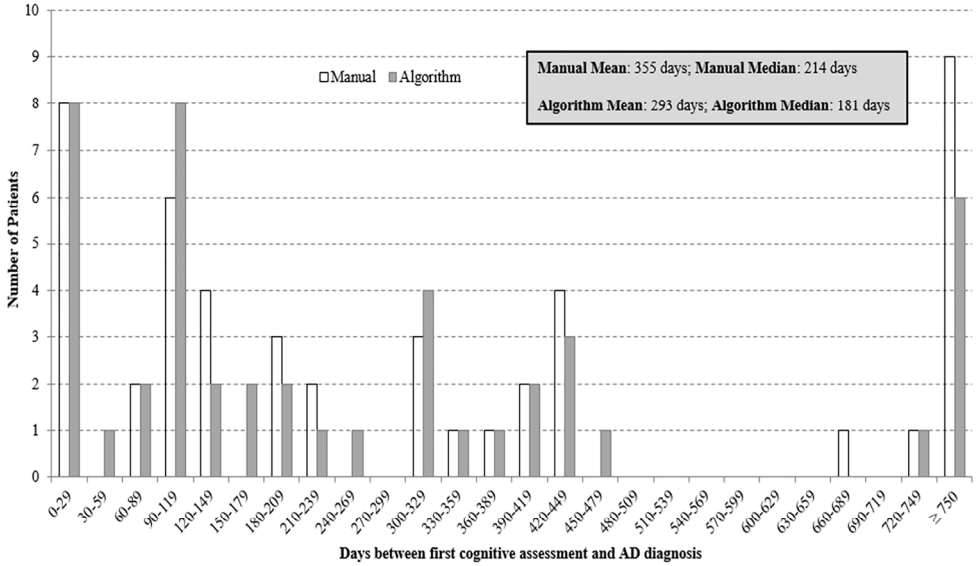


Abbreviations: AD = Alzheimer's disease, HES = Hospital Episode Statistics, CPRD = Clinical Practice Research Datalink

Note:
Please refer to Appendix Table 1 for the Read codes and ICD-10 codes used to identify AD and other dementia types.

Figure 1: Sample selection

282x217mm (300 x 300 DPI)

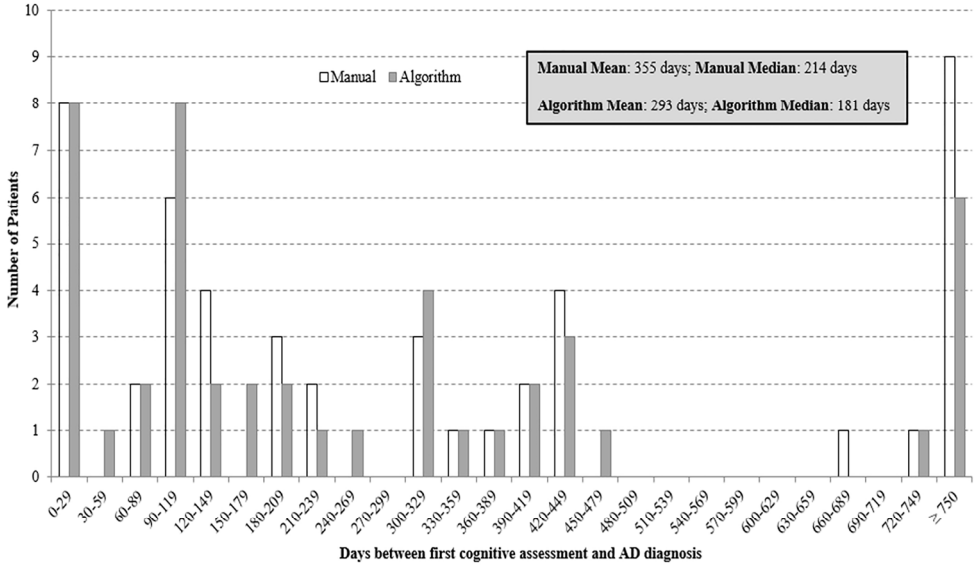


Abbreviation: AD = Alzheimer's disease

Note:
Manual review included the review of both structured data and text-based data

