
Article:
Campbell, P., Rathod-Mistry, T., Marshall, M. et al. (16 more authors) (2021) Markers of dementia-related health in primary care electronic health records. Aging and Mental Health, 25 (8). pp. 1452-1462. ISSN 1360-7863

https://doi.org/10.1080/13607863.2020.1783511

Title: Markers of dementia-related health in primary care electronic health records

Authors

Paul Campbell[1,2], Trishna Rathod-Mistry[1], Michelle Marshall[1], James Bailey[1], Carolyn A Chew-Graham[1,2,3], Peter Croft[1], Martin Frisher[4], Richard Hayward[1], Rashi Negi[2], Swaran Singh[5], Shula Tantalo-Baker[1], Suhail Tarafdar[1], Opeyemi O Babatunde[1,9], Louise Robinson[6], Athula Sumathipala[1,2], Nwe Thein[2], Kate Walters[7], Scott Weich[8], Kelvin P Jordan[1,9*]

Affiliations

[1]School of Primary, Community and Social Care, Keele University, Keele, Staffordshire, ST5 5BG, UK

[2]Midlands Partnership NHS Foundation Trust, St. George's Hospital, Corporation Street, Stafford, ST16 3AG, UK

[3]West Midlands ARC, UK

[4]School of Pharmacy and Bioengineering, Keele University, Keele, Staffordshire, ST5 5BG, UK

[5]Division of Mental Health and Wellbeing, Warwick Medical School, University of Warwick, Coventry CV4 7AL, UK

[6]Institute of Health and Society and Newcastle University Institute for Ageing, Campus for Ageing and Vitality, Newcastle upon Tyne, NE4 5PL, UK

[7]Research Department of Primary Care & Population Health, University College London, Royal Free Campus, Rowland Hill St, London NW3 2PF, UK

[8]Mental Health Research Unit, School of Health and Related Research (ScHARR), University of Sheffield, Regent Court, 30 Regent Street, Sheffield S1 4DA, UK

[9]Centre for Prognosis Research, Keele University, Keele, Staffordshire, ST5 5BG, UK

***Corresponding author**: Kelvin Jordan, School of Primary, Community and Social Care, Keele University, Staffordshire ST5 5BG, UK. Tel. +44 1782 733924; Fax: +44 1782 734719. Email: k.p.jordan@keele.ac.uk

Total word count: 3872

Funding

This work was supported by The Dunhill Medical Trust [RPGF 1711/11]. The views and opinions expressed are those of the authors and not necessarily the views of The Dunhill Medical Trust. CiPCA funded by the Keele University School of Primary, Community and Social Care, and the Primary Care Research Consortium

Acknowledgement and Patient and Public Involvement

Informatics team at Keele University. The study team would also like to acknowledge the Patient and Public Involvement and Engagement Dementia Group within the School of Primary, Community and Social Care for their input into the development of the rapid literature review.

Abstract

Objectives

Identifying routinely recorded markers of poor health in patients with dementia may help treatment decisions and evaluation of earlier outcomes in research. Our objective was to determine whether a set of credible markers of dementia-related health could be identified from primary care electronic health records (EHR).

Methods

The study consisted of (i) rapid review of potential measures of dementia-related health used in EHR studies; (ii) consensus exercise to assess feasibility of identifying these markers in UK primary care EHR; (iii) development of UK EHR code lists for markers; (iv) analysis of a regional primary care EHR database to determine further potential markers; (v) consensus exercise to finalise markers and pool into higher domains; (vi) determination of 12-month prevalence of domains in EHR of 2328 patients with dementia compared to matched patients without dementia.

Results

Sixty-three markers were identified and mapped to 13 domains: Care; Home Pressures; Severe Neuropsychiatric; Neuropsychiatric; Cognitive Function; Daily Functioning; Safety; Comorbidity; Symptoms; Diet/Nutrition; Imaging; Increased Multimorbidity; Change in Dementia Drug. Comorbidity was the most prevalent recorded domain in dementia (69%). Home Pressures was the least prevalent domain (1%). Ten domains had a statistically significant higher prevalence in dementia patients, one (Comorbidity) was higher in non-dementia patients, and two (Home Pressures, Diet/Nutrition) showed no association with dementia.

Conclusions

EHR captures important markers of dementia-related health. Further research should assess if they indicate dementia progression. These markers could provide the basis for identifying individuals at risk of faster progression and outcome measures for use in research.

Keywords: Dementia, Prognosis, Electronic Health Records, Primary Care, Outcomes

Introduction

Dementia significantly impacts individuals, their families, and health and social care services (World Health Organisation, 2017; Cahill, 2019). Over 800,000 people live with dementia in the UK (Pink, O'Brien, Robinson & Longson, 2018), and its impact is likely to increase with a growing ageing population (Prince et al, 2013; Matthews et al, 2016). In response, the UK government initiated a National Dementia Strategy (Department of Health, 2009; Pickett et al, 2018), and Prime Minister's National Dementia Challenge (Department of Health, 2012) that included calls for delaying its consequences, such as nursing home admissions, hospital admissions, and early mortality (Dodd, Cheston, Ivanecka, 2015). It was recognised that primary care would have a central role in delivering this dementia strategy (Burns, 2012; Parmar et al, 2014; Greaves et al, 2015; Thyrian et al, 2016).

An important contribution to primary care management of a condition is understanding of its course and the factors that influence prognosis. Central to this is knowledge about markers relevant to primary care that indicate poorer health and progression. One potential longitudinal data resource for identifying common patterns of disease-related health is primary care Electronic Health Records (EHR). Primary care EHR contain information routinely recorded from patient contacts with primary care services including coded reason for consultation, prescriptions, referrals, investigations, and tests. The vast majority of the UK population are registered with a general practitioner (GP) providing a rich source of data on individuals over time. EHR databases have been used in dementia research previously, notably to ascertain factors associated with onset (Dunn, Mullee, Perry & Holmes, 2005; Rait et al, 2010; Cooper et al, 2015; Walters et al, 2016; Dell'Agnello et al, 2018). However, to date there has been little research using EHR to examine changes in health and prognosis in patients with dementia, despite evidence of patient variability over time in relation to long

term outcomes such as hospital admissions and mortality (Poblador-Plou et al, 2014; Alzheimer's Association, 2015).

