

23 **Abstract**

24 *Introduction*

25 Ethnic differences in dementia are increasingly recognised in epidemiological measures and
26 diagnostic biomarkers. Nonetheless, ethnic diversity remains limited in many study populations Here we
27 provide insights into ethnic diversity in open access neuroimaging dementia datasets.

28 *Methods*

29 Datasets comprising dementia populations who underwent neuroimaging assessment with
30 available data on ethnicity were included. Statistical analyses of sample and effect
31 sizes were based on the Cochrane Handbook.

32 *Results*

33 14 databases were included, with 12 studies of healthy and MCI groups, and 11 of dementia groups.
34 Combining all studies, the largest ethnic group was Caucasian (21,512 participants) with the next most
35 common being Afro-Caribbean (1,960), followed by Asian (780). The smallest effect size detectable within
36 the Caucasian group was 0.03, compared to Afro-Caribbean (0.1) and Asian (0.16).

37 *Discussion*

38 Our findings quantify the lack of ethnic diversity in openly available neuroimaging dementia datasets. More
39 representative data would facilitate the development and validation of neuroimaging biomarkers relevant
40 across ethnicities.

41 (150 words)

42

43

44

45

46 **Introduction**

47 The past few decades have seen growing interest in the field of biomarkers for neurodegenerative
48 conditions. The neuroimaging community has led the way in open data ¹, facilitating an
49 explosion of research in neuroimaging biomarkers for dementia ². This interest is in the context
50 of an increasing global burden of neurodegenerative disorders, particularly in relation to the impact of
51 Alzheimer's disease and other dementias on an increasingly ageing population ³. Crucially, it
52 has been estimated that the prevalence of dementia will increase from 57.4 million cases globally
53 in 2019 to 152.8 million cases in 2050 ⁴, posing a considerable risk to global healthcare and society in
54 the near future.

55
56 There has been emerging evidence of ethnic differences amongst dementia populations, not only
57 in prevalence and incidence, but also in cerebrospinal fluid (CSF) and imaging biomarkers ⁵⁻⁶.
58 Nonetheless, many studies remain homogenous in the ethnicity of participants ⁷. This may
59 hinder the translation of results to real world applications. As such, we aimed to provide insights
60 into the ethnic diversity of currently available open neuroimaging dementia databases worldwide.

61

62 **Materials and methods**

63 We compiled and analysed demographic data reported by open access neuroimaging dementia
64 databases. Databases were included if they consisted of (a) patients with a diagnosis of dementia or mild
65 cognitive impairment, (b) had available neuroimaging data, and (c) demographic data including the
66 breakdown of ethnicities. Datasets were identified through online research platforms including

67 the Global Alzheimer’s Association Interactive Network (GAAIN) (<https://www.gaain.org/>),
68 individual database repositories, and via peer-reviewed journal articles. We excluded datasets of
69 solely genetic forms of dementia since these may be associated with specific ethnicities, or include
70 large families that might bias the estimate of the distribution of ethnicities. A total of 46 databases
71 were found but 32 were subsequently excluded as they either only included healthy controls, had
72 no available data on demographics, or did not include neuroimaging data. Given the different
73 definitions of ethnicities available, we took a pragmatic approach using the most widely used terms
74 in the literature that permitted comparison between studies.

75
76 Statistical analyses on combined mean and standard deviation was performed as laid out by the
77 Cochrane Handbook ⁸, and effect sizes calculations were using the pwr package in R (version
78 4.2.2) with ⁹. To compare samples of presumed equal sizes, we performed a power calculation
79 for a two sample t test, estimating the effect size or sample size detectable with 90% power at a
80 significance level (p value) of 0.05. Sample sizes were initially computed by setting a range of
81 effect sizes, while minimum detectable effect sizes for single ethnic groups were then calculated
82 using the aggregated dementia patient populations of different ethnicities from the open access
83 dementia databases.

84

85 **Data availability**

86 Data sharing is not applicable to this article as no new data were created or analysed in this study.

87

88 **Results**

89 **Demographics of dementia databases**

90 A total of 14 dementia neuroimaging datasets were included, separated into the three
91 diagnostic groups, with 12 including healthy participants (Supplementary Table
92 1), 12 including patients with mild cognitive impairment (MCI) (Supplementary Table 2) and 11
93 including patients with dementia (Table 1)¹⁰⁻²². In these tables, two entries for the Alzheimers'
94 Disease Neuroimaging Initiative (ADNI) dataset were made due to the separation of the ADNI-1
95 from ADNIGO and ADNI-2 cohorts. The majority of patients were from North America and Europe,
96 with the two largest databases being from the National Alzheimer's Coordinating Center (NACC)
97 and UK Biobank respectively, in which there were a considerably higher percentage of
98 Caucasians compared to other ethnicities.