Figure 2: Distribution of days between first cognitive symptom to cognitive assessment: code-based algorithm vs. comprehensive data review (N=50)

235x171mm (300 x 300 DPI)



Abbreviation: AD = Alzheimer's disease

Note:
Manual review included the review of both structured data and text-based data

Figure 3: Distribution of days between first cognitive assessment to AD diagnosis: code-based algorithm vs. comprehensive data review (N=50)

235x171mm (300 x 300 DPI)

Appendix Table 1: Read codes and ICD-10 codes used to identify AD and other dementia types

Disease	Code	Description
Read codes		
Alzheimer's disease	Eu00.00	[X]Dementia in Alzheimer's
	Eu00000	Dementia in Alzheimer's disease with early onset
	Eu00011	[X]Presenile dement,Alzheimer
	Eu00012	Primary degen dementia, Alzheimer's type, presenile onset
	Eu00013	[X]Alzheimer's disease type 2
	Eu00100	[X]Late onset Alzheim dementia
	Eu00111	[X]Alzheimer's disease type 1
	Eu00112	[X]Senile dementia, Alzheimer
	Eu00113	Primary degen dementia of Alzheimer's type, senile onset
	Eu00200	[X]Atypical/mixed Alzheimer's
	Eu00z00	[X]Alzheimer's disease unspec
	Eu00z11	[X]Alzheimer's dementia unspec
	F110.00	Alzheimer's disease
	F110000	Alzheimer dis wth early onset
	F110100	Alzheimer's dis wth late onset
	Fyu3000	[X]Other Alzheimer's disease
Vascular dementia	E004.00	Arteriosclerotic dementia
	E004.11	Multi infarct dementia
	E004000	Arterioscl.dementia-uncomplic.
	E004100	Arterioscl.dementia+delirium
	E004200	Arterioscl.dementia+paranoia
	E004300	Arterioscl.dementia+depression
	E004z00	Arteriosclerotic dementia NOS
	Eu01.00	[X]Vascular dementia
	Eu01.11	[X]Arteriosclerotic dementia
	Eu01000	[X]Vascular dementia of acute onset
	Eu01100	[X]Multi-infarct dementia
	Eu01111	[X]Predom cortical dementia
	Eu01200	[X]Subcortical vascular dement
	Eu01300	[X]Mix cort/subcor vasc dement
	Eu01y00	[X]Other vascular dementia
	Eu01z00	[X]Vascular dementia unspecif
Dementia with Lewy bodies	Eu02500	[X]Lewy body dementia
	F116.00	Lewy body disease
Frontotemporal dementia	Eu02000	[X]Dementia in Pick's disease
	F111.00	Pick's disease
	F118.00	Frontotemporal degeneration

Disease	Code	Description
Normal-Pressure Hydrocephalus	F113000	Normal pressure hydrocephalus
Parkinson's dementia	Eu02300	[X]Dementia in Parkinson's
	F11x900	Cerebral degen Parkinson dis
ICD-10 codes		
Alzheimer's disease	G30.x	Alzheimer disease
	F00.x	Dementia in Alzheimer disease
Vascular dementia	F01.x	Vascular dementia
Dementia with Lewy bodies	G31.8	Other specified degenerative diseases of nervous system (Grey-matter degeneration, Lewy body disease, subacute necrotizing encephalopathy)
Frontotemporal dementia	G31.0	Circumscribed brain atrophy (frontotemporal dementia, Pick disease, progressive isolated aphasia)
	F02.0	Dementia in Pick disease
Normal-Pressure Hydrocephalus	G91.2	Normal pressure hydrocephalus
Parkinson's dementia	F02.3	Dementia in Parkinson disease

Appendix Table 2: Terms from text-data that are most frequently associated with the earliest dates of cognitive symptoms, cognitive assessment, and AD diagnosis

