HHS Public Access

Author manuscript

Clin Geriatr Med. Author manuscript; available in PMC 2024 February 01.

Published in final edited form as:

Clin Geriatr Med. 2023 February ; 39(1): 177–190. doi:10.1016/j.cger.2022.07.009.

Harmonizing ethno-regionally diverse datasets to advance the global epidemiology of dementia

Darren M. Lipnicki, PhD[1], **Ben C. P. Lam, PhD**[1], **Louise Mewton, PhD**[1], **John D. Crawford, PhD**[1], **Perminder S. Sachdev, MD, PhD**[1,2]

[1]Centre for Healthy Brain Ageing, University of New South Wales, Sydney, Australia

[2]Neuropsychiatric Institute, The Prince of Wales Hospital, Sydney, Australia

Keywords

harmonization; dementia; prospective; retrospective; integrative data analysis; neuropsychological tests; neuroimaging

Introduction

Age is the biggest risk factor for cognitive impairment (CI) and dementia, and the global societal and financial burdens these conditions impose are rising as the world's population ages.[1] Globally, the number of people with dementia is estimated to reach around 150 million by 2050, with the greatest increases expected to occur in developing regions, including Africa.[1] Research on CI and dementia is lacking in many low- and middle-income countries (LMICs).[2] However, the research results from one country or population do not apply to another, with reported differences in the epidemiology of CI and dementia between countries,[3] as well as between different races/ethnicities within countries.[4] Research into how cognitive decline can be slowed and CI and dementia ultimately prevented are thus necessarily a global effort, using large samples with data from different ethno-regions. Resource and co-ordination limitations mean that data on this scale will typically not come from a single source. Rather, such data must be collated from across multiple unique sources focused on particular countries or regions.

The data needed to understand the epidemiology, etiology, and risk and protective factors for CI and dementia comprise a vast array of types, including, but not limited to, demographics, diagnoses, cognitive or neuropsychological test results, medical histories, lifestyle variables like physical activity, substance use and diet, functional status, neuroimaging, and biomarkers. Each data type can be assessed in many ways, and collaborative efforts that

Corresponding Author: University of New South Wales, Level 1, AGSM (G27), Gate 11, Botany Street, UNSW SYDNEY NSW 2052, Darren Lipnicki; d.lipnicki@unsw.edu.au.

Disclosure Statement: Nothing to disclose.

use data from multiple sources are faced with the challenge of making these data comparable so they can be pooled for analysis or more accurately compared.

In this review we discuss how data used in dementia and cognitive impairment research can be made more comparable by harmonization. We cover the benefits and challenges of harmonization, and outline broad retrospective and prospective approaches. We also describe harmonization for particular data types, focusing on neuropsychological test results and neuroimaging, but also including dementia diagnoses, behavioral and psychological symptoms of dementia instruments, and electroencephalography measures.

Discussion

What is data harmonization? Qualitative and quantitative approaches

Harmonization is the process by which data for similar measures or constructs from different sources are made more comparable, or inferentially equivalent.[5] The type of harmonization process needed to achieve comparability depends upon the sort of data involved, and may be qualitative or quantitative.

Qualitative approaches lead to data from different sources having a common format, such as the same range of response options or categories, sometimes requiring a transformation process.[6,7] Examples of this approach include:

- Choosing an item from each source that best represents the measure or construct of interest, e.g., different questions addressing subjective cognitive decline (for a more detailed account, see Box 1)
- Creating a categorical variable by choosing cut-points for different continuous scales measuring the same construct, e.g., classifying the presence of current depression based on a score of 6+ on the Geriatric Depression Scale (GDS) used by one source, and a score of 16+ on the Centre for Epidemiological Studies depression scale (CES-D) used by another source (for a more detailed account see Table S11 in Lipnicki et al.[8])
- Collapsing response categories in the data for some sources to make them similar to those for data from another source with fewer response categories, e.g., self-rated health scales with different numbers of response categories (see Table 1).

Quantitative harmonization is needed for more complex data types and often requires statistical processing to bring them to a common format.[9] Statistical harmonization is typically required for data from cognitive or neuropsychological tests, of which there are hundreds that differ on characteristics like the particular cognitive abilities assessed, and the depth to and mode by which they are assessed.[10] A detailed account of approaches to the statistical harmonization of neuropsychological test scores is given in a later section.

