Author/s:
Brodaty, H;Mothakunnel, A;de Vel-Palumbo, M;Ames, D;Ellis, KA;Reppermund, S;Kochan, NA;Savage, G;Trollor, JN;Crawford, J;Sachdev, PS

Title:
Influence of population versus convenience sampling on sample characteristics in studies of cognitive aging

Date:
2014-01-01

Citation:
Brodaty, H., Mothakunnel, A., de Vel-Palumbo, M., Ames, D., Ellis, K. A., Reppermund, S., Kochan, N. A., Savage, G., Trollor, J. N., Crawford, J. & Sachdev, P. S. (2014). Influence of population versus convenience sampling on sample characteristics in studies of cognitive aging. Annals of Epidemiology, 24 (1), pp.63-71. https://doi.org/10.1016/j.annepidem.2013.10.005.

Accepted Manuscript

Influence of population versus convenience sampling on sample characteristics in studies of cognitive aging

Henry Brodaty, MD, DSc Annu Mothakunnel, BPsych Melissa de Vel-Palumbo, MSc David Ames, MD Kathryn A. Ellis, PhD Simone Reppermund, PhD Nicole A. Kochan, PhD Greg Savage, PhD Julian Trollor, MD John Crawford, PhD Perminder S. Sachdev, MD, PhD

PII: S1047-2797(13)00375-X

DOI: 10.1016/j.annepidem.2013.10.005

Reference: AEP 7563

To appear in: *Annals of Epidemiology*

Received Date: 31 May 2013

Revised Date: 16 September 2013

Accepted Date: 3 October 2013

Please cite this article as: Brodaty H, Mothakunnel A, de Vel-Palumbo M, Ames D, Ellis KA, Reppermund S, Kochan NA, Savage G, Trollor J, Crawford J, Sachdev PS, Influence of population versus convenience sampling on sample characteristics in studies of cognitive aging, *Annals of Epidemiology* (2013), doi: 10.1016/j.annepidem.2013.10.005.

Influence of population versus convenience sampling on sample characteristics in studies of cognitive aging

Henry Brodaty, MD, DSc, [1,2] Annu Mothakunnel, BPsych, [1] Melissa de Vel-Palumbo, MSc, [1] David Ames, MD, [3,4] Kathryn A. Ellis, PhD, [3,4,5] Simone Reppermund, PhD,[1] Nicole A. Kochan, PhD,[1,6] Greg Savage, PhD,[7] Julian Trollor, MD,[1,8] John Crawford, PhD,[1] and Perminder S. Sachdev, MD, PhD[1,6]

[1]Centre for Healthy Brain Ageing, University of New South Wales, Sydney, NSW
[2]Dementia Collaborative Research Centre, University of New South Wales, Sydney, NSW
[3] National Ageing Research Institute, Parkville, VIC
[4] University of Melbourne Academic Unit for the Psychiatry of Old Age, Kew, VIC
[5] Mental Health Research Institute, Parkville, VIC
[6] Neuropsychiatric Institute, Prince of Wales Hospital, Randwick, NSW
[7]Department of Psychology and ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney, NSW
[8]Department of Developmental Disability Neuropsychiatry, University of New South Wales, Sydney, NSW

Corresponding Author: Professor Henry Brodaty, Dementia Collaborative Research Centre, 302 AGSM, UNSW, Sydney, NSW 2052 Australia. h.brodaty@unsw.edu.au.

Running title: Sampling method in cognitive aging studies

Abstract word count: 200

Main text word count: 4481

Number of tables: 3

Number of figures: 3

Abstract

Purpose: We examined whether differences in findings of studies examining mild cognitive impairment (MCI) were associated with recruitment methods by comparing sample characteristics in two contemporaneous Australian studies, using population-based and convenience sampling.

Method: The Sydney Memory and Aging Study invited participants randomly from the electoral roll in defined geographic areas in Sydney. The Australian Imaging, Biomarkers and Lifestyle Study of Ageing recruited cognitively normal (CN) individuals via media appeals and MCI participants via referrals from clinicians in Melbourne and Perth. Demographic and cognitive variables were harmonized, and similar diagnostic criteria were applied to both samples retrospectively.

Results: CN participants recruited via convenience sampling were younger, better educated, more likely to be married and have a family history of dementia and performed better cognitively than those recruited via population-based sampling. MCI participants recruited via population-based sampling had better memory performance and were less likely to carry the APOE ε4 allele than clinically referred participants, but did not differ on other demographic variables.

Conclusion: A convenience sample of normal controls is likely to be younger and better functioning and that of an MCI group likely to perform worse than a purportedly random sample. Sampling bias should be considered when interpreting findings.

MeSH heading keywords: Mild cognitive impairment, Aging, Epidemiologic Studies, Epidemiologic Research Design, Selection Bias, Patient Selection, Apolipoprotein E4, Neuropsychological Tests

List of abbreviations and acronyms

- MCI: Mild Cognitive impairment
- CN: Cognitively normal
- AD: Alzheimer's disease
- MMSE: Mini-Mental State Examination
- MAS: Memory and Ageing Study
- AIBL: Australian Imaging, Biomarkers and Lifestyle
- NESB: Non-English speaking background
- SMC: Subjective memory complaint
- LM: Logical Memory
- RAVLT: Rey Auditory Verbal Learning Test
- CVLT-II: California Verbal Learning Test-Second Edition
- BNT: Boston Naming Test

Influence of population versus convenience sampling on sample characteristics in studies of cognitive aging

Epidemiological studies differ regarding findings about rates of decline and prognosis of mild cognitive impairment (MCI), an intermediate state between normal aging and dementia. Differences in study findings could be associated with differences in sampling methods. Studies may employ population-based sampling which aims to select a random group of participants who are representative of the population of interest or convenience sampling which involves engaging volunteers who are selected due to ease of recruitment and willingness to participate and clinical referrals who are selected to maximize the sampling of specific types of disorders.