In order to determine common patterns of progression in dementia and understand prognostic factors based on EHR, there is a need to first identify markers of health that are related to dementia that can be reliably detected using primary care EHR. This will not only help to identify those with poorer prognosis and help guide management of the disease but would also allow the use of routine data to evaluate interventions that aim to improve outcomes for those with dementia, significantly reducing time and cost of research studies in dementia. The aim of the current study was to investigate whether a set of credible markers of dementia-related health can be identified from routine EHR primary care data, and whether these individual markers can be grouped into larger domains. We also determined whether the recorded prevalence of the derived domains were higher in patients with dementia compared to those without dementia.

Methods

The study consisted of four stages (Figure 1).

Stage 1: Initial identification of pool of potential markers of dementia-related health

A rapid literature review was conducted including a systematic search to identify markers of dementia-related health (for example, dementia severity and outcomes) that had been used in previous EHR studies worldwide. The review was conducted in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), and registered on PROSPERO (CRD: 42016053455). Searches were undertaken from inception date to 2017 in AGELINE, AMED, CINAHL, EMBASE, Ethos, PsychINFO, MEDLINE and Web of Science, and the bibliographies of two UK-based primary care EHR databases (THIN [https://www.ucl.ac.uk/epidemiology-health-care/research/primary-care-and-population-

health/research/thin-database] and CPRD [https://www.cprd.com/Bibliography]). The search consisted of keywords and MeSH terms for dementia, and individual EHR databases (CPRD, GPRD, Kaiser Permanente, QResearch, ResearchOne, SAIL, THIN). Bibliographies of included research studies were hand searched. Studies had to follow participants post diagnosis of dementia, and use data from EHRs. Papers were excluded if they were: 1) not freely available in English language, 2) non-human studies, 3) editorials, guidelines, policies and/or non peer-reviewed, 4) not available in full-text after contacting the author. EHR data were defined as information extracted from routinely collected records of patient interactions with healthcare providers which were recorded in electronic format.

Retrieved citations were screened against these inclusion and exclusion criteria independently by two reviewers (STB, ST), with consensus reached where disagreements arose via discussion with a third reviewer (PCa). All extracted markers were systematically sorted (STB, ST, PCa) into domains specified by the International Consortium for Health Outcome Measurement (ICHOM) Standard Set for Dementia (ICHOM, 2016). These domains include clinical status, safety, sustainability, carer, symptoms, medication, quality of life, functioning. This allowed assessment of the coverage of markers compared to this comprehensive framework for dementia.

Stage 2: Consensus-based evaluation of potential marker identification in UK primary care EHR systems

An expert consensus meeting [*n*=7, consisting of general practitioners (GPs), experts in dementia, primary care EHR researchers] considered whether the markers identified from the review could feasibly be identified within UK primary care EHR, and whether there were other potential markers of dementia-related health that may be identifiable within the EHR.

At this stage, all markers were treated independently (i.e. markers of a similar or overlapping nature were not combined unless considered identical in nature).

In UK primary care, the Read code hierarchical system is currently used to record processes of care, symptoms, and diagnoses within EHR. The consensus group considered whether the markers identified from the review were likely to be coded and used within UK primary care EHR. Broad lists of Read codes were derived for those markers thought to be feasibly identifiable within primary care EHR based on previous UK-based EHR research studies (for example [Baker, Cook, Arrighi & Bullock, 2010; Davies, Kehoe, Ben-Shlomo & Martin, 2011; Grant, Drennan, Rait, Petersen & Illiffe, 2013; Imfeld , Bodmer , Schuerch, Jick & Meier, 2013; Cook et al, 2015]), existing databases of Read code lists,[Springate et al, 2014; Keele Medical Record Research website, https://www.keele.ac.uk/mrr], and additional searches of the Read code hierarchy. The consensus group felt that neuropsychiatric symptoms (for example, psychosis, depression) associated with dementia may be understood as part of dementia and therefore not coded separately. Therefore, we also included prescriptions of relevant medication to define these neuropsychiatric symptoms as in previous EHR studies (Dennis et al, 2017; Lewis, Werbeloff, Hayes, Howard & Osborn, 2018).

Stage 3: Refinement of markers, determination of further markers, and final allocation to domains

The application of the Read code lists developed in Stages 1 and 2 were tested on the records of patients with dementia included in a regional primary care EHR database (Consultations in Primary Care Archive (CiPCA)). CiPCA has ethical approval as a research database from North West Haydock Research Ethics Committee (ref 17/NW/0232) and contains the pseudonymised EHR of patients attending 9 GP practices in North Staffordshire (annual registered population approximately 90,000 patients). These practices have undergone regular

assessments on quality of their electronic morbidity recording since 1998 (Porcheret et al, 2004; Jordan et al, 2007; Jordan et al 2014).

We included the medical records of patients in CiPCA with a recorded diagnosis (new or ongoing) of any type of dementia between 2000 and 2015. Diagnosis of dementia was based on a Read code list developed previously through consensus of GP and EHR researchers (Burton, Campbell, Jordan, Strauss & Mallen, 2012), and code lists used in other studies (Brown et al, 2016; Browne et al, 2017; Pham et al, 2018). Coding of dementia in UK primary care EHR has been validated previously (Pujades-Rodriguez et al, 2018). Index date was defined as the date of earliest record of dementia within the period 2000-2015. Patients were censored at the end of 2015 or at the point their registration with the general practice ended if prior to the end of 2015.

We determined the number of patients with dementia who had at least one recorded code for each marker after their index date, and the most commonly recorded codes from the relevant code lists for each marker. Based on patterns of use of codes, markers were renamed, if necessary, and code lists refined and reduced from the initial broad lists. Markers deemed to be similar in terms of care, symptoms, or diagnosis were merged at this stage.