Benefits of harmonized data

Harmonization is an often-necessary step before integrative data analysis, in which individual participant level data from multiple sources are analyzed simultaneously. Integrative data analysis techniques such as mega-analysis and individual participant data

meta-analyses can overcome some of the limitations associated with single studies or meta-analyses of aggregated study data.[6,9] The benefits of harmonized data thus include the capacity to:

- Pool data from different sources, which increases the sample size and thereby the statistical power:
  - This is particularly important when analyzing rare conditions, characteristics, or outcomes, given the increased absolute numbers of individuals with these (for details see Hussong et al.[6])
  - Pooling data can similarly increase the number of participants from subgroups that may be typically underrepresented in single studies.[6]
- Make more accurate comparisons across data sources using measures that are more similar:
  - This is particularly relevant for investigations of commonalities and differences in factors contributing to CI and dementia across different countries, regions of different economic development, or different races/ethnicities[11] (for examples of relevant research studies see Box 2).
- Conduct validation of results or replication across multiple data sources.[5]

Other benefits associated with harmonization include the opportunity for extended use of existing datasets through collaborative projects where data are shared.[7] Indeed, data sharing has become of increasing importance, with many publishers and funders now encouraging or requiring data sharing, for example, the publishing company Elsevier[12] and the National Institutes of Health, USA.[13]

General challenges of data harmonization

The potential for different data sources to have used considerably different methods to measure the same construct often makes harmonizing data challenging, particularly for cognitive data, given the vast range of tests available.[10] The process can be time consuming and resource intensive,[5] even more so when done on a global scale where translation and cultural differences may need to be considered. It should also be noted that harmonization is often specific to the requirements of a certain research question.[5] Further, transformation of raw data to harmonized data can involve some loss or distortion of information, such as when a variable with five response options or a continuous scale is collapsed to a common format variable with three response categories (see Table 1).

Retrospective and prospective approaches to data harmonization

Most of our discussion of harmonization refers to *retrospective data harmonization*, which is applied to pre-existing data where constructs or characteristics of interest were obtained or recorded differently by different sources (for an example see Box 1). An alternative approach is *prospective data harmonization*, which is the implementation of uniform protocols across different studies or research centers before data collection occurs, so that data are collected in a harmonized way. Examples of prospective data harmonization on an

international scale are the 10/66 dementia research group protocols for addressing dementia epidemiology in Latin America, China and India,[14] the Harmonized Cognitive Assessment Protocol designed to enhance comparisons across international sister studies of the U.S. Health and Retirement Study,[15] and the Latin America and the Caribbean Consortium on Dementia (LAC-CD), which aims to facilitate comparisons of dementia between countries with harmonized dementia diagnoses.[16] Similarly prescriptive approaches have been developed for retrospective data harmonization, including a set of guidelines outlining the procedural steps.[5] There have also been attempts to develop systems that facilitate retrospective data harmonization, such as DataSHaPER[17] and the BioSHaRE Project.[18] Full adherence to a harmonized protocol can be compromised by context-dependent requirements, such as the need to replace a cognitive task requiring spelling ability in populations with low rates of literacy.[19] In addition, it has been suggested that the evidence produced by repeated implementation of a protocol across samples may be weaker than evidence from studies using different methodologies.[5]

Harmonizing neuropsychological test data

Neuropsychological test data can be complex to harmonize. There are more than 500 neuropsychological tests[20] and 70 different tests commonly used to assess dementia.[21] A simple method to harmonize such data is to analyze a *common* test or set of tests and treat raw scores as equivalent across sources. In aging research, this has been done for a limited number of widely used measures, like the Mini-Mental State Examination (MMSE).[22] However, this approach excludes potentially useful studies that do not use the same test(s) as others. Also, when evaluating cognition across different ethno-racial populations, it is traditional to base assessments on standardized scores (using appropriate norms) rather than regard raw scores as being equivalent. When different sources use different tests, harmonization requires a statistical approach, of which there are three broad methods:[9,23]

- Standardization
- Latent variable modelling
- Use of multiple imputation.