Convenience sampling of cognitively normal (CN) participants is vulnerable to self-selection bias as those who seek out opportunities to participate in cognitive research may be more capable and motivated than randomly recruited CN participants. Consistent with this, studies have shown that CN convenience samples tend to be younger [1-3] and better educated [1-4] than those recruited via population-based sampling and more likely to have a family history of Alzheimer's disease (AD) [3], probably reflecting their personal interest and motivation.

Clinically referred samples are also susceptible to bias as they may contain people who have better access to health care due to socioeconomic factors or have more complex or severe conditions [5]. Consistent with such a bias, clinically referred MCI participants tend to be better educated, [3, 6-9] and more likely to be married and living independently than people with MCI in the wider population, [6, 9]. They also tend to be younger, possibly because doctors are more likely to refer younger patients to specialty clinics [3, 6, 7, 10, 11], though some studies have found them to be older [8, 12]. Additionally, clinically recruited MCI and AD participants are more likely to carry the APOE ε4 allele [3, 7] and more likely to decline faster suggesting more aggressive brain pathology [3].

Such demographic differences between population-based and convenience samples could lead to invalid research conclusions. For example, younger age of convenience samples could affect the

validity of research examining neuropathology of MCI, effects of anti-AD medications, and APOE genotype [7, 11]. Similarly, higher levels of education observed in convenience samples may be associated with greater levels of cognitive reserve and could lead to incorrect conclusions regarding MCI progression rates.

There are mixed findings as to whether sampling methods are associated with differences in cognitive performance of CN samples. CN convenience samples outperformed population-based participants on the Mini-Mental State Examination (MMSE) [2, 3], and on a vocabulary task possibly due to higher education levels [4], but not on reasoning or word recall tasks [4].

Similarly, there is mixed evidence as to whether sampling methods are associated with differences in cognitive performance of MCI samples. There is some evidence that population-based MCI samples outperform clinic samples (solely based on MMSE) possibly because participants from clinics have a more aggressive or advanced form of MCI [6, 7, 12]. By contrast, others found no difference between clinic and population samples on the MMSE, memory tasks, or executive function tasks [8], or found that clinic samples performed better possibly due to higher levels of cognitive reserve though this result was not corrected for differences in sample age and education [3].

Potential cognitive differences merit further investigation. If there is consistent evidence that CN convenience samples outperform population-based samples then studies comparing MCI participants against a convenience sampled normal reference group would exaggerate their degree of cognitive impairment. Additionally, evidence indicating that clinically-referred MCI samples cognitively underperform population-based MCI samples would suggest that clinic samples consist of a select group of patients with a form of MCI more likely to progress to dementia and do not represent the heterogeneity of MCI in the general population.

Additionally, as convenience sampling is more selective than population-based sampling, one may expect less inter-individual variability among convenience samples. In one study convenience

samples showed less variance than population-based samples in some quality of life and social relationship variables but not cognitive measures [4].

This study examined the relationship between recruitment method and demographic and cognitive characteristics of the purportedly random electoral roll based sample used in the Sydney Memory and Aging Study (MAS) [13] and a convenience sample of CN participants recruited via media advertisement and clinical referrals with MCI used in the Australian Imaging and Biomarkers Lifestyle (AIBL) study of ageing [14]. We hypothesized that CN and MCI participants in the MAS sample would be older, less educated, less likely to be married and less likely to be living independently than those in the AIBL study. Additionally we hypothesized that the AIBL study would contain more CN participants with a family history of memory problems or dementia and more MCI participants who were APOE ε4 carriers than the MAS. There were no clear predictions regarding differences between the samples on cognitive performance or on inter-individual variability on cognitive measures.

Methods

Protocols

Baseline data were obtained from two Australian longitudinal studies of cognitive aging: the MAS and the AIBL study. The MAS [13] was initiated in 2005 and conducted in Sydney. Participants were recruited from the community via the electoral roll; (in Australia, voting is compulsory). A random sample of 8914 people living in the federal government electorates of Kingsford-Smith and Wentworth aged between 70 and 90 years were invited by letter to participate. Of these, 1772 people (20%) agreed to participate and were screened over the phone to assess their eligibility; 735 people were excluded because they were ineligible or no longer agreed to participate. The final sample had 1037 participants.

The AIBL study [14], which was initiated in 2006, aimed to recruit 200 participants with AD, 100 participants with MCI, and 700 healthy participants over the age of 60 from Melbourne and

Perth. Healthy participants were largely recruited via a media appeal and participants with MCI or AD largely via clinical referral. The total sample contained 1112 participants.

There were some differences in study exclusion criteria. The AIBL study excluded people with non-AD dementia whereas the MAS excluded those with any form of dementia. Unlike the MAS, the AIBL study excluded people with current depression, Parkinson's disease, symptomatic stroke, uncontrolled diabetes, or regular alcohol use exceeding two standard drinks per day for women or four for men. The AIBL study did not contain participants from non-English speaking backgrounds (NESBs) whereas the MAS included people from NESBs who spoke sufficient English to complete the assessment. Full details of MAS and AIBL exclusion criteria have been reported previously [13, 14].

Ethics Approval

Written informed consent was obtained from all participants. The MAS was approved by the Ethics Committees of the University of New South Wales and the South Eastern Sydney and Illawarra Area Health Service. The AIBL study was approved by the institutional ethics committees of Austin Health, St Vincent's Health, Hollywood Private Hospital and Edith Cowan University.

Sample reclassification process

To allow comparison between the samples, 211 AIBL participants diagnosed with AD at baseline were excluded (by definition no MAS participant had dementia at baseline). A further 445 AIBL participants outside the 70-90 year age range were excluded in order to match the samples' age ranges.