Category	Key Phrases		
Symptom	memory	disturbance	xalzheimers
	dementia	senile	decline
	mental	presenile	dconfusion
	alzheimers	dysfunction	impaired
	30	27	symptoms
	loss	neurology	29
	symptom	mmts	senility
	mmse	difficulties	24
	cognitive	deterioration	confusesd
	confused	22	forgetfulness
	poor	26	worsening
	problems	deteriorated	losing
	forgetful	difficulty	15
	impairment	disorder	20
	xdementia	deteriorate	23
	confusion	memoy	28
	cognition	problem	psychoses
Cognitive assessment	memory	mmts	review
	dementia	psychiatry	psychology
	mental	examination	psychogeriatrics
	alzheimers	team	psych
	30	referral	exam
	mmse	psychiatrist	phychological
	cognitive	psychogeriatrician	test
	xdementia	screening	screen
	cognition	assessment	psychological
	neurology		
Diagnosis	alzheimers	xalzheimers	

Appendix Table 3: Final code-based algorithm to identify early indications of cognitive symptoms, cognitive assessment, and AD diagnosis

Category	Diagnosis code	Description
Read codes		
Symptom	1B1A.12	memory loss symptom
	F110.00	alzheimer's disease
	Eu00.00	[x]dementia in alzheimer's disease
	Eu02z00	[x] unspecified dementia
	28G..00	forgetful
	Eu00100	[x]dementia in alzheimer's disease with late onset
	E2A1000	mild memory disturbance
	E00z.00	senile or presenile psychoses nos
	1B1A.13	memory disturbance
	Z7CF800	poor short-term memory
	Z7C1.00	impaired cognition
	R009.00	[d]confusion
	Eu00z11	[x]alzheimer's dementia unspec
	Eu05700	[X]Mild cognitive disorder
	2841.00	Confused
	2841.11	Confusion
	1461.00	H/O: dementia
	168..14	C/O 'Muzzy head'
	1JA2.00	Suspected dementia
	28E..00	Cognitive decline
	28H..00	Mentally vague
	E00..11	Senile dementia
	E00..12	Senile/presenile dementia
	Eu01y00	[X]Other vascular dementia
	Eu02500	[X]Lewy body dementia
	F116.00	Lewy body disease
	R00z011	[D]Memory deficit
	Z7CEH14	Memory problem
Cognitive assessment	9N1T.00	seen in psychiatry clinic
	388m.00	mini-mental state examination
	388V.00	mini mental state score
	6AB..00	dementia annual review
	9N1M.00	seen in psychology clinic
	ZL9D.00	seen by psychiatrist
	9Nk1.00	seen in memory clinic
	3AD3.00	six item cognitive impairment test