Standardization—Standardized scores can be used to interpret an individual's test performance. Some test manuals present standardized scores (z-scores with a mean of 0 and a standard deviation [SD] of 1) for different demographic groups, defined by sex and ranges of age and/or education. These demographically adjusted standardized scores are the ones most commonly used when neuropsychologists determine diagnoses of Mild Cognitive Impairment (MCI) or dementia. However, when harmonizing test scores across studies from different ethno-racial populations, such manuals are usually not available. In this situation, regression models have been used to produce demographically adjusted standardized scores, using an appropriate normative sample. In community based longitudinal studies, the baseline sample (excluding those with serious illness or dementia) has been used as the normative sample. Demographically adjusted scores can then be obtained as the standardized residuals in regression models, with demographic variables (usually age, sex and education) as the independent variables and the raw test scores as the dependent variable. Equations used to obtain these standardized scores at baseline can then be applied

to raw scores at later waves, to produce scores which are comparable across waves. This method of harmonizing cognitive tests across cohorts has been done previoulsy[24,25] (see Box 2 for an illustration on harmonizing MCI diagnoses based on standardized cognitive scores).

When research examines the associations of age, sex and education with cognitive performance, demographically adjusted scores would not serve as the appropriate outcome variables. If analyses are confined to a single study, z-scores with means and SDs calculated using the baseline sample (or other appropriate normative sample) could be used. However, such within-study z-scores would not be comparable across studies owing to their different distribution of demographic characteristics.

One solution is to form “demographic category-centred scores” (or C-scores).[9,26] Here, subsamples with the same sex and ranges of age and/or education are selected in each study, and their means and SDs are used to calculate C-scores within each study. For example, subsamples of women aged 70—74 years with 8—13 years of education were used to harmonize cognitive test scores in three Canadian studies.[26] A limitation of this method is the possibility of not obtaining subsamples of sufficient size to reliably estimate the means and SDs required. To overcome this, a modified procedure uses regression models to estimate means and SDs withing each study, conditional on common values of the demographics, chosen to be close to the mean or median values across all studies.[27]

Latent variable modelling—Latent variable modelling assumes the existence of latent factors (or constructs) underlying a set of neuropsychological tests or test items (or more generally, observed indicators). Two modelling methods are the use of Item Response Theory (IRT) based models and Linear Factor Analysis (LFA).

IRT is a framework for understanding the psychometric properties of a test and its items.[28,29] IRT is especially relevant in integrative data analysis,[30,31] because it allows the identification of item biases across studies and demographic groupings, referred to as *differential Item functioning* (DIF), and it uses tests that are both common and noncommon across studies to estimate the underlying construct (for an illustration of linking see Box 3). IRT-based latent variable modelling has been used for harmonizing longitudinal cognitive data.[32] An example of employing LFA in structural equation modelling to obtain latent cognitive factors can be found in Salthouse et al.[33]

Recently, a Moderated Nonlinear Factor Analysis (MNLFA) model has been developed to handle mixed distributions of observed indicators (e.g., binary, ordinal and continuous).[30] This method has the additional advantages of modelling non-linear associations between items and the latent factor, and allowing the model parameters to be moderated by categorical (e.g., sex, study membership) and continuous (e.g., age) covariates simultaneously for testing DIF.

Multiple Imputation—Tests or test items that are not assessed in a particular study can be considered as missing by design, and handled using statistical models like multiple imputation. Values for missing items/tests in one study can be imputed using information

from items/tests overlapping across studies as well as other related variables in the combined data set, but does not require the overlapping items/tests to be in every study. Typically, multiple imputed data sets are generated, and each analysed separately en route to a pooled estimate. Alternatively, values can be averaged across the imputed data sets to generate a full data set. Burns et al.[34] shows how missing MMSE item scores across studies can be imputed and a full data set analyzed.