Hypothesis-free analysis was performed to identify other symptoms, morbidities, or processes of care recorded in the first 12 months after index date that were associated with dementia. Each patient with dementia was age, gender, and practice-matched to a randomly selected patient without a recorded diagnosis of dementia in the period 2000-2015. Patients without dementia were given the same index date as their matched patient's index date. All codes at the third level of the five level Read code hierarchy were examined (for example, code N05 "Osteoarthritis and allied disorders" is a third level code within Chapter N "Musculoskeletal and connective tissue diseases"). A code was taken forward if its prevalence was ≥1%

(indicating relevance at a population level), and its association with dementia was either statistically significant (unadjusted $p<0.05$) or the odds ratio was greater than 1.3 or less than 0.77 (suggesting a potential association with dementia with a small effect size, Olivier & Bell, 2013) derived from conditional logistic regression. Codes meeting the criteria were then combined with existing markers (from Stages 1 and 2) if they were judged as similar, or included as separate markers if deemed plausible.

Markers were then remapped to domains based on the ICHOM Standard Set for Dementia (ICHOM, 2016). A final expert consensus group (n=15) made final adjustments by combining or splitting markers, making definitions of markers and domains more precise, and allocating markers to domains as necessary. Read codes and medications lists are available from www.keele.ac.uk/mrr/morbiditydefinitions/.

Stage 4: Analysis of final list of domains

The final stage used the records of patients with dementia and matched non-dementia patients to determine the period prevalence of the final list of markers and domains in the 12 months after index date. In order to allow all patients the opportunity for a full 12 months follow-up prior to the end of the study period (end of 2015), we excluded patients with dementia (and their matched unexposed patients) with their first recorded diagnosis in 2015. All other patients were censored at the earliest of 12 months follow-up or the date their registration ended at the practice (therefore patients who left the practice or died before the end of the 12 months were retained to avoid healthy survivor bias). Prevalence of a marker was estimated as the number of patients with at least one coded record of the marker in the 12 months after index date, divided by the total number of patients with (or without) dementia. To determine prevalence by domain, the numerator was defined as the number of patients with at least one coded record of any marker from that domain in the 12 months after index date, and the

denominator was the total number of patients with (or without) dementia. Associations of domains with a dementia diagnosis were assessed using univariable binary logistic regression, with robust variance estimators to account for the matching of dementia and non-dementia patients.

Patient and Public Involvement

The Patient and Public Involvement and Engagement Dementia Group within the School of Primary, Community and Social Care at Keele University contributed to the development of the rapid literature review. The Dementia Group consisted of people with a diagnosis of dementia and caregivers of those with dementia (current and previously).

Results

Stage 1: Initial identification of a pool of potential markers of dementia-related health

3167 unique citations were retrieved as a result of the review. Screening removed 3133 papers (appendix 1), leaving 34 relevant papers (53% USA, 35% UK, 12% other). Data extraction identified 153 potential markers which were mapped to eight ICHOM domains (35% to symptoms domain, 22% sustainability and time to full time care, 17% medication, 15% safety, 3% clinical status, 3% functioning, 3% quality of life, 2% carer).

Stage 2: Consensus-based evaluation of potential marker identification in UK primary care EHR systems

The consensus exercise group agreed that 115 of the 153 markers were feasibly identifiable within UK primary care EHR either directly (for example, a fall or fracture Read code) or as a proxy (for example, a Read code indicating bedbound as an indication of daily functioning ability). Example items that were considered not feasible included changes in brain volume

using magnetic resonance imaging (MRI), quality of life measures, and details related to the caregiver. Initial Read code lists were then derived for the included markers.

Stage 3: Refinement of markers, determination of further markers, and final allocation to domains

There were 2714 patients with a recorded dementia diagnosis (new or ongoing) in the CIPCA database between 2000 and 2015. The lists of potential marker Read codes developed in Stage 2 were tested in the records of these patients. Further, 1622 codes at the third level of the Read code hierarchy were found to be recorded in the twelve-month period from index date and hence included in the hypothesis-free analyses. Of these 1622 Read codes, 282 had a prevalence of ≥1% and had an association (odds ratio>1.3 or <0.77, or p<0.05) with a dementia diagnosis. Of these 282 codes, 93 were already included in code lists for markers previously identified in Stages 1-2. A further 139 codes were excluded, mainly as they were codes for laboratory tests or general investigations being undertaken, routine monitoring, or administration and so unlikely to indicate aspects of dementia-related health. The remaining 50 codes included a range of comorbidities and symptoms and additional codes to those already identified around advanced directives and shared decision making. These were assessed alongside the 115 markers and their code lists derived in Stages 1-2, and the markers and domains further refined. This included merging markers of a similar or overlapping nature, updating code lists for markers, renaming of markers, and remapping markers to domains. The provisional markers and domains were then presented at the final consensus meeting. This led to a final set of 63 markers mapped to 13 domains (table 1).

Stage 4: Analysis of final list of domains

2328 patients had a recorded dementia diagnosis (new or ongoing) prior to 2015 and hence were included in the Stage 4 analysis. Mean age was 80.8 (SD 8.31) years and 65% were

female (table 2). The majority (98% of those with a type recorded) of dementia patients were diagnosed with Alzheimer's or vascular dementia. The proportion of patients that had a full 12 months of follow-up after index date was lower in dementia patients compared to the matched non-dementia patients (75% vs 85%).

The 12-month period prevalence for the final 13 domains is given in table 3. Having a record of one of the specified comorbidities (for example, cardiovascular, diabetes) was the most prevalent domain in both dementia (69%) and non-dementia patients (74%), with Home Pressures being the least prevalent domain (1% in both groups). Ten of the domains had a statistically significant higher prevalence in dementia patients compared with non-dementia patients. Aside from the Change in Dementia Drug domain, the Cognitive Function (odds ratio (OR) 9.25; 95% CI 7.47, 11.47), Severe Neuropsychiatric (7.14; 5.73, 8.89) and Care (4.87; 3.92, 6.05) domains had the strongest associations with dementia. The odds of having a recorded comorbidity were lower in dementia compared to non-dementia patients (OR 0.77, 95% CI 0.69, 0.87), and two domains (Home Pressures and Diet/Nutrition) showed no association with dementia.