Participants were reclassified as CN or MCI using common MCI diagnostic criteria: cognitive impairment and subjective memory complaint (SMC) in the absence of dementia or significant functional impairment [15]. As the studies differed in how they originally defined cognitive impairment, criteria were harmonized so that cognitive impairment was defined for all participants as scores lower than or equal to 1.5 standard deviations below published normative data on at least *one* of the cognitive measures outlined in table 1 (excluding estimated IQ measures) . SMC was

harmonized by using responses to a similar question in both studies which asked about memory difficulties. As this question was not asked of clinically referred AIBL study participants, we inferred that these participants also had SMCs. Our MCI classification encompassed participants with impairments on both amnestic and non-amnestic measures and on single or multiple cognitive domains. Participants without cognitive impairment were classified as CN regardless of SMC or functional impairment.

Participants were deemed unclassifiable if they were from NESBs (because of the questionable validity of using normative data derived from mainly native English speakers); if they had cognitive impairment *and* functional impairment (defined as a score of one or more on the Clinical Dementia Rating Scale [35] domains of home and hobbies, personal care or community affairs); or if they had cognitive impairment but *no* SMC. Participants who had missing scores on more than one neuropsychological measure but had no cognitive impairment were also unclassifiable.

Figures 1a, b show how the MAS and AIBL samples were reclassified. There were 569 MAS participants classified as CN, 161 classified as MCI, and 307 deemed unclassifiable of whom 113 were from NESBs, 10 had cognitive impairment and functional impairment, 11 had cognitive impairment but missing functional impairment data, 126 had cognitive impairment but no SMC, 28 had cognitive impairment but missing SMC data, and 19 were missing neuropsychological data. In the AIBL study, there were 256 participants meeting CN criteria of whom 13 were excluded as they entered the study as clinical referrals rather than healthy participants recruited via media advertisement, 135 participants meeting MCI criteria of whom 82 were excluded as they entered the study as participants recruited via media advertisement rather than as clinical referrals, and 65 participants deemed unclassifiable of whom 9 had cognitive impairment and functional impairment, 2 had cognitive impairment but missing functional impairment data, 38 had cognitive impairment but no SMC, 11 had cognitive impairment but missing SMC data, and 5 were missing neuropsychological data.

Harmonization Process

Variables of interest were harmonized retrospectively as the MAS and AIBL study used different tests and scoring procedures. Harmonization occurred in three steps. First, variables measured identically in both studies required no adjustments, namely: age, sex, APOE genotype, MMSE [36], Logical Memory (LM) [25], Digit Symbol Coding [18] and Phonemic fluency[28, 30] scores. Secondly, where studies measured similar constructs but differed in how these were assessed, responses were re-coded into a comparable scoring system. These were marital status, education, living arrangements, and family history of dementia/memory problems. Thirdly, where cognitive measures differed in the neuropsychological tests or versions of the neuropsychological tests, decisions were made by expert consensus (from experienced neuropsychologists: NAK, GS, KAE) to consider measures as acceptable equivalents (see table 1). Manual derived z-scores were used to compare the Rey Auditory Verbal Learning Test (RAVLT) [19] and California Verbal Learning Test Second Edition (CVLT-II) [24]. For semantic fluency, the AIBL score [30] (an aggregate for "animals" and "boys' names") was halved to allow comparison with the MAS [31] (which used "animals" only); the validity of halving the aggregate score was checked in a subsample of 39 AIBL participants who had separate scores available for animals and boys' names. The ratio of animals to boys' names was 1:1 suggesting that halving the aggregate score was a reasonable approximation. Raw scores were used when comparing other measures.

Measures

At the conclusion of harmonization the following variables were available for analysis: age, sex, education, marital status, living arrangements, APOE genotype, family history of dementia/memory problems, MMSE scores and neuropsychological test scores shown in table 1.

Statistical Analyses

Comparisons between the MAS and AIBL samples were conducted separately for CN and MCI participants. Scores on neuropsychological variables were examined for normality; all variables had skew less than ±1.5. Missing data were estimated via single imputation using the Expectation

Maximisation method for imputing missing data [37]. Single imputation is justified as being adequate since the percentage of missing data was 4.3% for the Boston Naming Test (BNT) and less than 1% for other neuropsychological measures, well below the commonly accepted upper level for single imputation of about 10% [38].

Levene's test was used to determine whether variances in the two samples were equivalent for neuropsychological measures. Where Levene's test was significant, we used t-tests that did not assume equal variance rather than ANOVA. Analyses were repeated with age, sex and education as control variables. ANCOVA was used for variables with equal variances; for variables with unequal variances we compared adjusted values (obtained via regression analyses) using the t-test procedure that does not assume equal variances. Participants included in the analyses (i.e., classified as CN or MCI) were compared on age, sex and education with those deemed unclassifiable. SPSS version 20 was used for statistical analysis.

Results

CN comparisons

Table 2 shows comparisons between MAS and AIBL study CN samples. The MAS sample was older and less educated than the AIBL study sample and had fewer participants who were married or in de facto relationships and more participants who were widowed or had never been married. The samples did not differ in sex ratios, living arrangements or the percentage of APOE ε4 carriers. Fewer MAS participants had a family history of dementia/memory problems.

The AIBL study sample outperformed the MAS sample on the MMSE, Digit Symbol Coding, List Learning, List memory (short and long delay), Semantic fluency and the BNT. The samples did not differ on LM (I and II) or phonemic fluency or on estimated IQ. These results were unchanged when analyses were repeated adjusting for age, sex and education, with the exception of estimated IQ on which the MAS sample were significantly better, $F(1, 806) = 4.32$, $p = .04$.

MCI comparisons

As shown in table 3, the MAS and AIBL study MCI samples did not differ on age, sex ratios, education, marital status, living arrangements, or family history of dementia/memory problems. However, there were more APOE ε4 carriers in the AIBL study sample.

The MAS sample outperformed the AIBL study sample on the MMSE, List learning, List memory (short and long delay), and LM (I and II). The AIBL study sample outperformed the MAS sample on the BNT. The samples did not differ on estimated IQ, Digit symbol coding, phonemic or semantic fluency performance. The results remained unchanged after controlling for age, sex and education.

