Multi-time-point data preparation robustly reveals MCI and dementia risk factors

Multi-time-point data preparation robustly reveals MCI and dementia risk factors

Daman Kaur[a*], Magda Bucholc[b], David P. Finn[c], Stephen Todd[d], KongFatt Wong-Lin[b], Paula L. McClean[a]

a Northern Ireland Centre for Stratified Medicine, Biomedical Sciences Research Institute, Ulster University, Clinical Translational Research and Innovation Centre (C-TRIC), Altnagelvin Hospital Site, Derry~Londonderry BT47 6SB, Northern Ireland, UK. Ph: +44 (0) 2871675675

b Intelligent Systems Research Centre, School of Computing, Engineering and Intelligent Systems, Ulster University, Northland Rd, Derry~Londonderry BT48 7JL, Northern Ireland, UK. Ph: +44 (0) 2871675320

c Pharmacology and Therapeutics, School of Medicine, Galway Neuroscience Centre, National University of Ireland Galway, University Road, Galway H91 W5P7, Republic of Ireland. Ph: +353-91-495586

d Altnagelvin Area Hospital, Western Health and Social Care Trust, Derry~Londonderry BT47 6SB, Northern Ireland, UK. Ph: +44 (0) 2871345171

*Corresponding author: Daman Kaur, kaur-d1@ulster.ac.uk, Northern Ireland Centre for Stratified Medicine, Biomedical Sciences Research Institute, Ulster University, C-TRIC, Altnagelvin Hospital Site, Derry~Londonderry BT47 6SB, Northern Ireland, UK.

Declarations of interest: none

Abstract

INTRODUCTION: Conflicting results on dementia risk factors have been reported across studies. We hypothesize that variation in data preparation methods may partially contribute to this issue.

METHODS: We propose a comprehensive data preparation approach comparing individuals with stable diagnosis over time, to those who progress to mild cognitive impairment (MCI)/dementia. This was compared to the often-used 'baseline' analysis. Multivariate logistic regression was employed to evaluate both methods.

RESULTS: The results obtained from sensitivity analyses were consistent with those from our multi-time-point data preparation approach, exhibiting its robustness. Compared to analysis using only baseline data, the number of significant risk factors identified in progression analyses was substantially lower. Additionally, we found that moderate depression increased Healthy-to-MCI/Dementia risk, while hypertension reduced MCI-to-Dementia risk.

DISCUSSION: Overall, multi-time-point based data preparation approaches may pave the way for a better understanding of dementia risk factors, and address some of the reproducibility issues in the field.

Keywords: Dementia progression, mild cognitive impairment (MCI), cardiometabolic risk factors, multi-time-point data preparation, multivariate logistic regression, NACC data, longitudinal data, baseline.

1 Background

Identifying risk factors for dementia is important not only for understanding its underlying pathologies, but also for suggesting potential interventions [1]. In particular, cardiometabolic risk factors have been suggested to play a significant role in dementia [1-3]. However, there remain considerable gaps in knowledge, given that several studies have reported contradictory results regarding the impact of such risk factors on cognitive decline and dementia [2]. For instance, a study by Solomon and colleagues analysing midlife cholesterol levels in an American cohort (n=9,844), at baseline (i.e. single-time-point), found an increased risk of dementia associated with elevated cholesterol [4]. However, another study analysing a Swedish cohort (n=1,462) found no association between midlife cholesterol and dementia risk [5]. Similarly, another study on data from the University of Alabama at Birmingham (UAB) Study of Aging [6], analysing baseline data on 624 individuals showed significant association between diabetes with cognitive decline. Whereas a longitudinal study using the National Alzheimer's Coordinating Center (NACC) dataset (n=11,777) observed no association between diabetes and cognitive decline [7]. In the case of hypertension, a study on the Neurological Disorders of Central Spain (NEDICES) cohort (n=3,824) showed increased risk of dementia with untreated baseline hypertension [8]. In contrast, the *90+ Study* (n=559) found lower dementia risk associated with baseline hypertension [9]. Both studies applied the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV) criteria to determine dementia diagnosis.

Closely related to cardiometabolic risk factors is midlife obesity, which has been reported to increase dementia risk [1-3]. However, this has been challenged by the UK Clinical Practice Research Datalink (CPRD) cohort (n=1,958,191) findings, that revealed reduced dementia risk associated with midlife obesity [10]. Furthermore, considerable

differences have been observed across studies on lifestyle-related risk factors, such as alcohol consumption and cigarette smoking, and their relationship with dementia risk [11-14]. Drinking and smoking patterns are differently recorded, for instance smoking in some studies is categorized as former, current or never, whereas other studies measure cigarette pack-years [13,14]. With regards to the role of depression as a risk factor or a prodrome for cognitive impairment, it is still disputable [15-17]. The fact that the pathophysiological processes that lead to dementia occur decades before an official diagnosis is made, further complicates our understanding of the dementia-depression relationship. Similarly, the association between depression and MCI, and the accompanying acceleration in progression to dementia is evident in research, however, the cause-and-effect aspect remains debatable [18,19].

Based on the plethora of conflicting findings relating to risk reported in the dementia field, it is clear we have a significant reproducibility crisis, and ambiguity regarding the nature of association between various risk factors and outcome needs to be addressed. There can be several potential explanations, however a key issue is that methodologies used to calculate risk are not consistently applied across studies. For example, there are differences in sample sizes [4-9], inconsistent use of covariate/outcome definitions [11-14], and differences in consideration of treated/untreated groups [8,9], and diagnostic criteria used [4-7].