The period prevalence of the individual markers is shown in table 4. The most prevalent markers in dementia patients were 'Depression, Anxiety, Stress' (39%), 'Musculoskeletal pain' (31%), 'Hypertension' (26%), and 'Severe Mental Illness' and 'Poor Diet' (both 24%). Markers that were more prevalent in dementia compared to non-dementia patients included 'Carer', and 'Advanced Directive' in the Care domain; 'Severe Mental Illness' in the Severe Neuropsychiatric domain; 'Depression, Anxiety, Stress' and 'Sleep Problems' in the Neuropsychiatric domain; 'Memory Loss' and 'Confusion' in the Cognitive Function domain; and 'Dietary supplement' in the Diet/Nutrition domain.

Discussion

This study shows that it is feasible to identify a number of potentially important markers of health in patients with diagnosed dementia using information routinely recorded in UK primary care. The markers and domains were generally found to be commonly recorded in the primary care records of patients with dementia (12-month period prevalence of twelve of the thirteen domains exceeded 10%) suggesting these markers and domains are relevant to this population.

This study combined a systematic search of EHR dementia-based research, mapping to a robust framework of outcomes related to dementia (ICHOM, 2016), expert consensus meeting to assess the relevance and applicability to primary care data in the UK, and testing within a localised primary care EHR database including exploratory (hypothesis-free) analysis to identify further markers. All codes and domains were further refined by a concluding expert consensus meeting. This led to a set of markers/domains identified within EHR that relate to the primary health care experience of those with dementia, and map to broader (non-primary care) health care outcomes. The reported 12-month period prevalence of the domains indicates these factors occurred at a frequency that make the potential to use them in future development of progression measures practical and plausible.

Comparison to other recent large EHR UK-based studies show comparable figures on age, gender, dementia type, prescriptions, and general comorbidity rates in both those with dementia and comparable aged cohorts without dementia (Guthrie, Clark & McCowan, 2010; Dennis et al, 2017; Lewis, Werbeloff, Hayes, Howard & Osborn, 2018). This suggests the findings of this current study are likely to be generalizable to the primary care consultation population within the UK, however further research should assess generalisability, particularly in countries where different primary care healthcare systems operate.

There are some limitations to the study. Whilst improvements have been made in the detection of dementia within primary care (Eichler et al, 2015), there may be individuals in our comparison group with undiagnosed or unrecorded dementia (Eichler et al, 2014; Lang et al, 2017). Any such misclassification might mean the strength of the associations of domains with dementia are underestimated. There may be differences in the prevalence of markers and domains by dementia sub-type, as has been shown with mortality rates (Garcia-Ptacek et al, 2014). Another limitation associated with the use of consultation-based EHR is that often only one condition is coded (usually the most prominent) in a consultation, even though older adults may consult for multiple reasons, therefore some information on multi-morbidity may be missing or have been added as "free text" which was not included in this current analysis.

Examination of the study findings in relation to the ICHOM Dementia Standard Set (ICHOM, 2016) shows successful identification of markers and domains indicative of aspects of clinical status, safety, functioning, care, symptoms, and medication. Results show associations with dementia were particularly strong for the domains relating to care, severe neuropsychiatric conditions, and cognitive function, all of which have been shown previously to relate to long term outcomes such as nursing home admission, hospitalisation, and mortality in non-EHR based studies (Connors et al, 2016; Lewis, Werbeloff, Hayes, Howard, & Osborn, 2018; Christensen & White, 2006). Further findings are also clinically informative, for example, common markers in the Daily Functioning domain are mobility and limitation problems, and in the Safety domain are falls and fractures. Similarly, our findings show a high prevalence of comorbidity and particularly symptoms in those with dementia indicating increased health burden, with previous EHR research showing an association of increasing number of chronic comorbidities with hospitalisation in patients with dementia (Browne, Edwards, Rhodes, Brimicombe, & Payne, 2017). These findings are all likely to reflect the increased vulnerability in this population (Kulmala, Nykänen, Mänty, &

Hartikainen, 2014), and whilst general population EHR-based measures exist to identify vulnerability (e.g. Electronic Frailty Index; Clegg et al, 2016), these have not been tested in the dementia population and the domains/markers within this study are specific to dementia.

While the domains mapped onto key dementia outcome criteria proposed by the ICHOM Dementia Standard Set (ICHOM, 2016), there are potential markers not routinely recorded in primary care EHR. For example, although the Cognitive Function domain contains markers of cognition and memory loss, it does not contain information regarding the actual level of cognitive ability or activities of daily living (ADL). The Care domain includes markers of shared decision making and additional care, but does not include information on social relationships, levels of formal and informal care, and caregiver burden. Improved collection of cognitive and ADL function, care provision and caregiver issues would help primary care further identify poorer dementia-related health. This may be aided in the future by technology which allows the incorporation of data from personal mobile devices that monitor health and wellbeing, and that allows a more widespread and smarter sharing of information between service providers (health, social care, third sector, government).

This is the first study to establish a feasible pool of markers of dementia-related health that are retrievable from primary care EHR, that have been grouped into wider domains to allow assessment across key areas of dementia-related health, and may allow assessment of dementia progression in individual patients. The next step in terms of research is to establish whether they are valid predictors of future adverse outcomes such as hospitalisation and death, work that we are undertaking currently in a UK-wide national primary care EHR dataset. If this validity is established, then the potential usefulness of these markers to clinical practice will lie in their capacity to provide evidence-based information to support shared decision-making by patients, their carers and clinicians, either by highlighting early in the course of the disease the presence of modifiable risk factors that can be targeted for

intervention or by highlighting patients at higher risk of poor long-term outcomes who would benefit from prioritisation of resources and care at an early stage. Recording of these markers and domains may assist in tailoring care to individuals' health needs based on areas of vulnerability, as well as highlighting patients early after diagnosis who appear to be accumulating markers which may indicate increased progression of dementia and may benefit from more targeted management including referral to other services. Such information could also be used at the population level for supporting planning and policy decisions on care provision. These domains and markers may also have potential value as outcomes in clinical trials and prognostic studies of the progression of dementia in the short and long-term, where studying long-term outcomes such as mortality and nursing home admission is unrealistic or impractical. The advantage of EHR data is that they are available in all primary care settings and are by definition updated as part of routine care. Our set of markers provide a clear starting-point for studies to evaluate the usefulness of prognostic data in practice and a framework for improving the range and quality of data within these domains.