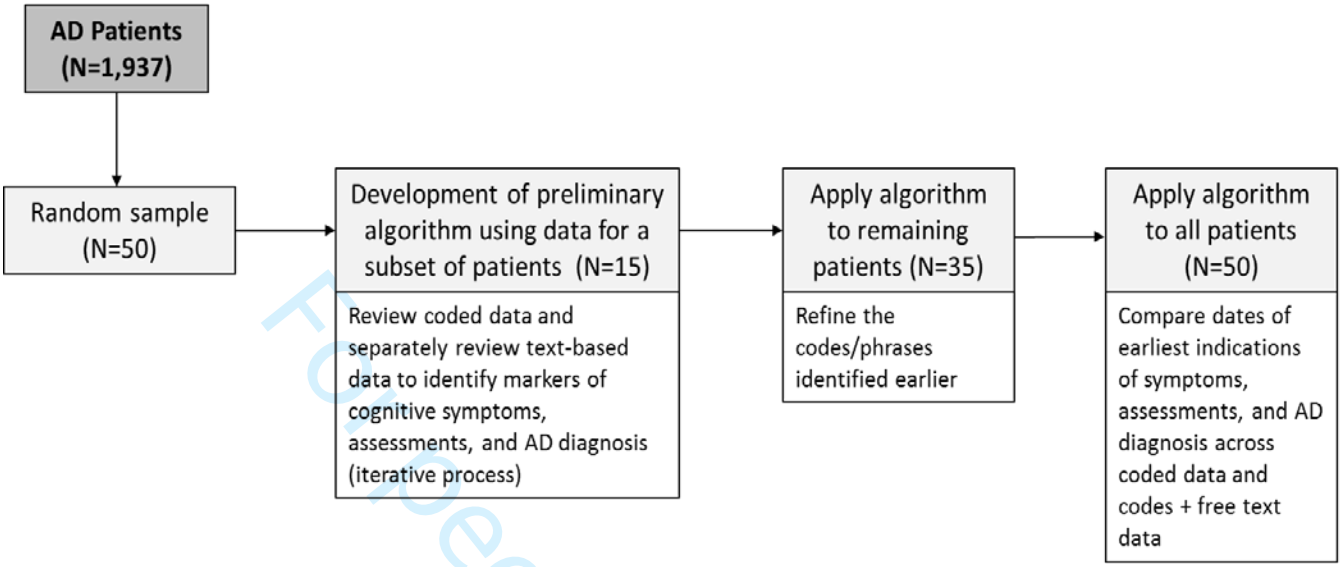
Category	Diagnosis code	Description
	9Nk6.00	seen in mental health clinic
	6A6..00	mental health review
	388m.11	mmse score
	9N1R.00	seen in neurology clinic
	ZRaA.00	mini-mental state examination
	9N2a.11	Seen by CPN
	ZL9D412	Seen by old age psychiatrist
	ZQ3E.00	Mental health assessment
	3A...11	Memory assessment
	8CM2.00	Psychiatry care plan
	ZL9D400	Seen by psychogeriatrician
	38C1000	Assessment for dementia
	38Dv.00	GPCOG - general practitioner assessment of cognition
	3A...12	Dementia assessment
	3AF..00	Addenbrooke's cognitive examination revised
	66h..00	Dementia monitoring
	8A2..00	Psychiatric monitoring
	8CMZ.00	Dementia care plan
	8HLC.00	Psychogeriatric D.V. done
	9N1yA00	Seen in psychogeriatric clinic
	9NN7.00	Under care of mental health team
	ZLA2E00	Seen by psychiatric nurse
	ZLA3111	Seen by CPN
	ZLB5.00	Seen by mental health counsellor
Relevant referral	8H4D.00	Referral to psychogeriatrician
	8H47.00	Geriatric referral
	8HKC.00	Psychogeriatrics D.V. requestd
	8HTY.00	Referral to memory clinic
	8Hc..00	Referral to mental health team
	8H49.00	Psychiatric referral
	8HHo.00	Referral to older age community mental health team
	ZL5B.00	Referral to psychiatrist
Encounter	9N1C.11	Home visit
	9N33.11	Letter encounter
	9N33.00	Letter encounter from patient
	9N35.00	Letter encounter to patient
	9N36.11	Letter from consultant
	9N36.00	Letter from specialist
	8H87.00	Follow-up 1 month
	9NV..00	Follow-up encounter

Category	Diagnosis code	Description
	9N32.00	Third party encounter
	6A...00	Patient reviewed
	9N3D.00	Letter received
	2....11	Examination of patient
	9H...00	Mental health administration
	ZL9AL00	Seen by care of the elderly physician
	68Q..00	Geriatric screening
	69D1.00	Geriatric health exam.
	9N1U.00	Seen in elderly assessment clinic
	9Nk5.00	Seen in elderly care clinic
	3876.00	Multidisciplinary assessment
	3891.00	Initial patient assessment
	3Z...00	Diagnostic procedure NOS
	67...11	Counselling
	671C.00	Discussed with doctor
	68P..00	Adult screening
	68Q3.00	Geriatric 75 year screen
	9N02.00	Seen in geriatric clinic
	9N0c.00	Seen in private clinic
	9N11.00	Seen in GP's surgery
	9N1C.00	Seen in own home
	9N22.00	Seen by practice nurse
	9N2G.00	Seen by consultant
	9N2N.00	Seen by Rota Doctor
	9N2R.00	Seen by co-operative doctor
	9N2o.00	Seen by health support worker
	9N7..11	Follow-up consultation
	9NFA.00	District nurse visit
	9NY..00	Appointment
	9Na..00	Consultation
	ZL23300	Under care of district nurse
	ZV67.00	[V]Follow-up examination
Other referral	8HR1.00	Refer for ECG recording
	8H7Y.00	Refer to acupuncture
	8H77.00	Refer to physiotherapist
	8H...00	Referral for further care
	8H68.00	Referral to haematologist
	8HTb.00	Referral to male urology clinic
	8H7..12	Referral to nurse
	8H4J.00	Referred to anaesthetist
	8H4K.00	Referred to endocrinologist