More importantly, for many risk factors, underlying pathologies and disease status vary over time, hence baseline values are not necessarily reflective of measurements at follow-ups [20]. Indeed, many individuals who are disease-free at baseline, subsequently acquire various medical conditions, including mild cognitive impairment (MCI) and dementia as well as cardiovascular disease and stroke. Analysis on baseline values alone may

therefore lead to misleading results [20]. Despite this, most studies do not account for temporal changes in risk factors [4,6-9,20]. Specifically, they underestimate the real strength of associations between risk factors and disease progression by relating the baseline value of a risk factor to outcome, even though it may substantially differ from the follow-up values (e.g. changing body mass index (BMI)).

Several studies have previously adapted a multi-time-point analysis approach and developed predictive models for progression to dementia [21-23]. However, these studies analysed risk factors that are numerical in nature such as cognitive test scores, hippocampal volume, total active voxels etc [21-23]. Categorical risk factors, such as presence/absence of comorbidities are not generally analysed at multiple time-points.

In the present study, we analyse several cardiometabolic comorbidities of dementia and other related risk factors using a comprehensive, multi-time-point data preparation approach. In this approach, data collated from several visits per individual is used to determine risk factors for Healthy-to-MCI, Healthy-to-Dementia and MCI-to-Dementia conversion. Thereafter, given the measurements of risk factors from successive patient visits, we create a consistent set of rules for defining longitudinal changes, and accordingly estimate the effect of a set of potential risk factors on progression of disease severity. Subsequently, this was compared to the single-time-point analysis method. We believe that our multi-time-point approach better represents risk factor changes over time and helps minimize bias introduced by varying data preparation methods.

2 Methods

2.1 Data source

The National Alzheimer's Coordinating Center (NACC) dataset, one of the largest and most comprehensive longitudinal databases for dementia research, collated across the United States of America was used in this study. It consists of over 500 variables on lifestyle, genetic and clinical data from over 34,000 individuals.

Details about the NACC consortium and design and implementation of the NACC database have been described previously [24].

The dataset used in our longitudinal investigation was the NACC Uniform Data Set (UDS; n=34,848), collected from UDS visits conducted between September 2005 and June 2018. Written and informed consent was obtained from all participants and co-participants for the UDS by the Alzheimer's Disease Centers (ADCs). Among the risk factors in the NACC data, we selected age at visit, gender, years smoked, alcohol dependence, stroke, cardiac-arrest/heart-attack, diabetes, hypertension, hypercholesterolemia, BMI and Geriatric Depression Scale (GDS) score. These were selected based on evidence from previous studies regarding their role in cognitive impairment [1-4,6,8-14,25]. Depression and lifestyle factors such as smoking and alcohol dependence are known to be strongly associated with metabolic disorders [15-19,26], hence were included in this study. Incidence of MCI and all-cause dementia was determined based on clinical diagnosis. Due to low numbers of participants in progression groups, all-cause dementia was analysed instead of specific dementia subtypes.

2.2 Data preparation

Two data preparation approaches were compared in this study i.e., the traditional baseline approach where data was collected from the first patient visit (Fig. 1; Section 2.2.1), and a multi-time-point progression approach where data from multiple visits was collated for each participant (Fig. 2; Section 2.2.2). This was done to reduce bias associated with a single measurement of a given risk factor, and to identify change in cognitive status over time. Individuals included in the analyses were aged ≥40 years.

2.2.1 Baseline analyses

Observations from the first visit were analysed for baseline groups. Three comparisons were made: Healthy (n=12,622) vs. MCI (n=6,685), Healthy (n=12,622) vs. Dementia (n=7,948) and MCI (n=6,685) vs. Dementia (n=7,948). BMI was categorized as underweight-1 (<18.5kg/m^2), normal-2 (18.5-24.99kg/m^2) or overweight-3 (>24.99kg/m^2), and GDS scores were categorized as no depression-1 (<5), mild depression-2 (5-9) and moderate depression-3 (>9). Individuals with missing baseline values (n=7,593) were excluded from the analysis (Fig. 1).

2.2.2 Progression analyses

Three comparisons were assessed for progression analyses: Remained Healthy (n=5,431) vs. Healthy-to-MCI (n=543), Remained Healthy (n=5,431) vs. Healthy-to-Dementia (n=400), and Remained MCI (n=1,141) vs. MCI-to-Dementia (n=809). Figure 2 depicts the data preparation process for the progression groups, and figure 3 depicts the transitions between diagnostic groups over time. The length of time from the first visit to subsequent diagnosis varied among the progression groups, ranging between 3.10-7.00 years

(Supplementary Table 1). Individuals with single observations, dementia at baseline, and those with alternating diagnosis between visits were excluded (n=21,090). Moreover, those with missing values at baseline (n=560) or only having complete observations for a single visit (n=4,871) were also excluded (Fig. 2). Next, a multi-time-point approach was used to determine the status of risk factors given the data from multiple visits as explained in Section 2.2.2.1.

2.2.2.1 Adjusting observations with respect to progression

Numerous participants acquired conditions such as stroke, hypertension, and depression after their baseline visit. Accordingly, these variables are categorised differently at baseline and in subsequent visits. To account for this, we adjusted the values of independent variables to reflect how a particular risk factor developed beyond baseline. Instances where a risk factor developed after the individual progressed to MCI/dementia were not considered.