Effect size

Figure 2 shows the effect size (Cohen's d) of differences between the MAS and AIBL study samples on cognitive measures. Among CN participants, most differences were small to moderate whereas among MCI participants they were mostly moderate to large.

Comparison of sample variances

Table 2 shows that CN participants in the MAS displayed more variance than AIBL study participants on estimated IQ, the MMSE and BNT, whereas AIBL study participants displayed greater variance on LM (I and II). When controlling for age, sex and education the AIBL study sample also displayed greater variance on list memory (short delay), $F(1,865) = 5.10$, $p=0.02$. Table 3 shows that among MCI participants, the MAS sample showed more variance on List Memory (short delay), though this was no longer significant when controlling for age, sex and education. The AIBL study sample displayed more variance on the MMSE, and also on digit symbol coding after controlling for age, sex and education, $F(1,212) =3.94$, $p=0.05$.

Comparisons between unclassifiable and included participants

Unclassifiable MAS participants were older, $t(1035)=4.13$, $p <.001$, and less educated, χ^2 (3, N=1037) = 8.69, $p=.03$, than MAS participants who were included in the analyses, but did not differ on sex ratios, $p=.17$. Unclassifiable AIBL study participants were older than those who were included, $t(454)= 2.21$, $p=.03$, but did not differ on education, χ^2 (3, N=453) =4.066, $p=.25$, or sex ratios, $p=.92$.

Sensitivity analysis using altered MCI diagnostic criteria

As a substantial proportion of participants were unclassifiable because of lack of SMC or missing SMC data, a sensitivity analysis was conducted using altered MCI diagnostic criteria that did not require presence of SMC. Results on demographic measures remained largely unchanged except that the trend for AIBL MCI participants to be more likely to be married or in a de facto relationship became significant, *p*=.04*,* as did the trend for AIBL MCI participants to be more likely to have a family history of dementia or memory problems, *p*=.01. Results on cognitive measures remained unchanged, except that the AIBL MCI sample showed greater variance than the MAS MCI sample on digit symbol coding, F=3.83, *p=.05,* and that there were no longer significant differences in variance between the MAS and AIBL MCI samples on list memory at short delay, F=1.92, *p*=.17.

Sensitivity analyses comparing AIBL MCI participants recruited via advertisement with clinically referred AIBL MCI and with MAS MCI participants

As only clinically referred AIBL MCI participants were included in the analyses and AIBL MCI participants recruited via advertisement were excluded, a sensitivity analysis was conducted to compare the demographic and cognitive characteristics of these two groups. There were no significant differences between the groups on demographic measures, however clinically referred AIBL MCI participants performed significantly worse than AIBL MCI participants recruited via advertisement on the following cognitive measures: MMSE, t(128)=-6.40, p<.001; Digit Symbol Coding, t(128)=3.35, p=.001; Logical Memory I , t(128)=5.31,p<.001; Logical Memory II, t(128)=6.53, p<.001; Semantic fluency, t(128)=3.27, p<.001; List learning, t(128)=7.34, p<.001; List Memory (Short delay), t(128)=7.86, p<.001; and List Memory (Long delay), t(128)=8.59, p<.001.

An additional sensitivity analysis was conducted comparing AIBL MCI participants recruited via advertisement and MAS MCI participants. AIBL MCI participants recruited via advertisement were significantly younger than MAS MCI participants, t(241)=2.78, *p*=.006, and significantly more likely to have had a family history of memory problems of dementia, *p*=.01. The groups did not differ on any other demographic measures. However, AIBL MCI participants recruited via advertisement

significantly outperformed MAS MCI participants on a number of cognitive measures including the MMSE t(241)=3.6, p<.001, digit symbol coding t(241)=4.45, p<.001, List learning t(241)=.2.66, p<.01, BNT t(241)=6.04, p<.001, Logical Memory I t(241)=2.21, p=.03, Logical Memory II t(241)=2.99, p=.003, and Semantic fluency t(214)=5.84, p<.001.

Discussion

We compared MAS and AIBL study samples to examine whether differences in their recruitment methods were associated with differences in the demographic and cognitive characteristics of their samples. The hypothesis that the MAS sample would be older, less educated, less likely to be married and less likely to be living independently than the AIBL study sample was supported among CN participants but not among MCI participants. The hypothesis that AIBL study CN participants would be more likely to have a family history of dementia/memory problems than MAS CN participants was supported, as was the hypothesis that AIBL study MCI participants would be more likely to be APOE ε4 carriers than MAS MCI participants. Additionally we observed cognitive differences between the MAS and AIBL study samples.

As expected and confirming previous research [1-4], MAS CN participants were older, less educated, and less likely to be married or in a de facto relationship than AIBL study CN participants. Also we confirmed that CN convenience samples had more participants with a family history of dementia/memory problems [3]. Contrary to expectations, the MAS sample did not contain more participants in assisted living arrangements compared to the AIBL study. Both samples had few participants in assisted living arrangements, possibly because these people tend to have impairments or co-morbid conditions which excluded them from participating in either study. Our results suggest that healthy participants who respond to media appeals for participants in cognitive research are more capable (being younger and better educated) and possibly more motivated to do so (due to a family history of dementia) compared to those recruited via population-based sampling.

Contrary to our predictions and to previous research [3, 6, 7, 9-11], the MAS MCI sample were not older, less educated, less likely to be married or less likely to be living independently than

the AIBL study MCI sample. Some previous studies specifically recruited population-based MCI participants from minority groups [8], or from non-respondents to the initial study invitation [3], possibly resulting in a less selective population-based sample thereby increasing demographic differences with clinic MCI samples. However, we did confirm our hypothesis that the AIBL study sample would contain more APOE ε4 carriers than the MAS sample, which was also consistent with earlier research [3, 7].