Harmonizing neuroimaging data

Magnetic resonance imaging (MRI) data can be valuable for understanding and diagnosing neurodegenerative diseases.[35] The cost and time associated with collecting neuroimaging data mean it is often necessary to combine data collected from multiple sites and across diverse populations and experimental conditions to enhance both statistical power and generalisability of findings. This multisite approach to the collection and analysis of neuroimaging data for dementia research includes the Alzheimer's Disease Neuroimaging Initiative (ADNI),[36] ENIGMA[37] and CHARGE[38] consortia. A major challenge for pooling multi-site neuroimaging is the lack of standardisation in both technical aspects (i.e., scanner platforms, image acquisition and processing protocols), as well as differences in sample characteristics (i.e., inclusion/exclusion criteria and sample size).[39]

Methods for the prospective harmonisation of neuroimaging data in the dementias field have been developed by consortia, multi-centre studies and working groups and can include standardisation of: definitions and frameworks (e.g., for imaging of white matter hyperintensities[40]), imaging acquisition protocols (e.g., for vascular dementia[41]) and segmentation procedures (e.g., for hippocampal volume[42]). Data quality control procedures can also be standardised,[43] while containerised software packages can be distributed to ensure consistency in software across sites and time.[44] However, studies have shown that even after careful prospective harmonisation, systematic differences in images and sample characteristics across sites may lead to bias in MRI-derived measures.[45] Retrospective data harmonisation approaches have therefore been developed that allow the pooling of imaging datasets from heterogeneous sources in an unbiased manner.

One of the most widely used methods for retrospective harmonisation of neuroimaging data is the ComBaT approach, a technique originally developed to remove batch effects in genomics data.[46] ComBaT was first extended to the harmonisation of diffusion tensor imaging data,[46] and has recently been applied to the harmonization of structural neuroimaging data in both cross-sectional[39] and longitudinal contexts,[47] as well as functional neuroimaging data.[48] ComBat corrects for site (or scanner) differences via an empirical Bayes algorithm that estimates and removes location (mean) and scale (variance) differences across sites prior to downstream analysis. Clinically-relevant variations are preserved by defining covariates of interest and incorporating their effect on the variance. ComBAT has been applied to the harmonisation of dementia datasets[49] and shown to outperform other site correction techniques.[39]

Other approaches to the harmonisation of multi-site neuroimaging data include Neuroharmony, a supervised machine learning approach that predicts ComBaT correction factors from imaging quality metrics.[50] In a process akin to pediatric growth charts,

normative modelling uses percentiles to chart the variation of an outcome brain measure normed to the variation of a set of clinically-relevant covariates which, in a multisite framework, can include site as a covariate of interest.[51] Recent reviews have identified the potential of this normative approach to address heterogeneity in neuroimaging models of dementia.[52] *Deep learning approaches* have also been developed that are based on generative adversarial networks. These aim to extract a set of imaging features that are maximally informative for an outcome of interest (e.g., Alzheimer's disease) while also being maximally uninformative about the site or scanner where the data originated.[53] These approaches to the retrospective harmonisation of neuroimaging data have their advantages and disadvantages,[54] but each has the potential to provide more powerful and generalisable research into neurodegenerative disorders.

Harmonizing dementia diagnoses

Autopsy-based diagnoses are the gold-standard for dementia and other neurodegenerative diseases. Recent advancements in brain imaging, such as positron emission tomography (PET) scans for amyloid beta and tau, have improved the accuracy of Alzheimer's disease diagnoses. However, this is expensive and not always possible for cohort studies of aging, especially in LMICs. Many research studies therefore rely on clinical diagnoses of dementia, but there are substantial differences in diagnostic procedures (e.g., consensus by an expert panel, assessment tools like the Clinical Dementia Rating scale, the Geriatric Mental State interview) and criteria (e.g., DSM-III-R, DSM-IV, ICD-10) across studies.[1,55] These methodological differences can result in varying estimations of dementia rates.[56] Dementia can be diagnosed from assessments of cognitive performance and instrumental functioning, and algorithms derived from these can be a standardized method of dementia classification across studies (see Prince et al.[57] for an algorithm developed in the 10/66 project). Recently, an IRT based model was used to harmonize dementia classifications in two cross-sectional studies,[58] but its application to a larger number of and more diverse studies has yet to be examined.

Harmonizing behavioral and psychological symptoms of dementia (BPSD) instruments

One challenge for the collection and pooling of BPSD data across studies is the large array of available tools that measure the same or similar constructs. In terms of prospective harmonization of BPSD measures, several consensus guidelines have been developed,[59] with many recommending the Neuropsychiatric Inventory for global assessment of BPSD, as well as more specific measures such as the Geriatric Depression Scale and the Dimension Apathy Scale.[59] Many of these recommended tools are available in multiple languages, including those from LMICs.