Observations for age at visit, number of years smoked, and gender were obtained from baseline. Diabetes, hypertension, hypercholesterolemia, alcohol dependence, stroke and heart attack/cardiac arrest are categorised in the NACC data as absent-0, active/recent-1 (occurred within the last year or requiring active management) or inactive/remote-2 (occurred in the past, i.e. more than one year ago but was resolved or there is no treatment currently under way). For these variables, values were set to 0 if all visits were 0, 1 if a single visit was 1, and as 2 if all visits were 2.

BMI was first categorized as underweight-1 ($<18.5 kg/m^2$), normal-2 ($18.5-24.99 kg/m^2$) or overweight-3 ($>24.99 kg/m^2$). Change in BMI was determined by calculating the average of BMI categories (underweight, normal or overweight) across all visits and qualitatively

comparing this to the baseline category to determine increase/decrease/stable progression. Given that BMI can increase/decrease within the same category, we decided to take the average of BMI categories (as opposed to average of absolute BMI values) to represent transition between the groupings. Similarly, GDS scores were categorized as no depression-1 (<5), mild depression-2 (5-9) and moderate depression-3 (>9). Change in GDS was detected by calculating the average of GDS categories (none, mild or moderate) across all visits and qualitatively comparing it to the baseline category to determine improvement/deterioration/maintenance of condition.

2.3 Statistical analysis

Univariate analyses were performed to assess differences in demographic characteristics. The normality of data was assessed with Shapiro-Wilk test. The significance of differences between continuous variables was evaluated using an independent t-test for normally distributed data, and the Mann-Whitney U test for non-normally distributed data. To compare differences for categorical variables a chi-square test was applied. A multivariate logistic regression model was used to explore the relative contributions of the risk factors to MCI and all-cause dementia incidence at baseline, and for progression groups. False discovery rate (FDR) was applied to adjust for multiple hypothesis testing by employing the Benjamini-Yekutieli correction method [27]. False discovery-adjusted *p*-values (FDR *p*) <0.01 were considered statistically significant. Statistical analyses were performed using the 'PredictABEL' package in R studio (Version 1.1.423).

2.3.1 Sensitivity analysis

Sensitivity analyses were performed to assess the validity of our multi-time-point data preparation approach. We analysed the baseline observations of non-converters vs. baseline observations of converters from progression groups in order to assess any statistically significant differences between these groups. Hence, the following comparisons were analysed: (1) baseline observations of individuals who Remained Healthy (n=5,431) vs. baseline observations of Healthy-to-MCI converters (n=543); (2) Baseline observations of individuals who Remained Healthy (n=5,431) vs. baseline observations of Healthy-to-Dementia converters (n=400); and finally, (3) baseline observations of individuals who Remained MCI over time (n=1,141) vs. baseline observations of MCI-to-Dementia converters (n=809).

3 Results

3.1 General characteristics of participants

3.1.1 Baseline groups

Individuals with MCI and dementia were significantly older compared to healthy controls ($p<2.2e-16$, $p<2.2e-16$ respectively), and a higher proportion of them were married ($p=4.1e-16$, $p<2.2e-16$ respectively) and had alcohol dependence ($p=2e-7$, $p=2e-18$ respectively). Univariate analysis showed that participants with MCI were more likely to suffer from various comorbidities compared to healthy controls or dementia patients (Supplementary Table 1). Furthermore, a higher proportion of men suffered from MCI ($p<2.2e-16$), whereas a higher proportion of women suffered from dementia ($p<2.2e-16$).

3.1.2 Progression groups

While analysing data from progression groups, we found that 543 (8.52%) healthy individuals developed MCI over an average duration of 6.7 years, and 400 (6.27%) developed dementia over a mean period of 7 years. Additionally, 809 (41.49%) individuals with MCI developed dementia over an average of 5.5 years. Individuals who remained healthy or MCI, had follow-up data available for an average of 5.4 and 3.1 years respectively (Supplementary Table 1, Supplementary Fig.1). The average number of visits for all the groups ranged from 3.5-6.7 (Supplementary Table 1).

Healthy participants who progressed to MCI or dementia over time were significantly older ($p<2.2e-16$, $p<2.2e-16$ respectively), less educated ($p=0.04$, $p=0.002$ respectively) and a smaller proportion of them were married ($p=0.002$, $p=0.003$ respectively), compared to those who remained healthy. Additionally, those who progressed from Healthy-to-MCI had a higher average of total years smoked ($p=0.003$).

Participants who progressed from MCI-to-Dementia, were significantly older ($p<0.001$), predominantly white ($p<0.001$), more educated ($p=0.01$), and a higher proportion of them were married ($p<0.001$), compared to those with stable MCI diagnosis (Supplementary Table 1).

3.2 Risk factors associated with baseline and progression analyses

Adjusted odds ratios for active/inactive stages of comorbidities were measured against absence of the disease. BMI (underweight/overweight or decreasing/increasing) and

GDS (mild/moderate or decreasing/increasing) categories were measured against normal or stable groups for baseline and progression analyses, respectively.