We observed cognitive differences between the samples. Among CN participants, the MAS sample generally had poorer cognitive performance than the AIBL study sample; they were worse on the MMSE echoing earlier findings [2, 3], and were also worse on other memory, processing speed/attention and language tasks. The opposite pattern was observed among MCI participants, the MAS sample outperformed the AIBL study sample on the MMSE consistent with earlier findings [6, 7, 12], they also outperformed AIBL study participants on other memory tasks. Poorer memory performance and more APOE ε4 carriers in the AIBL study sample are consistent with the likely greater prevalence of preclinical AD in clinically referred MCI samples compared to MCI samples recruited from the general population.

For most cognitive measures the MAS and AIBL study samples did not differ in their variances. Where differences occurred, the pattern of their direction was inconsistent. This is unsurprising as a previous study [4] which reported more variance in random than convenience samples only did so for quality of life and relationship measures and not for cognitive measures.

Study limitations must be addressed. First, the MAS sample was not a truly random sample. Only 20% of people responded to the study invitation. This was lower than response rates in other population-based studies of cognitive aging (e.g., 87% in the Baseline Study of Seniors [4] and 61% in the Mayo Clinic Study of Aging [39]) and may have introduced self-selection bias as MAS participants were better educated and more likely to live in their own homes compared to similarly aged people in the same geographical area [13]. Had response rates been higher, we may have obtained a more representative sample and observed greater demographic differences in the expected direction.

Furthermore, a large number of participants were deemed unclassifiable and excluded from analyses. Included participants were younger than excluded participants and within the MAS sample were also better educated, suggesting that those included in our analyses were not representative of the overall MAS and AIBL study samples. However, it should be noted that the majority of excluded participants had cognitive impairment but no SMC or missing SMC data. When MCI diagnostic criteria were altered so as to include these participants in the analysis the results remained largely unchanged indicating a reasonable degree of confidence in our findings despite the high exclusion rate.

Another limitation was that we restricted the AIBL MCI participants to the 53 (39.2%) who were clinically referred and excluded the 82 (60.8%) who were responders to advertisements in order to ensure homogeneity in groups being compared. These two groups did not differ on demographic measures however the clinically referred group performed worse on many cognitive measures. Additionally the excluded AIBL MCI participants were more likely to have a family history of dementia (implying a motivation for participation) and to outperform MAS MCI participants on a number of cognitive measures.

The effects of harmonization on our comparisons are unknown. For example, the assumption that manual derived normative scores for the RAVLT and CVLT-II were valid equivalents may be flawed. One study reported that patients with brain injuries attained lower standardized scores on the CVLT-II compared to the RAVLT possibly because they made less use of semantic clustering strategies available in the CVLT-II [40]. This finding, together with research suggesting that MCI is associated with reduced semantic clustering [41, 42], could provide an alternative account for why AIBL study MCI participants had lower standardized scores on list learning and list memory compared to MAS MCI participants. However, given that the MAS MCI sample also outperformed the AIBL study sample on identically administered memory measures (LM and the MMSE), it is unlikely that differences on list memory measures were wholly due to the effects of harmonization.

Also, we were unable to fully harmonize SMC criteria. SMCs were inferred among clinically referred AIBL study participants but, for harmonization purposes, were categorized according to MAS participants' response to a single question. Although other studies [43, 44] have determined SMC by endorsement of a single question about memory difficulties it is not known how well this captures SMC compared to referral information.

Finally, in addition to differences in sampling methods, differences observed between the MAS and AIBL study samples may have been influenced by differences between the geographical populations from which they were drawn. Australian census data for people aged 70-89 indicate a higher proportion of 80-89 year olds in the local electoral districts of Sydney sampled, compared to Melbourne and Perth, (39.04% vs. 35.03%) [45, 46]. Also, the electoral districts of Sydney sampled had a higher proportion of people with tertiary education compared to Melbourne and Perth (9.73% vs. 6.18%) [47, 48]. This could explain why we failed to find MAS MCI participants to be less educated than AIBL study MCI participants and also may have reduced the education differences we observed among CN participants.

An important consideration when extrapolating results of longitudinal studies to the general population is appreciation of differences in sampling methods which may differ for those with and without cognitive symptoms. Our findings indicate that among the cognitively normal, a convenience sample of responders to advertisements was younger and better educated and possibly more motivated because of their increased propensity to dementia (e.g. family history of dementia and so linked to their higher proportion who carry the apolipoprotein E ε4 allele) than a population based sample. Our results are conservative as the population based sample's 20% participation rate to the initial mail-out is likely to have biased recruitment to the more motivated. In general non-respondents to surveys have less education and lower cognitive scores and are more disabled [see 2]. The seemingly paradoxical reversals of differences between the two samples for cognitively impaired participants are explicable in that AIBL MCI participants were recruited via clinic referrals and hence more likely to perform worse on cognitive tests.

An implication for the design of future studies is that where samples are recruited from clinics there is a higher likelihood of decline in cognitive performance and progression to dementia. As disease modifying drug trials for Alzheimer's disease have been unsuccessful to date, efforts are now focusing on prevention and early intervention trials for MCI[49]. The biases demonstrated here could compromise extrapolation of results to the general population.

In conclusion, we that found differences in sampling methods were associated with differences in sample characteristics. Convenience sampling of CN participants led to self-selection bias whereby those recruited via media advertisement had more frequent family history of dementia, differences in demographic variables and better cognitive performance than those recruited by population-based sampling. Additionally, clinically referred MCI samples were biased towards including participants with poorer memory performance and who carried the APOE ε4 allele. Sampling bias should be taken into account when interpreting studies of MCI and cognitive aging, as studies which use convenience samples of CN participants recruited via advertisement and clinically referred MCI participants may exaggerate differences between CN and MCI categories compared to studies which use population-based sampling.

Acknowledgements

We thank all participants and their informants for their enthusiastic support. We also thank Kristan Kang, Joanne Robertson, Lance Macaulay, and the MAS and AIBL study research teams. This study was supported by CSIRO under the Preventative Health Flagship. The MAS is supported through NHMRC Program Grant (ID 568969). The AIBL study receives support from the Science Industry Endowment Fund. A complete account of AIBL study funding is available at www.aibl.csiro.au.