This study has shown that EHR capture many domains and specific markers that are important indicators of health for persons with dementia. This research has the potential to provide clinically useful information to identify individuals with dementia at risk of more rapid progression, and a readily available method that may be useful as an outcome measure in future research (e.g. trials) or in 'natural experiments' that evaluate changes to practice in dementia care.

Ethical Approval

CiPCA Research database ethics approval, North West - Haydock REC ref: 17/NW/0232.

Contributions

Study was derived and planned by PCa, CCG, PCr, MF, SS, AS, KW, SW, and KPJ. Rapid literature review was led by STB, ST, OB, CCG, and PCa. TR-M and KPJ performed analysis of CiPCA. All authors contributed to development of markers and domains and the codelists. PCa, KPJ, MM, and TR-M drafted the paper and all authors commented on subsequent draft versions and approved final version.

Disclosures of interest

LR has a National Institute for Health Research Senior Investigator award; no other relationships or activities that could appear to have influenced the submitted work.

Data sharing

The Read codes for dementia and the markers are available at www.keele.ac.uk/mrr/morbiditydefinitions or by contacting the authors. CiPCA data cannot be shared due to the conditions of its ethics approval.

Figures

Figure 1 – Stages of the study

References

Alzheimer's, Association. (2015). 2015 Alzheimer's disease facts and figures. *Alzheimer's & Dementia: the Journal of the Alzheimer's Association*, *11*(3), 332.

Baker, N. L., Cook, M. N., Arrighi, H. M., & Bullock, R. (2010). Hip fracture risk and subsequent mortality among Alzheimer's disease patients in the United Kingdom, 1988–2007. *Age and Ageing*, *40*(1), 49-54.

Brown, A., Kirichek, O., Balkwill, A., Reeves, G., Beral, V., Sudlow, C., ... & Green, J. (2016). Comparison of dementia recorded in routinely collected hospital admission data in England with dementia recorded in primary care. *Emerging themes in Epidemiology*, *13*(1), 11.

Browne, J., Edwards, D. A., Rhodes, K. M., Brimicombe, D. J., & Payne, R. A. (2017). Association of comorbidity and health service usage among patients with dementia in the UK: a population-based study. *BMJ Open*, *7*(3), e012546.

Burns, A. (2012). The benefits of early diagnosis of dementia. *BMJ*, *344*, e3556.

Burton, C., Campbell, P., Jordan, K., Strauss, V., & Mallen, C. (2012). The association of anxiety and depression with future dementia diagnosis: a case-control study in primary care. *Family Practice*, *30*(1), 25-30.

Cahill, S. (2019). WHO's global action plan on the public health response to dementia: some challenges and opportunities. *Aging & Mental Health*, 1-3.

Christensen, M. D., & White, H. K. (2006). Dementia assessment and management. *Journal of the American Medical Directors Association*, *7*(2), 109-118.

Clegg, A., Bates, C., Young, J., Ryan, R., Nichols, L., Ann Teale, E., ... & Marshall, T. (2016). Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age and Ageing*, *45*(3), 353-360.

Connors, M. H., Ames, D., Boundy, K., Clarnette, R., Kurrle, S., Mander, A., ... & Brodaty, H. (2016). Predictors of mortality in dementia: the PRIME study. *Journal of Alzheimer's Disease*, *52*(3), 967-974.

Cook, M., Baker, N., Lanes, S., Bullock, R., Wentworth, C., & Arrighi, H. M. (2015). Incidence of stroke and seizure in Alzheimer's disease dementia. *Age and Ageing*, *44*(4), 695-699.

Cooper, C., Lodwick, R., Walters, K., Raine, R., Manthorpe, J., Iliffe, S., & Petersen, I. (2015). Observational cohort study: deprivation and access to anti-dementia drugs in the UK. *Age and Ageing*, *45*(1), 148-154.

CPRD database - https://www.cprd.com/Bibliography [Accessed September 2019].

Davies, N. M., Kehoe, P. G., Ben-Shlomo, Y., & Martin, R. M. (2011). Associations of anti-hypertensive treatments with Alzheimer's disease, vascular dementia, and other dementias. *Journal of Alzheimer's Disease*, *26*(4), 699-708.

Dell'Agnello, G., Desai, U., Kirson, N. Y., Wen, J., Meiselbach, M. K., Reed, C. C., ... & Rasmussen, J. (2018). Reliability of coded data to identify earliest indications of cognitive decline, cognitive evaluation and Alzheimer's disease diagnosis: a pilot study in England. *BMJ Open*, *8*(3), e019684.

Dennis, M., Shine, L., John, A., Marchant, A., McGregor, J., Lyons, R. A., & Brophy, S. (2017). Risk of adverse outcomes for older people with dementia prescribed antipsychotic medication: a population based e-cohort study. *Neurology and Therapy*, *6*(1), 57-77.

Department of Health. (2009). *Living well with dementia: A national dementia strategy*. Department of Health.

Department of Health. (2012). Prime Minister's Challenge on Dementia: Delivering Major Improvements in Dementia Care and Research by 2015.

Dodd, E., Cheston, R., & Ivanecka, A. (2015). The assessment of dementia in primary care. *Journal of Psychiatric and Mental Health Nursing*, *22*(9), 731-737.

Dunn, N., Mullee, M., Perry, V. H., & Holmes, C. (2005). Association between dementia and infectious disease: evidence from a case-control study. *Alzheimer Disease & Associated Disorders*, *19*(2), 91-94.

Eichler, T., Thyrian, J. R., Hertel, J., Köhler, L., Wucherer, D., Dreier, A., ... & Hoffmann, W. (2014). Rates of formal diagnosis in people screened positive for dementia in primary care: results of the DelpHi-Trial. *Journal of Alzheimer's Disease*, *42*(2), 451-458.

Eichler, T., Thyrian, J. R., Hertel, J., Michalowsky, B., Wucherer, D., Dreier, A., ... & Hoffmann, W. (2015). Rates of formal diagnosis of dementia in primary care: The effect of screening. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *1*(1), 87-93.

Garcia-Ptacek, S., Farahmand, B., Kåreholt, I., Religa, D., Cuadrado, M. L., & Eriksdotter, M. (2014). Mortality risk after dementia diagnosis by dementia type and underlying factors: a cohort of 15,209 patients based on the Swedish Dementia Registry. *Journal of Alzheimer's Disease*, *41*(2), 467-477.