3.2.1 Baseline vs. progression analyses for Healthy and MCI

Baseline analysis, illustrated in Fig. 4A, found that age (FDR $p<0.0001$), gender (male; FDR $p<0.0001$), active diabetes (FDR $p<0.0001$), hypertension (FDR $p<0.001$) and hypercholesterolemia (FDR $p < 0.001$), history of stroke (FDR $p<0.0001$) and depression (mild and moderate; FDR $p<0.0001$) were significantly associated with an increased risk of MCI when compared to healthy individuals, while being overweight (BMI >24.99kg/m^2; $p<0.001$) was significantly associated with a reduced risk of MCI. Given the progression group with individuals who Remained Healthy across all visits, versus those who developed MCI (Fig. 4B), only age (FDR $p<0.0001$) and increasing GDS score (from no depression to mild/moderate, or from mild to moderate) (FDR $p=0.006$) were significantly associated with an increased risk of Healthy-to-MCI progression. Hence, there is a general reduction in the number of risk factors when analysing the progression groups, as compared to baseline analyses.

3.2.2 Baseline vs. progression analyses for Healthy and Dementia

Next, we compared baseline analysis for Healthy vs. Dementia with progression analysis i.e. Remained Healthy vs. Healthy-to-Dementia progression. As shown in Fig. 5A, at baseline, age (FDR $p<0.0001$), gender (male; FDR $p<0.0001$), active and inactive alcohol dependence (FDR $p<0.0001$), history of stroke (FDR $p<0.0001$), being underweight (BMI

<18.5kg/m^2; FDR p<0.001) and depression (mild and moderate; FDR p<0.0001) were significantly associated with an increased dementia risk. Moreover, being overweight (BMI >24.99kg/m^2; FDR p<0.0001) was significantly associated with a reduced dementia risk. These results are primarily in concordance with existing literature [1-3]. In contrast, upon comparing individuals who Remained Healthy, to those who progressed from Healthy-to-Dementia, age (FDR p<0.0001) and increasing GDS score (FDR p=0.0065) were associated with a significantly increased risk of progressing to dementia (Fig. 5B). Again, there is a general reduction in the number of risk factors when analysing the progression groups, as compared to baseline analyses. Moreover, the identified risk factors are consistent with those for Healthy-to-MCI progression.

3.2.3 Baseline vs. progression analyses for MCI and Dementia

We then focused on potential risk factors that were differentially associated with MCI and dementia. Fig. 6 illustrates baseline (MCI vs. Dementia) and progression analyses (Remained MCI vs. MCI-to-Dementia). At baseline, we found that age (FDR p<0.0001) was significantly associated with an increased risk of having dementia. Male gender (FDR p<0.0001), active diabetes (FDR p=0.003) and hypertension (FDR p=0.002) and being overweight (BMI >24.99kg/m^2; FDR p<0.0001) on the other hand were associated with a reduced risk of dementia. When we considered individuals, who remained MCI over time to MCI-to-Dementia converters, age (FDR p<0.0001) was significantly associated with increased risk of progression. Furthermore, active hypertension (FDR p=0.002) was significantly associated with a reduced risk of MCI-to-Dementia progression, compared to

individuals with stable MCI diagnosis. The reduced number of risk factors obtained, as compared to baseline analyses, was again observed.

3.3 Sensitivity analysis

Sensitivity analysis was conducted to examine whether there were any differences between the baseline values of stable and progression groups, and if they were consistent with the progression analyses. Individuals from the progression groups were identified in the baseline samples, and subsequently non-converters were analysed against converters.

3.3.1 Baseline of stable healthy vs. baseline of Healthy-to-MCI

Baseline analysis of individuals who remained healthy over time vs. those who progressed to MCI, revealed significant differences between these two groups in terms of depression. Magnitude of adjusted odds ratio for depression (GDS 5-9) 1.48 ($p=0.05$; Supplementary Fig. 2A) was consistent with the progression analysis (Fig. 4B).

3.3.2 Baseline of stable healthy vs. baseline of Healthy-to-Dementia

When we compared baseline observations of individuals who remain healthy over time to those who progressed to dementia, the magnitude of adjusted odds ratio of 1.35 ($p=0.22$) for depression (GDS 5-9) was again consistent with the progression analysis (Fig. 5B).

3.3.3 Baseline of stable MCI vs. baseline of MCI-to-Dementia

Lastly, comparing baseline observations of individuals who remained MCI over time versus those who progressed from MCI-to-Dementia, the adjusted odds ratio associated with active hypertension 0.85 (p=0.11; Supplementary Fig. 2C), was consistent with the progression analysis (Fig. 6B).

Overall, although variability in statistical significance was observed, the results from sensitivity analyses revealed similar magnitudes of effect sizes as those of the progression groups. This consistency in results increases confidence, that our proposed data preparation approach is more robust and minimizes bias.

4 Discussion

In this study, we showed that when identifying risk factors for MCI and dementia, analyses based solely on baseline data (at a single-time-point), generally reproduce the existing findings within the literature. Specifically, amongst the risk factors considered in this study, majority of significant results from baseline analyses (age, hypercholesterolemia, alcohol dependence, stroke, BMI and depression) were in accordance with existing literature [1-4,6,8,11-14,26,28-33]. However, in some cases contrasting outcomes were observed. We report that at baseline men are at higher risk of dementia (Fig. 5). Prevalence of dementia is known to be higher in women [31] due to longevity in women and faster rate of disease progression in men [34]. This may reflect inherent bias in self-selecting populations in clinical trials or research generally. However, some large population-based

studies have reported no gender differences in dementia incidence, or different risk profiles for dementia progression in men and women [35,36]. In the NACC cohort, a higher proportion of men had a parent with cognitive impairment compared to women. Additionally, men had higher average of BMI and number of years of smoking, and a greater proportion of them suffered from comorbidities such as diabetes, hypercholesterolemia, history of stroke, cardiac arrest/heart attack and alcohol dependence. Collectively, these factors might have influenced the outcomes of analyses associated with the data. These add further evidence in supporting our hypothesis that baseline analysis can be unreliable, and the outcomes may vary across different studies. Comparing MCI and dementia at baseline, active stages of diabetes and hypertension were associated with reduced dementia risk in contrast to existing literature [1-3,6,8].