Category	Diagnosis code	Description
	8H52.00	Ophthalmological referral
	8H53.00	ENT referral
	8H54.00	Orthopaedic referral
	8H43.00	Dermatological referral
	8H7R.00	Refer to chiropodist
	8H48.00	Gastroenterological referral
	8H4L.00	Referred to nephrologist
	8H58.00	Gynaecological referral
	8H59.00	Referred to plastic surgeon
	8H5B.00	Referred to urologist
	8H5D.00	Referred to vascular surgeon
	8H5J.00	Referral to colorectal surgeon
	8H72.00	Refer to district nurse
	8H7G.00	Refer to speech therapist
	8H7Q.00	Refer to surgical fitter
	8H7V.00	Refer to audiologist
	8H7X.00	Refer to podiatry
	8HBJ.00	Stroke / transient ischaemic attack referral
	8HD..00	Refer to hospital OPD
	8HH5.00	Refer to domiciliary physiotherapy
	8HHk.00	Referral to hospital phlebotomist
	8HHl.00	Referral to practice phlebotomist
	8HQ..00	Refer for imaging
	8HQ2.00	Refer for ultrasound investign
	8HQ8.00	Referral for dual energy X-ray photon absorptiometry scan
	8HR8.00	Referral for 24 hour blood pressure recording
	8HTX.00	Referral to incontinence clinic
	8HVQ.00	Private referral to rheumatologist
	8He..00	Referral to intermediate care
	8He0.00	Referral to intermediate care - hospital at home
	8Hj0.00	Referral to diabetes structured education programme
	ZL85111	Referral to community physiotherapist
Diagnosis	F110.00	alzheimer's disease
	Eu00.00	[x]dementia in alzheimer's disease
	Eu00100	[x]dementia in alzheimer's disease with late onset
	Eu00z11	[x]alzheimer's dementia unspec
ICD-10 codes		
Symptom	F03	unspecified dementia
	R418	other and unspecified symptoms and signs involving cognitive functions and awareness
	R54	senility
	G309	Alzheimer's disease, unspecified