Grant, R. L., Drennan, V. M., Rait, G., Petersen, I., & Iliffe, S. (2013). First diagnosis and management of incontinence in older people with and without dementia in primary care: a cohort study using The Health Improvement Network primary care database. *PLoS medicine*, *10*(8), e1001505.

Greaves, I., Greaves, N., Walker, E., Greening, L., Benbow, S. M., & Jolley, D. (2015). Gnosall primary care memory clinic: eldercare facilitator role description and development. *Dementia*, *14*(4), 389-408.

Guthrie, B., Clark, S. A., & McCowan, C. (2010). The burden of psychotropic drug prescribing in people with dementia: a population database study. *Age and Ageing*, *39*(5), 637-642.

Häyrinen, K., Saranto, K., & Nykänen, P. (2008). Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *International Journal of Medical Informatics*, *77*(5), 291-304.

ICHOM. The ICHOM Standard Set for Dementia version 1.0. (2016). *The International Consortium for Health Outcomes Measurement*. London, UK.

Imfeld, P., Bodmer, M., Schuerch, M., Jick, S. S., & Meier, C. R. (2013). Risk of incident stroke in patients with Alzheimer disease or vascular dementia. *Neurology*, *81*(10), 910-919.

Jordan, K. P., Jöud, A., Bergknut, C., Croft, P., Edwards, J. J., Peat, G., ... & Englund, M. (2014). International comparisons of the prevalence of health care for musculoskeletal disorders using population-based health care data from England and Sweden. *Ann Rheum Dis*, *73*(1), 212-8.

Jordan, K., Clarke, A. M., Symmons, D. P., Fleming, D., Porcheret, M., Kadam, U. T., & Croft, P. (2007). Measuring disease prevalence: a comparison of musculoskeletal disease using four general practice consultation databases. *Br J Gen Pract*, *57*(534), 7-14.

Keele Medical Record Website. https://www.keele.ac.uk/mrr [Accessed September 2019].

Kulmala, J., Nykänen, I., Mänty, M., & Hartikainen, S. (2014). Association between frailty and dementia: a population-based study. *Gerontology*, *60*(1), 16-21.

Lang, L., Clifford, A., Wei, L., Zhang, D., Leung, D., Augustine, G., ... & Chen, R. (2017). Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis. *BMJ Open*, *7*(2), e011146.

Lewis, G., Werbeloff, N., Hayes, J. F., Howard, R., & Osborn, D. P. (2018). Diagnosed depression and sociodemographic factors as predictors of mortality in patients with dementia. *The British Journal of Psychiatry*, *213*(2), 471-476.

Matthews, F. E., Stephan, B. C., Robinson, L., Jagger, C., Barnes, L. E., Arthur, A., ... & Dening, T. (2016). A two decade dementia incidence comparison from the Cognitive Function and Ageing Studies I and II. *Nature Communications*, *7*, 11398.

Olivier, J., & Bell, M. L. (2013). Effect sizes for 2× 2 contingency tables. *PLoS One*, *8*(3), e58777.

Parmar, J., Dobbs, B., McKay, R., Kirwan, C., Cooper, T., Marin, A., & Gupta, N. (2014). Diagnosis and management of dementia in primary care: exploratory study. *Canadian Family Physician*, *60*(5), 457-465.

Pham, T. M., Petersen, I., Walters, K., Raine, R., Manthorpe, J., Mukadam, N., & Cooper, C. (2018). Trends in dementia diagnosis rates in UK ethnic groups: analysis of UK primary care data. *Clinical Epidemiology*, *10*, 949.

Pickett, J., Bird, C., Ballard, C., Banerjee, S., Brayne, C., Cowan, K., ... & Knapp, M. (2018). A roadmap to advance dementia research in prevention, diagnosis, intervention, and care by 2025. *International Journal of Geriatric Psychiatry*, *33*(7), 900-906.

Pink, J., O'Brien, J., Robinson, L., & Longson, D. (2018). Dementia: assessment, management and support: summary of updated NICE guidance. *BMJ*, *361*, k2438.

Poblador-Plou, B., Calderón-Larrañaga, A., Marta-Moreno, J., Hancco-Saavedra, J., Sicras-Mainar, A., Soljak, M., & Prados-Torres, A. (2014). Comorbidity of dementia: a cross-sectional study of primary care older patients. *BMC Psychiatry*, *14*(1), 84.

Porcheret, M., Hughes, R., Evans, D., Jordan, K., Whitehurst, T., Ogden, H., & Croft, P. (2004). Data quality of general practice electronic health records: the impact of a program of assessments, feedback, and training. *Journal of the American Medical Informatics Association*, *11*(1), 78-86.

Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., & Ferri, C. P. (2013). The global prevalence of dementia: a systematic review and metaanalysis. *Alzheimer's & Dementia*, *9*(1), 63-75.

Pujades-Rodriguez, M., Assi, V., Gonzalez-Izquierdo, A., Wilkinson, T., Schnier, C., Sudlow, C., ... & Whiteley, W. N. (2018). The diagnosis, burden and prognosis of dementia: A record-linkage cohort study in England. *PloS one*, *13*(6), e0199026.

Rait, G., Walters, K., Bottomley, C., Petersen, I., Iliffe, S., & Nazareth, I. (2010). Survival of people with clinical diagnosis of dementia in primary care: cohort study. *BMJ*, *341*, c3584.

Springate, D. A., Kontopantelis, E., Ashcroft, D. M., Olier, I., Parisi, R., Chamapiwa, E., & Reeves, D. (2014). ClinicalCodes: an online clinical codes repository to improve the validity and reproducibility of research using electronic medical records. *PloS one*, *9*(6), e99825.

THIN database - https://www.ucl.ac.uk/epidemiology-health-care/research/primary-care-and-population-health/research/thin-database/publications [Accessed September 2019].

Thyrian, J. R., Eichler, T., Pooch, A., Albuerne, K., Dreier, A., Michalowsky, B., ... & Hoffmann, W. (2016). Systematic, early identification of dementia and dementia care management are highly appreciated by general physicians in primary care–results within a cluster-randomized-controlled trial (DelpHi). *Journal of Multidisciplinary Healthcare*, *9*, 183.