As opposed to the baseline approach, our proposed multi-time-point progression model highlights the features most significantly associated with cognitive decline. Cognitive status of participants and risk factors are prone to change over time, therefore risk factors measured longitudinally may have a different effect on risk associated with disease severity [15,37-39]. Additionally, collating information from multiple visits, and analysing trends of BMI and GDS scores, better represent physiological changes over time. We found a substantially reduced number of risk factors for progression groups compared to the baseline groups. Specifically, our multi-time-point data preparation approach in assessing temporal changes in depressive symptoms shows that increased GDS score (vs. stable) was significantly associated with an increased risk of Healthy-to-MCI and Healthy-to-Dementia progression. Additionally, the proportion of stable healthy individuals (19.7%) with a clinical diagnosis of depression was significantly lower than Healthy-to-MCI (35.4%) and Healthy-to-Dementia converters (46.5%, $p<0.001$).

Several studies adapting varying methodologies have explored the depression-MCI/dementia relationship. Consequently, evidence exists for depression as a prodrome, a risk factor and an accompanying symptom of cognitive impairment [15-19]. A clinical study showed no relationship between the level of depression and neuropathologic markers of dementia [40], whereas others have found common inflammatory pathology in both depression and dementia [41]. There are fewer studies on mechanism, with inconsistent findings reported [19]. In the case of MCI, a higher prevalence of depression is observed in hospital-based (vs. population-based) studies [42], highlighting the link between different diagnostic and selection criteria in different settings, and potentially contrasting outcomes. Studies analysing the trajectory of depression (based on relapsing-remitting and number of symptoms) found varying dementia risk depending on the course of depression [43]. This indicates the importance of optimising research design and approaches to improve reproducibility and reliability of research findings.

Despite contradictory reports, mid-life hypertension is a well-accepted risk factor for dementia [1-3,8]. However, in the present study, comparing stable MCI and MCI-to-Dementia converters exhibited surprising results. Active hypertension (vs. absent) was associated with a significantly reduced risk of MCI-to-Dementia progression. In the case of late-life hypertension, conflicting findings have been reported so far, even in clinical trial studies that evaluated late-life antihypertensive treatment [44]. Our findings suggest that, while a history of hypertension may be associated with dementia, late-life active hypertension is associated with reduced dementia risk, when compared to stable MCI. A cerebral blood-flow study suggested that in the case of essential hypertension, although there is an increase in cerebrovascular resistance, it is accompanied by a compensatory mechanism that maintains normal cerebral blood flow [45,46]. In chronic hypertension,

however, changes in cerebrovascular autoregulation occurred as a result of cerebrovascular resistance. It was observed, that due to the structural changes in cerebral small vessels, the limits of autoregulation were adjusted to high pressure levels. This indicated that despite the increased risk of ischaemia, this adaption of the brain protects it from high intravascular pressure [45,46]. Analysing sub-groups of individuals with active hypertension, based on the type of drug treatment might offer more insight [47] and will be explored in more detail in future work.

From a wider perspective, contradictory results with respect to dementia risk factors are manifold and may be explained by a combination of methodological differences. These include study design, diagnostic procedures used to determine grouping, non-standardized categorization of variables and outcomes, and selection criteria [18,48]. Moreover, factors such as variability in cohort characteristics, time of measurement, and referral patterns may also result in different estimates of dementia prevalence [18,48].

More importantly, differences in data preparation methods, and inconsistencies in reporting of such methods in published literature also contributes to outcome variability. Indeed, studies have shown that varying interpretations of risk factors, and relating 'disease risk' to risk factors that are measured at ≤2 time points can lead to misleading results and introduce bias [20]. Therefore, in the present study we utilised a multi-time-point data preparation approach, using the NACC dataset and cardiometabolic risk factors as an exemplar, and demonstrated that few established risk factors are significantly associated with risk of progression to MCI and dementia.

A limitation of this study is the smaller sample size of progression groups which may have led to insignificant p-values associated with the risk factors [49]. It is possible that depression and hypertension have larger effect sizes compared to the other risk factors

analysed in this study, hence they were identified as significant variables in both the baseline and progression analyses. Therefore, there is a need for large population studies from birth cohorts which would be invaluable to better understand the relationship of different risk factors throughout the lifecycle. Further, we have investigated all-cause dementia and not specific dementia subtypes, such as Alzheimer's disease or vascular dementia, and this might potentially have led to some mixing effects. Further work should focus on specific diseases to reduce such effects, and application to other dementia datasets in order to validate our approach. Moreover, it is worth exploring risk factors associated with alternating diagnosis to identify which significant factors emerge. This will in turn help us understand the similarities, and differences in disease progression between stable and unstable converters.

Overall, our proposed multi-time-point data preparation approach in analysing risk factors for neurological disorders, such as dementia, may provide more robust and reproducible results, which are urgently needed in current biomedical and clinical research [50]. Subsequently, this approach can also be adapted in other scientific fields.