References

1. Dixon R, Wahlin A, Maitland S, Hultsch D, Hertzhog C, Backman L. Episodic memory change in late adulthood: generalizability across samples and performance indices. Mem Cognition. 2004;32(5):768-78. doi: 10.3758/BF03195867.

2. Ganguli M, Lytle M, Reynolds M, Dodge H. Random versus volunteer selection for a community-based study. J Gerontol A Bio Sci Med Sci. 1998;53(1): M39-46. doi: 10.1093/gerona/53A.1.M39.

3. Whitwell J, Wiste H, Weigand S, Rocca W, Knopman D, Roberts R, et al. Comparison of imaging biomarkers in the Alzheimer Disease Neuroimaging Initiative and the Mayo Clinic Study of Aging. Arch Neurol. 2012;69(5):614-23. doi: 10.1001/archneurol.2011.3029

4. Hultsch D, MacDonald S, Hunter M, Maitland S, Dixon R. Sampling and generalisability in developmental research: comparison of random and convenience samples of older adults. Int J Behav Dev. 2002;26(4):345-59.doi: 10.1080/01650250143000247.

5. Delgado-Rodriguez M, Llorca J. Bias. J Epidemiol Commun H.2004;58(8):635-41. doi: 10.1136/jech.2003.008466.

6. Barnhart R, Belle G, Edland S, Kukull W, Borson S, Raskind M, et al. Geographically overlapping Alzheimer's disease registries: comparisons and implications. J Geriatr Psychiatry Neurol. 1995;8(4):203-8.

7. Tsuang D, Kukull W, Sheppard L, Barnhart R, Peskind E, Edland S, et al. Impact of sample selection on APOE epsilon4 allele frequency: a comparison of two Alzheimer's disease samples. J Am Geriatr Soc. 1996;44(6):704-7.

8. Farias S, Mungas D, Reed B, Harvey D, DeCarli C. Progression of mild cognitive impairment to dementia in clinic- vs community-based cohorts.Arch Neurol. 2009;66(9):1151-7. doi: 10.1001/archneurol.2009.106.

9. Kokmen E, Ozsarfati Y, Beard M, O'Brien P, Rocca W. Impact of referral bias on clinical and epidemiological studies of Alzheimer's disease. J Clin Epidemiol. 1996;49(1):79-83. doi.org/10.1016/0895-4356(95)00031-3.

10. Schneider J, Aggarwal N, Barnes L, Boyle P, Bennett D. The neuropathology of older persons with and without dementia from community versus clinic cohorts. J. Alzheimers Dis. 2009;18(3):691-701. doi: 10.3233/JAD-2009-1227.

11. Schoenmaker N, Van Gool W. The age gap between patients in clinical studies and in the general population: a pitfall for dementia research. Lancet Neurol. 2004;3(10):627-30. doi: 10.1016/s1474-4422(04)00884-1.

12. Andersen F, Engstad T, Straume B, Viitanen M, Halvorsen D, Hykkerud S, et al. Recruitment methods in Alzheimer's disease research: general practice versus population based screening by mail. BMC Med Res Methodol. 2010;10(35). doi: 10.1186/1471-2288-10-35.

13. Sachdev P, Brodaty H, Reppermund S, Kochan N, Trollor J, Draper B, et al. The Sydney Memory and Ageing Study (MAS): methodology and baseline medical and neuropsychiatric characteristics of an elderly epidemiological non-demented cohort of Australians aged 70-90 years. Int Psychogeriatr.2010;22(8):1248-64. doi: 10.1017/S1041610210001067.

14. Ellis K, Bush A, Darby D, De Fazio D, Foster J, Hudson P, et al. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. Int Psychogeriatr. 2009;21(4):672-87. doi: 10.1017/S1041610209009405.

15. Petersen R. Challenges of epidemiological studies of mild cognitive impairment. Alzheimer Dis Assoc Disord. 2004;18(1):1-2. doi:10.1097/00002093-200401000-00001.

16. Nelson H, Willison J. National adult reading test (NART): test manual.2nd ed. Windsor: NFER Nelson; 1991.

17. Wechsler D. Wechsler test of adult reading: examiner's manual. San Antonio, TX: The Psychological Corporation; 2001.

18. Wechsler D. Wechsler adult intelligence scale-III. San Antonio, TX: The Psychological Corporation; 1997.

19. Rey A. L'examen clinique en psychologie. Paris: Presses Universitaires de France; 1964.

20. Harris M, Ivnik R, Smith G. (2002). Mayo's Older Americans Normative Studies: expanded AVLT Recognition Trial norms for ages 57 to 98. J Clin Exp Neuropsyc. 2002;24(2):214–20. doi: 10.1076/jcen.24.2.214.995.

21. Ivnik R, Malec J, Tangalos E, Petersen R, Kokmen E, Kurland L. The Auditory-Verbal Learning Test (AVLT): norms for ages 55 years and older. Psychological Assessment. 1990;2(3): 304–12.

22. Ivnik R et al. Mayo's Older Americans Normative Studies: updated AVLT norms for ages 56 to 97. Clinical Neuropsychology. 1992; 6(supplement): S83–S104. doi 10.1080/13854049208401880.

23. Ivnik, R. J. et al.. Mayo's Older Americans Normative Studies: WAIS-R norms for ages 56 to 97. Clinical Neuropsychology. 1992; 6 (supplement): S1–S30. doi: 10.1080/13854049208401877.

24. Delis D, Kramer J, Kaplan E, Ober B. California verbal learning test-second edition. San Antonio, TX: The Psychological Corporation; 2000.