99

00 **Effect size analyses**

01 To understand how the breakdown of ethnicity in these datasets could affect research studies,
02 we calculated the sample sizes required for a range of effect sizes. For example, based on a
03 recent systematic review and meta-analysis on fluid biomarkers for Alzheimer's disease⁶, it was
04 found that CSF p-tau₁₈₁ and t-tau levels were significantly higher in the Caucasian population
05 compared to African Americans with MCI, with a standard mean difference of -0.50 (95%CI -0.73
06 to -0.28) and -0.52 (95%CI -0.75 to -0.30) respectively – though bearing in mind these did not
07 necessarily inform the effect size in other biomarkers or ethnicities. Therefore, using an estimated
08 effect size of 0.50 and basing off a power calculation of 90% and significance level of 0.05, the

09 number of patients required to detect a difference was $N=86$ each for two groups of patients of
10 different ethnicities. We went on to calculate sample sizes for a range of effect sizes to obtain a
11 better idea of the sample size to consider when planning future studies. In addition, we assessed whether the
12 available data were sufficient to make comparisons between the Caucasian population and other ethnic
13 groups (Table 2).

14
15 In an alternate approach, using available data for patients with dementia in those datasets
16 combined with similar power calculation of 90% and significance of 0.05, we determined the
17 smallest detectable effect size given currently available data (Table 3). The Caucasian population
18 had the smallest minimum detectable effect size at 0.03 due to its size.

19

20 **Discussion**

21 With the increasing number of studies focusing on ethnic differences in dementia, there is little
22 doubt that more emphasis needs to be placed on the role that ethnic differences play in biomarker
23 research. Our findings suggest that despite the vast amount of comprehensive and high-quality
24 data available worldwide, most participants come from a Caucasian background, limiting
25 comparison to other populations. Considerable numbers of patients are required for assessing
26 small magnitude effect sizes – which becomes particularly important when trying to
27 identify potentially subtle differences between ethnicities. The minimum detectable effect size can
28 therefore act as a guide or threshold towards that end. In fact, the majority of the population sizes
29 were made up of two large databases in the UK and US. We hope these findings can act as a
30 starting point into deciding how to expand representation of different ethnic groups in future

31 studies on dementia.

32

33 Understanding the limitations of currently available data can provide an opportunity to uncover and

34 tackle the challenges associated with ensuring ethnic diversity in studies – stemming from a range

35 of barriers to participation including literacy and language barriers, lack of understanding or

36 misperceptions of the condition and services available, and cultural factors leading to mistrust of

37 services or stigma surrounding the condition ²³. There are some encouraging initiatives to

38 address these challenges, such as those by the UK Dementia Research Institute, which are

39 designed not only to facilitate recruitment of diverse patient cohorts by increasing awareness and

40 broadening inclusion criteria, but also to improve access to healthcare through collaborations with

41 local teams ²⁴.

42

43 We were only able to obtain data for openly available datasets. We know from published data and

44 from personal contacts that many studies use local cohorts, and some large national cohorts are

45 not shared with the wider community. We advocate exploring the barriers to sharing those data,

46 including the concerns of those who have collected and curate those datasets.

47

48 There are several limitations to our study – the first being that we were unable to comment on the

49 representativeness (as opposed to heterogeneity) of the combined characteristics of the

50 populations included in the database. Data on global ethnicity is not readily available and

51 classifications differ between different countries, making it difficult to draw comparisons. Secondly,

52 a considerable number of studies were excluded due to the lack of available demographic data,

53 and those that were included were mainly based in the Western hemisphere, which may mean we
54 have underestimate the non-Caucasian ethnicities actually available.
55
56 In our analysis we assume that datasets can easily be combined. In fact, harmonisation between
57 datasets presents a significant methodological challenge given that protocols differ and site effects
58 need to be modelled²⁵⁻²⁶. This is particularly a challenge for combining neuroimaging data
59 despite the increasing availability of tools for this purpose such as ComBat²⁷.

60

61 **Conclusion**

62 With increasing awareness of the differences between ethnicities in dementia, it is imperative that
63 we begin to prioritise and broaden research to better understand underlying mechanisms, to
64 address the challenges associated with ethnic diversity in studies, and ultimately to pave the way
65 for reliable translation into clinical practice.