Acknowledgements

1. This project was supported by the European Union's INTERREG VA Programme, managed by the Special EU Programmes Body (SEUPB (Centre for Personalised Medicine, IVA 5036)), with additional support by the Northern Ireland Functional Brain Mapping Project Facility (1303/101154803), funded by invest Northern Ireland and the University of Ulster (K.W.-L.), Alzheimer's Research UK (ARUK) NI Pump Priming (M.B.,S.T.,K.W.-L.,P.M.), Ulster University Research Challenge Fund (M.B.,S.T.,K.W.-L.,M.B.), the Dr George Moore Endowment for Data Science at Ulster University (M.B.), and the COST Action Open Multiscale Systems Medicine (OpenMultiMed) supported by COST (European Cooperation in Science and Technology) (K.W.-L.). The views and opinions expressed in this paper do not necessarily reflect those of the European Commission or the SEUPB.
2. The NACC database is funded by NIA/NIH Grant U01 AG016976. NACC data are contributed by the NIA-funded ADCs: P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P50 AG008702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P50 AG047266 (PI Todd Golde, MD, PhD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P50 AG005134 (PI Bradley Hyman, MD, PhD), P50 AG016574 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Thomas Wisniewski, MD), P30 AG013854 (PI M. Marsel Mesulam, MD), P30 AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P50 AG047366 (PI Victor Henderson, MD, MS), P30 AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P50 AG005131 (PI James Brewer, MD, PhD), P50 AG023501 (PI Bruce Miller, MD), P30 AG035982 (PI Russell Swerdlow,

MD), P30 AG028383 (PI Linda Van Eldik, PhD), P30AG053760 (PI Henry Paulson, MD, PhD), P30 AG010124 (PI John Trojanowski, MD, PhD), P50 AG005133 (PI Oscar Lopez, MD), P50 AG005142 (PI Helena Chui, MD), P30 AG012300 (PI Roger Rosenberg, MD), P30 AG049638 (PI Suzanne Craft, PhD), P50 AG005136 (PI Thomas Grabowski, MD), P50 AG033514 (PI Sanjay Asthana, MD, FRCP), P50 AG005681 (PI John Morris, MD), P50 AG047270 (PI Stephen Strittmatter, MD, PhD)."

3. We would like to acknowledge the efforts of our colleague Niamh McCombe, PhD researcher at Ulster University for proofreading this manuscript.

References

[1] G.A. Edwards Iii, N. Gamez, G.J. Escobedo, O. Calderon, I. Moreno-Gonzalez, Modifiable Risk Factors for Alzheimer's Disease. Frontiers in Aging Neuroscience 2019;11:146.

[2] K. Deckers, M.P.J. van Boxtel, O.J.G. Schiepers, M. de Vugt, J.L. Munoz Sanchez, K.J. Anstey, C. Brayne, J. Dartigues, K. Engedal, M. Kivipelto, K. Ritchie, J.M. Starr, K. Yaffe, K. Irving, F.R.J. Verhey, S. Kohler, Target risk factors for dementia prevention: a systematic review and Delphi consensus study on the evidence from observational studies. Int J Geriatr Psychiatry 2015;30:234-246.

[3] M. Baumgart, H.M. Snyder, M.C. Carrillo, S. Fazio, H. Kim, H. Johns, Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective. Alzheimer's & Dementia 2015;11:718-726.

[4] A. Solomon, M. Kivipelto, B. Wolozin, J. Zhou, R.A. Whitmer, Midlife serum cholesterol and increased risk of Alzheimer's and vascular dementia three decades later. Dementia & Geriatric Cognitive Disorders 2009;28:75-80.

[5] M.M. Mielke, P.P. Zandi, H. Shao, M. Waern, S. Ostling, X. Guo, C. Bjorkelund, L. Lissner, I. Skoog, D.R. Gustafson, The 32-year relationship between cholesterol and dementia from midlife to late life. Neurology 2010;75:1888-1895.

[6] M. Crowe, A. Sartori, O.J. Clay, V.G. Wadley, R. Andel, H. Wang, P. Sawyer, R.M. Allman, Diabetes and cognitive decline: investigating the potential influence of factors related to health disparities. Journal of Aging & Health 2010;22:292-306.

[7] M. Sano, C.W. Zhu, H. Grossman, C. Schimming, Longitudinal Cognitive Profiles in Diabetes: Results From the National Alzheimer's Coordinating Center's Uniform Data. J Am Geriatr Soc 2017;65:2198-2204.

[8] F. Bermejo-Pareja, J. Benito-Leon, E.D. Louis, R. Trincado, E. Carro, A. Villarejo, A.G. de la Camara, Risk of incident dementia in drug-untreated arterial hypertension: a population-based study. J Alzheimer's Dis 2010;22:949-958.

[9] M.M. Corrada, K.M. Hayden, A. Paganini-Hill, S.S. Bullain, J. DeMoss, C. Aguirre, R. Brookmeyer, C.H. Kawas, Age of onset of hypertension and risk of dementia in the oldest-old: The 90+ Study. Alzheimer's & Dementia 2017;13:103-110.

[10] N. Qizilbash, J. Gregson, M.E. Johnson, N. Pearce, I. Douglas, K. Wing, S.J.W. Evans, S.J. Pocock, BMI and risk of dementia in two million people over two decades: a retrospective cohort study. The Lancet Diabetes & Endocrinology 2015;3:431-436.

[11] J. Ilomaki, N. Jokanovic, E.C.K. Tan, E. Lonnroos, Alcohol Consumption, Dementia and Cognitive Decline: An Overview of Systematic Reviews. Current Clinical Pharmacology 2015;10:204-212.