Walters, K., Hardoon, S., Petersen, I., Iliffe, S., Omar, R. Z., Nazareth, I., & Rait, G. (2016). Predicting dementia risk in primary care: development and validation of the Dementia Risk Score using routinely collected data. *BMC Medicine*, *14*(1), 6.

World Health Organization. (2017). Global action plan on the public health response to dementia 2017–2025.

Table 1. Final list of markers nested within domains with examples

Domain	Marker	Examples
Care	Additional Help	Home help, day care
Care	Carer	Evidence has a carer in records
Care	Shared Decision Making	Shared decision making
Care	Advanced Directive	Advanced care planning
Home Pressures	Home Pressures	Marital problems, family bereavement/row
Severe Neuropsychiatric	Severe Mental Illness	Psychosis, schizophrenia, anti-psychotic drug
Severe Neuropsychiatric	Sectioned	Sectioned Form completed/fee paid
Severe Neuropsychiatric	Crisis	Mental crisis plan, referral to crisis team
Severe Neuropsychiatric	Suicidal	Suicidal, high/medium suicide risk
Neuropsychiatric	Depression, Anxiety, Stress	Depression, anxiety, stress, anti-depressant drug
Neuropsychiatric	Aggressive Behaviour	Aggressive/abusive behaviour
Neuropsychiatric	Sleep Problems	Insomnia, nightmares, hypnotic/anxiolytic drug
Neuropsychiatric	Behavioural Issues	Behavioural problem, disinhibited behaviour
Neuropsychiatric	Low Mood	Low mood, tearful, worried, lack of concentration
Neuropsychiatric	Wandering	Wanders during day/night
Cognitive Function	Cognition	Cognitive decline, mentally vague
Cognitive Function	Memory Loss	Memory loss, amnesia, poor memory
Cognitive Function	Confusion	Confusion, delirium, disorientated
Cognitive Function	Aphasia	Aphasia, speech therapy/defect, stammer
Daily Functioning	Bedbound	Bedbound, bed-ridden
Daily Functioning	Wheelchair	Provision of/independent in wheelchair
Daily Functioning	Severe mobility limitation	Housebound, chairbound, zimmer frame
Daily Functioning	Mobility – Less Severe Limitation	Mobility poor, walking stick, gait abnormality
Daily Functioning	Pressure Sore	Pressure sore, decubitus ulcer
Daily Functioning	Driving	Unfit to drive, advised about driving
Daily Functioning	Difficulty in Eating	Eating problem, dependent for eating
Daily Functioning	Difficulty Handling Finance	Needs help handling financial affairs
Daily Functioning	Personal Care Limitation	Dependent for dressing/toilet/bathing
Daily Functioning	Stairs Limitation	Difficulty managing stairs, need help on stairs
Safety	Fall	Recorded fall
Safety	Fracture	Recorded fracture (excl. skull)
Safety	Intracranial Injury	Skull fracture, concussion
Safety	Safety Assessment	Falls risk assessment, home safety advice
Comorbidity	Cardiovascular	Myocardial infarction, ischaemic heart disease
Comorbidity	Stroke	Stroke, cerebral infarction
Comorbidity	Parkinson's Disease	Parkinson's disease
Comorbidity	Motor Neurone Disease	Motor Neurone disease
Comorbidity	Diabetes	Diabetes mellitus (type I or II)
Comorbidity	Epilepsy	Epilepsy, grand mal/petiti mal, fit frequency
Comorbidity	Asthma / COPD	Asthma, COPD, chronic bronchitis

Comorbidity	Musculoskeletal Pain	Osteoarthritis, regional pain, rheumatoid arthritis
Comorbidity	Anaemia	Iron deficiency anaemia, Vitamin B12 deficiency
Comorbidity	Ocular	Cataract, retinopathy, glaucoma, blindness
Comorbidity	Hypertension	Essential hypertension, hypertensive disease
Comorbidity	Candidiasis	Candidiasis, thrush
Symptoms	Dizziness	Dizziness, vertigo, hypotension, giddiness
Symptoms	Incontinence	Incontinent of urine/faeces, urgency micturition
Symptoms	Constipation / IBS	Constipation, irritable bowel syndrome (IBS)
Symptoms	Diarrhoea	Diarrhoea, loose stools
Symptoms	Urinary	Retention of urine, haematuria, dysuria
Symptoms	Neurological	Fit (no epilepsy record), blackout
Symptoms	Chest pain (non-cardiovascular)	Costochondritis, musculoskeletal/unspecified chest pain
Symptoms	Oral Health	Stomatitis, poor oral hygiene, sore mouth
Symptoms	Swallowing	Difficulty swallowing liquids/solids, dysphagia
Symptoms	Hearing Loss	Deafness, hearing loss/impairment
Symptoms	"Feels Unwell"	Recorded 'Feels unwell'
Diet/Nutrition	Poor Diet	Advice re diet, high fat diet, dietician referral
Diet/Nutrition	Nutrition	Vitamin/iron deficiency, osteomalacia
Diet/Nutrition	Weight Loss	Weight decreasing/loss, underweight
Diet/Nutrition	Dietary Supplement	Dietary supplement
Imaging	Imaging	X-ray, MRI, ECG, DXA, angiogram, CAT scan
Increased Multimorbidity	Increase in Polypharmacy	Increase in count of different drugs prescribed
Change in Dementia Drug	Change in Dementia-related Drug	Change in dementia-related drug prescribed

Table 2 – Demographic characteristics for patients included in Stage 4 analysis, *n* (%) unless stated

		Dementia	Non-Dementia
n		2328	2328
Dementia type	Alzheimer's	910 (39) (60[a])	N/A
	Vascular	574 (25) (38[a])	
	Other[b]	29 (1) (2[a])	
	Unknown[c]	815 (35)	
Gender	Male	805 (35)	805 (35)
	Female	1523 (65)	1523 (65)
Age: Mean (SD)		80.8 (8.31)	80.8 (8.31)
Full 12 month follow-up		1752 (75)	1969 (85)
Length of follow-up: Mean (SD) days		319 (95)	334 (85)