66

67 **Funding**

68 This research did not receive any specific grant from funding agencies in the public, commercial,
69 or not-for-profit sectors. This research was supported by the NIHR Cambridge Biomedical
70 Research Centre (NIHR203312). The views expressed are those of the author(s) and not
71 necessarily those of the NIHR or the Department of Health and Social Care.

72

73

74 **Competing Interests**

75 The authors report no competing interests.

76

77 **Supplementary Materials**

78 Supplementary tables 1 and 2 will be available online.

79

80 **References**

- 81 1. Poldrack R., Gorgolewski K. Making big data open: data sharing in neuroimaging. *Nat Neurosci* 17,
82 1510–1517 (2014). <https://doi.org/10.1038/nn.3818>
- 83 2. Rittman T. Neurological update: neuroimaging in dementia. *J Neurol*. 2020 Nov;267(11):3429-3435.
84 doi: 10.1007/s00415-020-10040-0. Epub 2020 Jul 7. PMID: 32638104; PMCID: PMC7578138.
- 85 3. GBD 2016 Neurology Collaborators. Global, regional, and national burden of neurological disorders,
86 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol*. 2019;
87 18(5): 459-480.
- 88 4. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in
89 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019.
90 *Lancet Public Health*. 2022; 7(2): e105-e125.
- 91 5. Wilkins CH, Windon CC, Dilworth-Anderson P et al. Racial and Ethnic Differences in Amyloid PET
92 Positivity in Individuals With Mild Cognitive Impairment or Dementia: A Secondary Analysis of the
93 Imaging Dementia-Evidence for Amyloid Scanning (IDEAS) Cohort Study. *JAMA Neurol*. 2022;
94 79(11): 1139-1147.

- 95 6. Chaudhry A and Rizig M. Comparing fluid biomarkers of Alzheimer's disease between African
96 American or Black African and white groups: A systematic review and meta-analysis. *J Neurol Sci.*
97 2021; 421: 117270.
- 98 7. Vyas MV, Raval PK, Watt JA, Tang-Wai DF. Representation of ethnic groups in dementia trials:
99 systematic review and meta-analysis. *J Neurol Sci.* 2018; 394: 107-111.
- 00 8. Higgins JPT, Green S: *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley Online
01 Library. 2008.
- 02 9. R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical
03 Computing, Vienna, Austria. 2013. Available from: <http://www.R-project.org/>.
- 04 10. Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, Jack CR, Jagust WJ,
05 Shaw LM, Toga AW, Trojanowski JQ, Weiner MW. Alzheimer's Disease Neuroimaging Initiative
06 (ADNI). *Neurology.* 2010; 74(3): 201-209.
- 07 11. Aisen PS, Petersen RC, Donohue M, Weiner MW. ADNI 2 Clinical Core: Progress and Plans.
08 *Alzheimers Dement.* 2015; 11(7): 734-739.
- 09 12. The NIA Alzheimer's Disease Research Centers Program. National Alzheimer's Coordinating
10 Center. 2023. Available from: <https://naccddata.org/requesting-data/data-summary/uds>.
- 11 13. O'Bryant SE, Johnson LA, Barber RC, Braskie MN, Christian B, Hall JR, Hazra N, King K,
12 Kothapalli D, Large S, Mason D, Matsiyevskiy E, McColl R, Nandy R, Palmer R, Petersen M,
13 Philips N, Rissman RA, Shi Y, Toga AW, Vintimilla R, Vig R, Zhang F, Yaffe K; HABLE Study
14 Team. The Health & Aging Brain among Latino Elders (HABLE) study methods and participant
15 characteristics. *Alzheimers Dement (Amst).* 2021;13(1):e12202.
- 16 14. LaMontague PJ, Benzinger TLS, Morris JC, Keefe S, Hornbeck R, Xiong C, Grant E, Hassenstab J,
17 Moulder K, Vlassenko AG, Raichile ME, Cruchaga C, Marcus D. OASIS-3: Longitudinal
18 Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease. 2019.
19 medRxiv. doi: 10.1101/2019.12.13.19014902.