[12] J. Rehm, O. Hassan, S.E. Black, K.D. Shield, M. Schwarzinger, Alcohol use and dementia: a systematic scoping review. Alzheimer's Research & Therapy 2019;11:1.

[13] G. Zhong, Y. Wang, Y. Zhang, J.J. Guo, Y. Zhao, Smoking is associated with an increased risk of dementia: a meta-analysis of prospective cohort studies with investigation of potential effect modifiers. PLoS ONE [Electronic Resource] 2015;10:e0118333.

[14] S.L. Tyas, L.R. White, H. Petrovitch, G. Webster Ross, D.J. Foley, H.K. Heimovitz, L.J. Launer, Mid-life smoking and late-life dementia: the Honolulu-Asia Aging Study. Neurobiol Aging 2003;24:589-596.

[15] K.P. Muliyala, M. Varghese, The complex relationship between depression and dementia. Annals of Indian Academy of Neurology 2010;13:S69-73.

[16] W. Wiels, C. Baeken, S. Engelborghs, Depressive Symptoms in the Elderly-An Early Symptom of Dementia? A Systematic Review. Frontiers in Pharmacology 2020;11:34.

[17] S. Bennett, A.J. Thomas, Depression and dementia: Cause, consequence or coincidence? Maturitas 2014;79:184-190.

[18] F. Panza, V. Frisardi, C. Capurso, A. D'Introno, A.M. Colacicco, B.P. Imbimbo, A. Santamato, G. Vendemiale, D. Seripa, A. Pilotto, A. Capurso, V. Solfrizzi, Late-life depression, mild cognitive impairment, and dementia: possible continuum? American Journal of Geriatric Psychiatry 2010;18:98-116.

[19] D. Enache, B. Winblad, D. Aarsland, Depression in dementia: epidemiology, mechanisms, and treatment. Current Opinion in Psychiatry 2011;24:461-472.

[20] C. Frost, I.R. White, The effect of measurement error in risk factors that change over time in cohort studies: do simple methods overcorrect for ‘regression dilution’? Int J Epidemiol 2005;34:1359-1368.

[21] P.J. Moore, T.J. Lyons, J. Gallacher, Alzheimer's Disease Neuroimaging Initiative, Using path signatures to predict a diagnosis of Alzheimer's disease. PLoS ONE [Electronic Resource] 2019;14:e0222212.

[22] E. Baker, E. Iqbal, C. Johnston, M. Broadbent, H. Shetty, R. Stewart, R. Howard, S. Newhouse, M. Khondoker, R.J.B. Dobson, Trajectories of dementia-related cognitive decline in a large mental health records derived patient cohort. PLoS ONE [Electronic Resource] 2017;12:e0178562.

[23] X. Ding, M. Bucholc, H. Wang, D.H. Glass, H. Wang, D.H. Clarke, A.J. Bjourson, L.R.C. Dowey, M. O'Kane, G. Prasad, L. Maguire, K. Wong-Lin, A hybrid computational approach for efficient Alzheimer's disease classification based on heterogeneous data. Scientific Reports 2018;8:9774.

[24] D.L. Beekly, E.M. Ramos, G. van Belle, W. Deitrich, A.D. Clark, M.E. Jacka, W.A. Kukull, NIA-Alzheimer's Disease Centers, The National Alzheimer's Coordinating Center (NACC) Database: an Alzheimer disease database. Alzheimer Disease & Associated Disorders 2004;18:270-277.

[25] A.L. Jefferson, A.S. Beiser, J.J. Himali, S. Seshadri, C.J. O'Donnell, W.J. Manning, P.A. Wolf, R. Au, E.J. Benjamin, Low cardiac index is associated with incident dementia and Alzheimer disease: the Framingham Heart Study. Circulation 2015;131:1333-1339.

[26] L. Shi, J.A. Morrison, J. Wiecha, M. Horton, L.L. Hayman, Healthy lifestyle factors associated with reduced cardiometabolic risk. Br J Nutr 2011;105:747-754.

[27] Y. Benjamini, D. Yekutieli, The Control of the False Discovery Rate in Multiple Testing under Dependency. The Annals of Statistics 2001;29:1165-1188.

[28] R. Guerreiro, J. Bras, The age factor in Alzheimer's disease. Genome Medicine 2015;7:106.

[29] M. Ganguli, B. Fu, B.E. Snitz, T.F. Hughes, C.H. Chang, Mild cognitive impairment: incidence and vascular risk factors in a population-based cohort. Neurology 2013;80:2112-2120.

[30] D.L. Bachman, P.A. Wolf, R. Linn, J.E. Knoefel, J. CobbS, A. Belanger, R.B. D'Agostino, L.R. White, Prevalence of dementia and probable senile dementia of the Alzheimer type in the Framingham Study. Neurology 1992;42:115-119.

[31] C. Berr, J. Wancata, K. Ritchie, Prevalence of dementia in the elderly in Europe. European Neuropsychopharmacology 2005;15:463-471.

[32] T. Anttila, E. Helkala, M. Viitanen, I. Kareholt, L. Fratiglioni, B. Winblad, H. Soininen, J. Tuomilehto, A. Nissinen, M. Kivipelto, Alcohol drinking in middle age and subsequent risk of mild cognitive impairment and dementia in old age: a prospective population based study. Bmj 2004;329:539.