[a] Excluding Unknown from denominator; [b] Parkinson's, Lewy Body, Frontotemporal, Huntington's; [c] Includes where type of dementia is unspecified or recorded as "Senile Dementia" or "Presenile Dementia"; SD: standard deviation; N/A: not applicable

Table 3 – 12-month period prevalence of final derived marker domains, *n* (%)

	Dementia	Non-Dementia	OR[a] (95% CI)
n	2328	2328	
Care	414 (18)	99 (4)	4.87 (3.92, 6.05)
Home Pressures	28 (1)	29 (1)	0.97 (0.57, 1.63)
Severe Neuropsychiatric	574 (25)	102 (4)	7.14 (5.73, 8.89)
Neuropsychiatric	1170 (50)	574 (25)	3.09 (2.73, 3.50)
Cognitive Function	713 (31)	106 (5)	9.25 (7.47, 11.47)
Daily Functioning	267 (11)	148 (6)	1.91 (1.55, 2.35)
Safety	645 (28)	483 (21)	1.46 (1.30, 1.65)
Comorbidity	1600 (69)	1723 (74)	0.77 (0.69, 0.87)
Symptoms	803 (34)	634 (27)	1.41 (1.24, 1.60)
Diet/Nutrition	913 (39)	894 (38)	1.04 (0.93, 1.16)
Imaging	604 (26)	476 (20)	1.36 (1.19, 1.56)
Increased Multimorbidity[b]	1223 (53)	1028 (44)	1.40 (1.25, 1.57)
Change in Dementia Drug[b]	653 (28)	19 (<1)	Not calculated

[a] Unadjusted; [b] Compared to previous 12 months

Table 4 – 12-month period prevalence of final list of markers, *n* (%)

Domain	Marker	Dementia	Non-Dementia
	n	2328	2328
Care	Additional Help	24 (1)	20 (<1)
Care	Carer	236 (10)	37 (2)
Care	Shared Decision Making	60 (3)	11 (<1)
Care	Advanced Directive	232 (10)	62 (3)
Home Pressures	Home Pressures	28 (1)	29 (1)
Severe Neuropsychiatric	Severe Mental Illness	562 (24)	101 (4)
Severe Neuropsychiatric	Sectioned / Crisis / Suicidal[b]	19 (<1)	[a]
Neuropsychiatric	Depression, Anxiety, Stress	919 (39)	448 (19)
Neuropsychiatric	Aggressive Behaviour	21 (<1)	[a]
Neuropsychiatric	Sleep Problems	497 (21)	229 (10)
Neuropsychiatric	Behavioural Issues	22 (<1)	[a]
Neuropsychiatric	Low Mood	31 (1)	34 (1)
Neuropsychiatric	Wandering	10 (<1)	[a]
Cognitive Function	Cognition	229 (10)	[a]
Cognitive Function	Memory Loss	315 (14)	41 (2)
Cognitive Function	Confusion	199 (9)	55 (2)
Cognitive Function	Aphasia	57 (2)	13 (<1)
Daily Functioning	Mobility – Severe Limitation[c]	98 (4)	44 (2)
Daily Functioning	Mobility – Less Severe Limitation	108 (5)	70 (3)
Daily Functioning	Pressure Sore	53 (2)	15 (<1)
Daily Functioning	Driving	14 (<1)	16 (<1)
Daily Functioning	Personal Care Limitation[d]	24 (1)	5 (<1)
Daily Functioning	Stairs Limitation	24 (1)	8 (<1)
Safety	Fall	362 (16)	211 (9)
Safety	Fracture	154 (7)	98 (4)
Safety	Intracranial Injury	63 (3)	22 (<1)
Safety	Safety Assessment	275 (12)	269 (12)
Comorbidity	Cardiovascular	395 (17)	483 (21)
Comorbidity	Stroke	125 (5)	61 (3)
Comorbidity	Parkinson's Disease	66 (3)	27 (1)
Comorbidity	Motor Neurone Disease	[a]	[a]
Comorbidity	Diabetes	343 (15)	329 (14)
Comorbidity	Epilepsy	48 (2)	21 (<1)
Comorbidity	Asthma / COPD	217 (9)	293 (13)
Comorbidity	Musculoskeletal Pain	720 (31)	813 (35)
Comorbidity	Anaemia	158 (7)	134 (6)
Comorbidity	Ocular	210 (9)	228 (10)
Comorbidity	Hypertension	604 (26)	852 (37)
Comorbidity	Candidiasis	59 (3)	37 (2)
Symptoms	Dizziness	138 (6)	125 (5)
Symptoms	Incontinence	172 (7)	73 (3)

Symptoms	Constipation / IBS	195 (8)	125 (5)
Symptoms	Diarrhoea	129 (6)	86 (4)
Symptoms	Urinary	144 (6)	129 (6)
Symptoms	Neurological	19 (<1)	[a]
Symptoms	Chest pain (Non-cardiovascular)	96 (4)	102 (4)
Symptoms	Oral Health	17 (<1)	5 (<1)
Symptoms	Swallowing	38 (2)	30 (1)
Symptoms	Hearing Loss	101 (4)	126 (5)
Symptoms	"Feels Unwell"	47 (2)	32 (1)
Diet/Nutrition	Poor Diet	554 (24)	721 (31)
Diet/Nutrition	Nutrition	138 (6)	78 (3)
Diet/Nutrition	Weight Loss	89 (4)	36 (2)
Diet/Nutrition	Dietary Supplement	369 (16)	174 (7)
Imaging	Imaging	604 (26)	476 (20)
Increased Multimorbidity	Increase in Polypharmacy[e]	1223 (53)	1028 (44)
Change in Dementia Drug	Change in Dementia-related Drug[e]	653 (28)	19 (<1)

[a] Less than 5 cases; [b] Markers combined due to low frequency; [c] Includes Wheelchair, Bedbound, and Severe mobility limitation due to low frequency; [d] Personal care limitation includes Difficulties in eating and handling finance due to low frequency; [e] Compared to previous 12 months

COPD: Chronic obstructive pulmonary disease; IBS: Irritable Bowel Syndrome

Figure 1 – Stages of the study