25. Wechsler D. Wechsler memory scale-III manual (WAIS-3). San Antonio, TX: Harcourt Brace & Company; 1997.

26. Grundman M. et al.(2004). Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials.Arch Neurol. 2004;61(1):59–66. doi: 10.1001/archneur.61.1.59

27. Petersen R. C. Personal communication, 3 October 2006.

28. Benton A. Problems of test construction in the field of aphasia. Cortex. 1967;3:32–58.

29. Tombaugh T, Kozak J, Rees L. Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. Arch Clin Neuropsych. 1999; 14(2): 167–77.

30. Delis D, Kaplan E, Kramer J. The Delis-Kaplan executive function system (D-KEFS). San Antonio, TX: Psychological Corporation; 2001.

31. Spreen O, Benton A. Neurosensory centre comprehensive examination for aphasia manual (NCCEA). Victoria: University of Victoria; 1996.

32. Kaplan E, Goodglass H, Weintraub S. The Boston Naming Test. Baltimore: Lippincott, Williams & Wilkins; 2001.

33. Fastenau P, Denburg N, Mauer B. Parallel short forms of the Boston Naming Test: Psychometric properties and norms for older adults. J Clin Exp Neuropsychol. 1998; 20(6): 828-834. doi:10.1076/jcen.20.6.828.110.

34. Saxton J, Ratcliff G, Munro C, Coffey E, Becker J, Fried L, Kuller L. Normative data on the Boston naming test and two equivalent 30-item short forms. Clin Neuropsychol. 2000;14(4): 526–34. doi:10.1076/clin.14.4.526.7204.

35. Morris J. The Clinical dementia rating (CDR): current version and scoring rules. Neurology. 1993; 43(11):2412–4.

36. Folstein M, Folstein S, McHugh P. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12(3): 189–98.

37. Schafer J. Analysis of incomplete multivariate data, Book No. 72, Chapman & Hall series Monographs on Statistics and Applied Probability. London: Chapman & Hall; 1997.

38. Scheffer J. Dealing with Missing Data. Res. Lett. Inf. Math. Sci. 2002; 3, 153-60.

39. Roberts R, Geda Y, Knopman D, Cha R, Pankratz V, Boeve B, et al. The Mayo Clinic Study of Ageing: Design and sampling, participation, baseline measures and sample characterstics. Neuroepidemiology. 2008; 30(1): 58-69. doi: 10.1159/000115751.

40. Stallings G, Boake C, Sherer M. Comparison of the California verbal learning test and the Rey auditory verbal learning test in head-injured patients. J Clin Exp Neuropsychol. 1995; 17(5): 706-12. doi: 10.1080/01688639508405160.

41. Malek-Ahmadi M, Raj A, Small B. Semantic clustering as a neuropsychological predictor for amnestic-MCI. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn. 2011; 18(3): 280–92. doi: 10.1080/13825585.2010.540642.

42. Ribeiro F, Guerreiro M, De Mendonca A. Verbal learning and memory deficits in mild cognitive impairment. J Clin Exp Neuropsychol. 2007; 29(2): 187-97. doi: 10.1080/13803390600629775.

43. Ganguli M, Dodge H, Shen C, DeKosky S. Mild cognitive impairment, amnestic type an epidemiologic study. Neurology. 2004; 63(1): 115-121.

44. Jungwirth S, Weissgram S, Zehetmayer S, Tragl K, Fischer R. VITA: subtypes of mild cognitive impairment in a community-based cohort at the age of 75 years. Int J Geriatr Psychiatry. 2005; 20(5): 452–458. doi: 10.1002/gps.1311.

45. Age 10 Year Age Groups (AGEP) by Age 5 Year Age Groups (AGEP) and Commonwealth Electoral Divisions 2004. [Internet]. Australian Bureau of Statistics; 2006 [generated 2013 April 19]. Available from http://www.abs.gov.au/websitedbs/censushome.nsf/home/tablebuilder.

46. Age 10 Year Age Groups (AGEP) by Age 5 Year Age Groups (AGEP) and ASGC Upper . [Internet]. Australian Bureau of Statistics; 2006 [generated 2013 April 19]. Available from http://www.abs.gov.au/websitedbs/censushome.nsf/home/tablebuilder

47. Non-School Qualification: Level of Education (QALLP) by Age 5 Year Age Groups (AGEP) and Commonwealth Electoral Divisions 2004. [Internet]. Australian Bureau of Statistics; 2006 [generated 2013 April 19]. Available from http://www.abs.gov.au/websitedbs/censushome.nsf/home/tablebuilder.

48. Non-School Qualification: Level of Education (QALLP) by Age 5 Year Age Groups (AGEP) and ASGC Upper**.** [Internet]. Australian Bureau of Statistics; 2006 [generated 2013 April 19]. Available from http://www.abs.gov.au/websitedbs/censushome.nsf/home/tablebuilder.

49. DeKosky S, Williamson J, Fitzpatrick A, Kronmal R, Ives D, Saxton J et al. Gingko biloba for prevention of dementia: a randomized controlled trial. JAMA. 2008; 300 (19):2253-62. doi: 10.1001/jama.2008.683.Legends

Legends

Table 1. Harmonized neuropsychological measures used in the MAS and AIBL study

[a] Estimated IQ measures were not used to determine cognitive impairment.

Table 2. Comparison of MAS and AIBL study CN samples on demographic and cognitive variables

Note. All comparisons shown are based on raw unadjusted scores.

*p<.05, **p<.01, †p<.05 (following Bonferroni correction)

Table 3. Comparison of MAS and AIBL study MCI samples on demographic and cognitive variables

Note. All comparisons shown are based on raw unadjusted scores.

*p<.05, **p<.01

Figure 1a. Flow-chart outlining the sample reclassification process in the MAS

Figure 1b. Flow-chart outlining the sample reclassification process in the AIBL study.

Figure 2. Effect size of significant differences observed between the MAS and AIBL samples for CN and MCI participants. Positive values indicated that the MAS sample outperformed the AIBL sample; negative values indicated that the AIBL sample outperformed the MAS sample. A Cohen's d value of 0.2 indicates a small effect size, 0.5 a moderate effect size, and 0.8 a large effect size. Significant differences between the samples are indicated * p<.05, **p<.01.