- 20 15. Méndez PC, Calandri I, Nahas F, Russo MJ, Demey I, Martín ME, Clarens MF, Harris P, Tapajoz F,
21 Campos J, Surace EI, Martinetto H, Ventrice F, Cohen G, Vázquez S, Romero C, Guinjoan S,
22 Allegri RF, Sevlever G. Argentina-Alzheimer's disease neuroimaging initiative (Arg-ADNI):
23 neuropsychological evolution profile after one-year follow up. *Arq Neuropsiquiatr.* 2018;76(4):231-
24 240.
- 25 16. Cavedo E, Redolfi A, Angeloni F, Babiloni C, Lizio R, Chiapparini L, Bruzzone MG, Aquino D,
26 Sabatini U, Alesiani M, Cherubini A, Salvatore E, Soricelli A, Vernieri F, Scrascia F, Sinforiani E,
27 Chiarati P, Bastianello S, Montella P, Corbo D, Tedeschi G, Marino S, Baglieri A, De Salvo S,
28 Carducci F, Quattrocchi CC, Cobelli M, Frisoni GB. The Italian Alzheimer's Disease Neuroimaging
29 Initiative (I-ADNI): validation of structural MR imaging. *J Alzheimers Dis.* 2014;40(4):941-52.
- 30 17. Swaddiwudhipong N, Whiteside DJ, Hezemans FH, Street D, Rowe JB, Rittman T. Pre-diagnostic
31 cognitive and functional impairment in multiple sporadic neurodegenerative diseases. *Alzheimers*
32 *Dement.* 2022.
- 33 18. NeuGRID2 consortium. NeuGRID. 2012. Available from:
34 <https://www.neugrid2.eu/index.php/introduction/>
- 35 19. Brueggen K, Grothe MJ, Dyrba M, Fellgiebel A, Fischer F, Filippi M, Agosta F, Nestor P,
36 Meisenzahl E, Blautzik J, Frölich L, Hausner L, Bokde ALW, Frisoni G, Pievani M, Klöppel S,
37 Prvulovic D, Barkhof F, Pouwels PJW, Schröder J, Hampel H, Hauenstein K, Teipel S. The
38 European DTI Study on Dementia - A multicenter DTI and MRI study on Alzheimer's disease and
39 Mild Cognitive Impairment. *Neuroimage.* 2017;144(Pt B):305-308.
- 40 20. NBDC Human Database. NBDC Research ID: hum0043.v1. 2016. Available from:
41 <https://humandbs.biosciencedbc.jp/en/hum0043-v1>
- 42 21. Byun MS, Yi D, Lee JH, Choe YM, Sohn BK, Lee JY, Choi HJ, Baek H, Kim YK, Lee YS, Sohn
43 CH, Mook-Jung I, Choi M, Lee YJ, Lee DW, Ryu SH, Kim SG, Kim JW, Woo JI, Lee DY; KBASE
44 Research Group. Korean Brain Aging Study for the Early Diagnosis and Prediction of Alzheimer's
45 Disease: Methodology and Baseline Sample Characteristics. *Psychiatry Investig.* 2017;14(6):851-863.

- 46 22. Albani D, Marizzoni M, Ferrari C, Fusco F, Boeri L, Raimondi I, Jovicich J, Babiloni C, Soricelli A,
47 Lizio R, Galluzzi S, Cavaliere L, Didic M, Schönknecht P, Molinuevo JL, Nobili F, Parnetti L,
48 Payoux P, Bocchio L, Salvatore M, Rossini PM, Tsolaki M, Visser PJ, Richardson JC, Wiltfang J,
49 Bordet R, Blin O, Forloni G, Frisoni GB; PharmaCog Consortium. Plasma A β 42 as a Biomarker of
50 Prodromal Alzheimer's Disease Progression in Patients with Amnesic Mild Cognitive Impairment:
51 Evidence from the PharmaCog/E-ADNI Study. *J Alzheimers Dis.* 2019;69(1):37-48. doi:
52 10.3233/JAD-180321. PMID: 30149449.
- 53 23. Kenning C, Daker-White G, Blakemore A, Panagioti M, Waheed W. Barriers and facilitators in
54 accessing dementia care by ethnic minority groups: a meta-synthesis of qualitative studies. *BMC*
55 *Psychiatry.* 2017; 316.
- 56 24. UK Dementia Research Institute. Diversity and dementia: how is research reducing health disparities?
57 2022. https://ukdri.ac.uk/uploads/UK-DRI_Dementia_Health_Inequalities_Report_2022.pdf
- 58 25. Lipnicki DM, Lam BCP, Mewton L, Crawford JD, Sachdev PS. Harmonizing Ethno-Regionally
59 Diverse Datasets to Advance the Global Epidemiology of Dementia. *Clin Geriatr Med.* 2023
60 Feb;39(1):177-190. doi: 10.1016/j.cger.2022.07.009. Epub 2022 Oct 18. PMID: 36404030; PMCID:
61 PMC9767705.
- 62 26. Shishegar R, Cox T, Rolls D, Bourgeat P, Doré V, Lamb F, Robertson J, Laws SM, Porter T, Fripp J,
63 Tosun D, Maruff P, Savage G, Rowe CC, Masters CL, Weiner MW, Villemagne VL, Burnham SC.
64 Using imputation to provide harmonized longitudinal measures of cognition across AIBL and ADNI.
65 *Sci Rep.* 2021 Dec 10;11(1):23788. doi: 10.1038/s41598-021-02827-6. PMID: 34893624; PMCID:
66 PMC8664816.
- 67 27. Pomponio R, Erus G, Habes M, Doshi J, Srinivasan D, Mamourian E, Bashyam V, Nasrallah IM,
68 Satterthwaite TD, Fan Y, Launer LJ, Masters CL, Maruff P, Zhuo C, Völzke H, Johnson SC, Fripp J,
69 Koutsouleris N, Wolf DH, Gur R, Gur R, Morris J, Albert MS, Grabe HJ, Resnick SM, Bryan RN,
70 Wolk DA, Shinohara RT, Shou H, Davatzikos C. Harmonization of large MRI datasets for the