[33] A. Viswanathan, E.A. Macklin, R. Betensky, B. Hyman, E.E. Smith, D. Blacker, The Influence of Vascular Risk Factors and Stroke on Cognition in Late Life: Analysis of the NACC Cohort. Alzheimer Disease & Associated Disorders 2015;29:287-293.

[34] S. Todd, S. Barr, M. Roberts, A.P. Passmore, Survival in dementia and predictors of mortality: a review. Int J Geriatr Psychiatry 2013;28:1109-1124.

[35] S. Artero, M.L. Ancelin, F. Portet, A. Dupuy, C. Berr, J.F. Dartigues, C. Tzourio, O. Rouaud, M. Poncet, F. Pasquier, S. Auriacombe, J. Touchon, K. Ritchie, Risk profiles for mild cognitive impairment and progression to dementia are gender specific. Journal of Neurology, Neurosurgery & Psychiatry 2008;79:979-984.

[36] A. Ruitenberg, A. Ott, J.C. van Swieten, A. Hofman, M.M. Breteler, Incidence of dementia: does gender make a difference?. Neurobiol Aging 2001;22:575-580.

[37] W. Willett, An overview of issues related to the correction of non-differential exposure measurement error in epidemiologic studies. Stat Med 1989;8:1031-1040.

[38] K.C. Cain, R.A. Kronmal, A.S. Kosinski, Analysing the relationship between change in a risk factor and risk of disease. Stat Med 1992;11:783-797.

[39] M.W. Knuiman, M.L. Divitini, J.S. Buzas, P.E. Fitzgerald, Adjustment for regression dilution in epidemiological regression analyses. Ann Epidemiol 1998;8:56-63.

[40] R.S. Wilson, A.W. Capuano, P.A. Boyle, G.M. Hoganson, L.P. Hizel, R.C. Shah, S. Nag, J.A. Schneider, S.E. Arnold, D.A. Bennett, Clinical-pathologic study of depressive symptoms and cognitive decline in old age. Neurology 2014;83:702-709.

[41] B.E. Leonard, Inflammation, depression and dementia: are they connected? Neurochem Res 2007;32:1749-1756.

[42] Z. Ismail, H. Elbayoumi, C.E. Fischer, D.B. Hogan, C.P. Millikin, T. Schweizer, M.E. Mortby, E.E. Smith, S.B. Patten, K.M. Fiest, Prevalence of Depression in Patients With Mild Cognitive Impairment: A Systematic Review and Meta-analysis. JAMA Psychiatry 2017;74:58-67.

[43] S.S. Mirza, F.J. Wolters, S.A. Swanson, P.J. Koudstaal, A. Hofman, H. Tiemeier, M.A. Ikram, 10-year trajectories of depressive symptoms and risk of dementia: a population-based study. The Lancet.Psychiatry 2016;3:628-635.

[44] S. Oveisgharan, V. Hachinski, Hypertension, executive dysfunction, and progression to dementia: the canadian study of health and aging. Archives of Neurology 2010;67;2:187-192.

[45] S.S. Kety, J.H. Hafkenschiel, W.A. Jeffers, I.H. Leopold, H.A. Shenkin, The blood flow, vascular resistance, and oxygen consumption of the brain in essential hypertension. J Clin Invest 1948;27:511-514.

[46] S.S. Kety, C.F. Schmidt, Cerebral blood flow and cerebral oxygen consumption in five patients with hypertension. Am J Med Sci 1946;212:124.

[47] B. McGuinness, S. Todd, P. Passmore, R. Bullock, Blood pressure lowering in patients without prior cerebrovascular disease for prevention of cognitive impairment and dementia. Cochrane Database of Systematic Reviews 00403:(4)-2009 Ot 07.

[48] M. Ganguli, W.A. Kukull, Lost in translation: epidemiology, risk, and Alzheimer disease. Arch Neurol 2010;67:107-111.

[49] M.S. Thiese, B. Ronna, U. Ott, P value interpretations and considerations. Journal of Thoracic Disease 2016;8:E928-E931.

[50] M. Baker, 1,500 scientists lift the lid on reproducibility. Nature 2016;533:452-454.

Figure legends

1. Fig. 1: Data preparation process for baseline groups.
2. Fig. 2: Data preparation process for progression groups within the NACC dataset. For categorical variables 0 represents absent, 1 represents active/recent state (occurred within the last year or requiring active management) and 2 represents inactive/remote state (occurred in the past, more than one year ago but was resolved or there is no treatment currently under way). (NACC: National Alzheimer's Coordinating Center; MCI: mild cognitive impairment; BMI: Body Mass Index; GDS: Geriatric Depression Scale).
3. Fig. 3: Transitions between different diagnostic groups over time in the NACC dataset.
4. Fig. 4: Forest plots of adjusted odds ratios for potential risk factors of baseline Healthy vs. MCI, and stable Healthy vs. conversion to MCI. (A) Outcome: Healthy vs. MCI; (B) Outcome: Remained Healthy vs. Healthy-to-MCI progression.
5. Fig. 5: Forest plots of adjusted odds ratios for potential risk factors of baseline Healthy vs. Dementia, and stable Healthy vs. conversion to Dementia. (A) Outcome: Healthy vs. Dementia; (B) Outcome: Remained Healthy vs. Healthy-to-Dementia progression.
6. Fig. 6: Forest plots of adjusted odds ratios for potential risk factors of baseline MCI vs. Dementia, and stable MCI vs. conversion to Dementia. (A) Outcome: MCI vs. Dementia; (B) Outcome: Remained MCI vs. MCI-to-Dementia progression.