Table 1. Harmonized neuropsychological measures used in the MAS and AIBL study

Cognitive domain	MAS neuropsychological measures	AIBL study neuropsychological measures
Estimated IQ[a]	National Adult Reading Test [16] No adjustment	Wechsler Test of Adult Reading [17] No adjustment
Attention/ processing speed	Digit Symbol Coding [18] Adjusted for age (18)	Digit Symbol Coding [18] Adjusted for age (18)
Memory	Rey Auditory Verbal Learning test (RAVLT) [19] Adjusted for age [20-23] Logical Memory Test (Story A) [25] Adjusted for education [26]	California Verbal Learning Test-Second edition (CVLT-II) [24] Adjusted for age and sex using CVLT-II Scoring Assistant Software Logical Memory Test (Story A) [25] Adjusted for education [27]
Executive function (Phonemic Fluency)	Controlled Oral Word Association Test (COWAT) [28] Adjusted for age and education[29]	D-KEFS Letter Fluency [30] Adjusted for age [30]
Language	Semantic Fluency [31] (animals) Adjusted for age and education [29] Boston Naming Test (BNT) [32] (Different 30-item subsets of the BNT used) Adjusted for age [33]	Semantic Fluency [30] (composite score for animals and boys names) Adjusted for age[30] Boston Naming Test (BNT) [32] (Different 30-item subsets of the BNT used) Adjusted for age and education [34]

[a] Estimated IQ measures were not used to determine cognitive impairment.

Table 2. Comparison of MAS and AIBL study CN samples on demographic and cognitive variables

		MAS n=569	AIBL n=243	Levene's F	F or χ^2	p
Demographic Variables						
Age		78.18 (4.66)	76.25 (4.78)	0.07	28.94	<.001**
Sex (% females)		59.6	52.7			.08
Education (%)	0-8 years	13.0	8.3		53.45	<.001**
	9-12 years	51.1	36.4†			
	13-15 years	24.3	22.7			
	>15 years	11.6	32.6			
Marital status (%)	Never married	14.1	4.9†		54.56	<.001**
	Married / de facto	40.6	68.3†			
	Separated/ Divorced	13.2	9.1			
	Widowed	32.1	17.7†			
Living independently (%)		98.4	98.8			1.00
APOE ε4 carrier (%)		20.8	24.7			.23
Family history of dementia or memory problems (%)		24.5	41.2			<.001**
Cognitive Variables						
IQ		108.80 (9.50)	109.74 (6.96)	45.32**	2.46	.12
MMSE		27.85 (1.71)	28.75(1.26)	14.13**	69.56	<.001**
Digit symbol coding		50.57 (12.14)	56.95 (12.27)	0.08	46.70	<.001**
List learning		0.51 (0.98)	1.16 (1.02)	0.39	72.42	<.001**
List Memory short delay		0.70 (0.87)	0.99 (0.97)	2.44	16.81	<.001**
List memory long delay		0.77 (0.93)	0.97 (0.91)	0.71	7.18	.01*
Logical Memory 1		12.31 (3.51)	12.45 (3.98)	6.94**	0.24	.63
Logical Memory 2		10.76 (3.42)	11.05 (4.18)	12.53**	0.88	.35
Phonemic fluency		40.23 (11.55)	41.73 (10.72)	1.76	2.99	.08
Semantic fluency		17.06 (3.98)	19.33 (4.01)	0.17	55.20	<.001**

BNT	26.28 (2.10)	27.77 (1.85)	8.39**	101.40	<.001**

Note. All comparisons shown are based on raw unadjusted scores.

*p<.05, **p<.01, †p<.05 (following Bonferroni correction)

Table 3. Comparison of MAS and AIBL study MCI samples on demographic and cognitive variables

		MAS n=161	AIBL n=53	Levene's F	F or χ^2	p
Demographic Variables						
Age		79.34 (4.50)	78.23 (4.68)		2.37	.13
Sex (% females)		46.0	62.3			.06
Education (%)	0-8 years	15.5	15.1		0.33	.95
	9-12 years	41.6	45.3			
	13-15 years	18.0	15.1			
	>15 years	24.8	24.5			
Marital status (%)	Never married	13.7	7.8		3.35	.34
	Married / de facto	46.6	60.8			
	Separated/ Divorced	9.9	7.8			
	Widowed	29.8	23.5			
Living independently (%)		98.8	98.0			.56
APOE ε4 carrier (%)		31.4	54.7			.003**
Family history of dementia or memory problems (%)		30.4	43.4			.09
Cognitive Variables		M (SD)	M (SD)			
IQ		105.65 (10.81)	106.79 (9.04)	2.45	0.48	.49
MMSE		27.16 (1.92)	25.72 (2.87)	10.30**	17.21	<.001**
Digit symbol coding		44.65 (10.90)	43.78 (13.00)	3.35	0.23	.63
List learning		-0.21 (1.10)	-1.34 (0.94)	2.13	45.50	<.001**
List Memory short delay		-0.14 (1.17)	-1.49 (0.95)	4.00*	58.30	<.001**
List memory long delay		0.07 (1.04)	-1.82 (0.85)	2.26	145.06	<.001**
Logical Memory 1		8.74 (3.84)	6.37 (3.48)	0.32	15.90	<.001**
Logical Memory 2		6.60 (3.80)	3.58 (3.44)	0.47	26.40	<.001**
Phonemic fluency		32.79 (12.69)	35.39 (13.09)	0.01	1.65	.20

Semantic fluency	13.63 (4.09)	14.53 (4.14)	0.04	1.92	.17
BNT	22.58 (4.36)	24.93 (4.44)	1.43	11.46	.001**

Note. All comparisons shown are based on raw unadjusted scores.

*p<.05, **p<.01