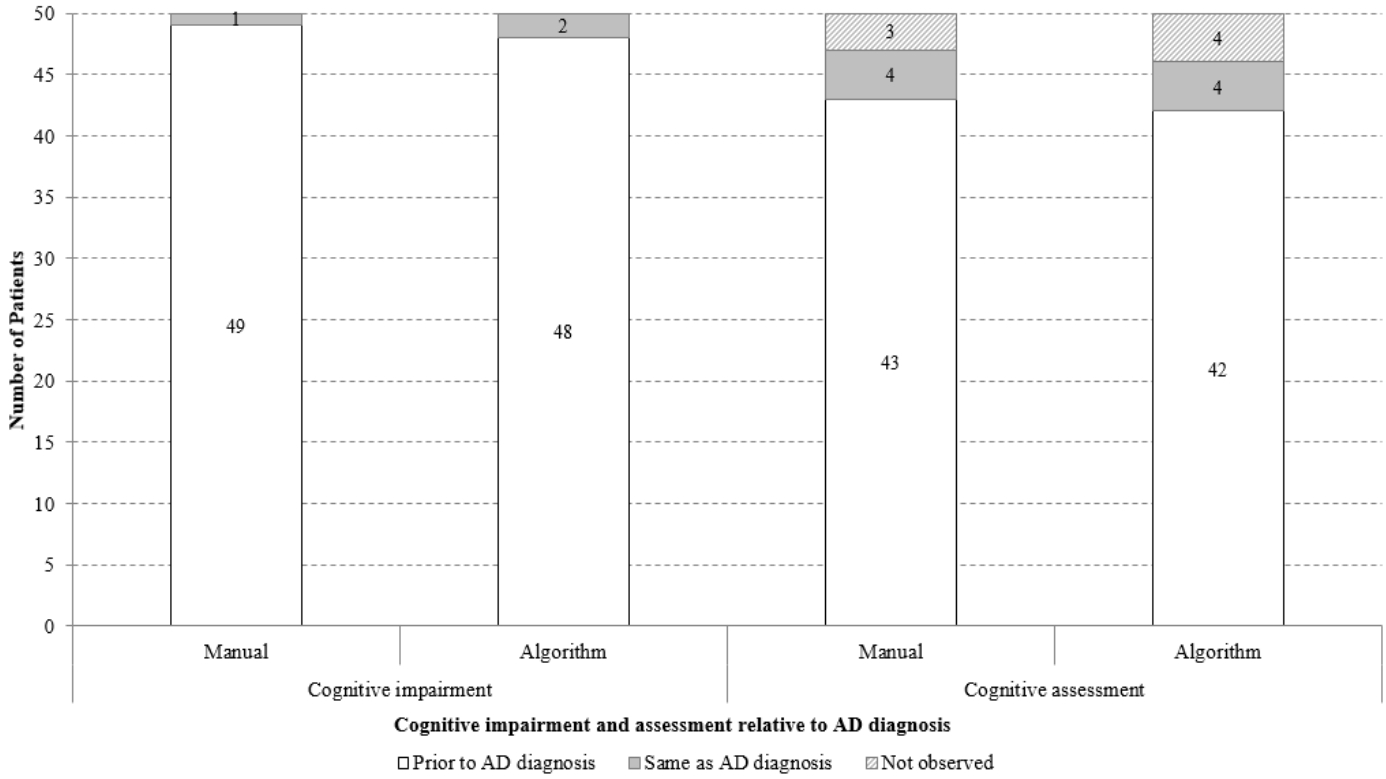
Category	Diagnosis code	Description
	G309D	Alzheimer's disease, unspecified
	R410	Disorientation, unspecified
	F051	Delirium superimposed on dementia
	F028	Dementia in other specified diseases classified elsewhere
	F067	Mild cognitive disorder
	F99	Mental disorder, not otherwise specified
Cognitive assessment - Encounter	Z139	Special screening examination, unspecified
Diagnosis	G309	Alzheimer's disease, unspecified
	G309D	Alzheimer's disease, unspecified

Appendix Figure 1: Study Schematic



Abbreviation: AD = Alzheimer's disease

Appendix Figure 2: Cognitive impairment and cognitive assessment relative to AD diagnosis



Abbreviation: AD = Alzheimer's disease

Note:
Manual review included the review of both structured data and text-based data

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	p.1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	pp.2-3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	pp. 5-6
Objectives	3	State specific objectives, including any prespecified hypotheses	p.6
Methods			
Study design	4	Present key elements of study design early in the paper	p. 1; pp.6-7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	pp.6-7
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	pp.7-8
		(*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	pp.8-10
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	pp. 6-10
Bias	9	Describe any efforts to address potential sources of bias	N/A
Study size	10	Explain how the study size was arrived at	pp.7-8
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	N/A
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	pp.8-10
		(*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	N/A
		(*e*) Describe any sensitivity analyses	N/A
Results			
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	pp. 10-11
		(b) Give reasons for non-participation at each stage	Figure 1
		(c) Consider use of a flow diagram	Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	pp. 10-11
		(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)	N/A

Outcome data	15*	Report numbers of outcome events or summary measures over time	N/A
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	pp.11-14
		(*b*) Report category boundaries when continuous variables were categorized	N/A
		(*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	N/A
Discussion			
Key results	18	Summarise key results with reference to study objectives	pp.14-15
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	pp.16-17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	p.17
Generalisability	21	Discuss the generalisability (external validity) of the study results	pp.16-17
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	p. 18

*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.