71 analysis of brain imaging patterns throughout the lifespan. Neuroimage. 2020 Mar;208:116450. doi:
72 10.1016/j.neuroimage.2019.116450. Epub 2019 Dec 9. PMID: 31821869; PMCID: PMC6980790.

1 Table 1. Table showing the breakdown of demographics data of patients with dementia in databases globally as separated by region.

S/N	Database	Number of patients with dementia	Mean age (SD)	Gender		Ethnicity				
				Male (%)	Female (%)	Caucasian (%)	Afro-Caribbean (%)	Asian (%)	Mixed (%)	Others (%)
North America										
1	ADNI-1 (USA) (10)	192	75.3 (7.5)	52.6	47.4	92.2	4.2	1.0	-	2.1% as Hispanic, 0.5% others
2	ADNIGO and ADNI-2 (USA) (11)	145	74.6 (8.1)	59.0	41.0	91.0	4.1	3.5	1.4	-
3	NACC (USA) (12)	20,053	75.9 (10.8)	48.0	52.0	83.3	9.2	2.2	2.6	0.5% American-Indian, 0.1% Hawaiian/Pacific Islander, 2.1% others
4	HABLE (USA) (13)	185	68.2 (9.9)	45.9	54.1	71.4	28.6	-	-	-
South America										
5	Argentina-ADNI (Argentina) (15)	12	77.9 (5.5)	41.7	58.3	100.0	-	-	-	-
Europe										
6	I-ADNI (Italy) (16)	201	71.8 (8.4)	38.8	61.2	100.0	-	-	-	-
7	UK Biobank (UK) (17)	2778	64.7 (4.2)	45.3	54.7	95.5	1.4	1.4	0.3	0.4% others
8	ARWIBO (Italy) (18)	402	73.5 (8.5)	36.8	63.2	100.0	-	-	-	-
9	EDSD (Italy, Germany, Netherlands) (19)	139	73.0 (8.0)	43.2	56.8	100.0	-	-	-	-
Asia										
10	J-ADNI (Japan) (20)	149	73.7 (6.6)	43.0	57.0	-	-	100.0	-	-
11	WMH-AD (Taiwan) (from GAAIN)	43	77.2 (7.7)	25.6	74.4	-	-	100.0	-	-
12	KBASE (South Korea) (21)	87	73.0 (8.1)	31.0	69.0	-	-	100.0	-	-
Total		24,368	74.4 (10.7)							

1
2
3 Table 2. Table showing the sample sizes required for specific effect sizes to be obtained based on power calculations of 90% and significance level of 0.05, with subsequent columns showing whether comparisons between ethnic groups can be performed based on currently available data.

Effect Size	Sample size required	Caucasian vs Afro-Caribbean	Caucasian vs Asian	Caucasian vs Mixed	Caucasian vs Others
0.5	86	☐	☐	☐	☐
0.3	235	☐	☐	☐	☐
0.2	527	☐	☐	☐	☐
0.1	2103	☐	☐	☐	☐
0.05	8407	☐	☐	☐	☐
0.01	210150	☐	☐	☐	☐

1
2 Table 3. Table showing the smallest effect sizes detected using population sizes based on available ethnicity data from dementia databases with similar parameters of 90% power and significance level of 0.05.

Ethnicity group	Total number of patients with dementia	Smallest effect size detected
Caucasian	21512	0.03
Afro-Caribbean	1960	0.10
Asian	780	0.16
Mixed	532	0.20
Others	138	0.39