Quantitative approaches to the retrospective harmonization of BPSD measures also hold great promise for pooling data that have already been collected or when the adoption of consensus guidelines is not appropriate. Harmonization across BPSD measures often necessitates the identification of common items for linking purposes, and this process for BPSD measures has been detailed recently in a reproducible manner.[60] However, when compared with quantitative harmonization of cognitive measures, the application of these approaches to BPSD instruments has been limited.

Quantitative harmonization has been used to develop common metrics, or crosswalks, which link various measures of neuropsychiatric symptoms,[61] though this approach has not yet been initiated in the dementias field.

Harmonizing electroencephalography (EEG) measures

As a low cost and minimally invasive measure of brain connectivity, EEG represents a viable option for measuring dementia biomarkers in LMICs. To encourage multicentre harmonization of EEG data, the Electrophysiology Professional Interest Area and Global Brain Consortium have endorsed recommendations for EEG measures in clinical trials of Alzheimer's Disease, including for stratification of participants and the monitoring of disease progression.[62] Meanwhile, recent efforts have focused on developing standardised guidelines and best practices for EEG data acquisition, preprocessing and data analysis that can be applied to multicentre EEG studies of brain connectivity more broadly.[63]

Summary

Dementia research is enhanced by bringing together data from multiple sources. However, methodological heterogeneity means that the data typically need to be retrospectively harmonized, sometimes even when prospective approaches to minimize heterogeneity have been implemented. The particular harmonization methods required depend on the data type, and range from a relatively simple choice of comparable items across sources, to the statistical and technology-driven methods needed to harmonize neuropsychological test scores and neuroimaging data, respectively. While often a resource intensive process, harmonization can facilitate data pooling and thereby enhance statistical power. Harmonization can also enable more accurate comparisons, such as comparisons of the prevalence and effects of risk factors for dementia across diverse ethno-regional groups.

References

1. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health. 2022;7(2):e105–e125. [PubMed: 34998485]
2. Ravindranath V, Sundarakumar JS. Changing demography and the challenge of dementia in India. Nat Rev Neurol. 2021;17(12):747–758. [PubMed: 34663985]
3. Prince M, Wimo A, Guerchet M, Ali GC, Wu YT, Prina M. World Alzheimer report 2015: the global impact of dementia. An analysis of prevalence, incidence, cost and trends. London: 2015.
4. Shiekh SI, Cadogan SL, Lin LY, Mathur R, Smeeth L, Warren-Gash C. Ethnic Differences in Dementia Risk: A Systematic Review and Meta-Analysis. J Alzheimers Dis. 2021;80(1):337–355. [PubMed: 33554910]
5. Fortier I, Raina P, Van den Heuvel ER, et al. Maelstrom Research guidelines for rigorous retrospective data harmonization. Int J Epidemiol. 2017;46(1):103–105. [PubMed: 27272186]
6. Hussong AM, Curran PJ, Bauer DJ. Integrative data analysis in clinical psychology research. Annu Rev Clin Psychol. 2013;9:61–89. [PubMed: 23394226]
7. Shishegar R, Cox T, Rolls D, et al. Using imputation to provide harmonized longitudinal measures of cognition across AIBL and ADNI. Sci Rep. 2021;11(1):23788. [PubMed: 34893624]
8. Lipnicki DM, Makkar SR, Crawford JD, et al. Determinants of cognitive performance and decline in 20 diverse ethno-regional groups: A COSMIC collaboration cohort study. PLoS Med. 2019;16(7):e1002853. [PubMed: 31335910]

9. Griffith L, van den Heuvel E, Fortier I, et al. In: Harmonization of Cognitive Measures in Individual Participant Data and Aggregate Data Meta-Analysis. Rockville (MD)2013.
10. Briceno EM, Gross AL, Giordani BJ, et al. Pre-Statistical Considerations for Harmonization of Cognitive Instruments: Harmonization of ARIC, CARDIA, CHS, FHS, MESA, and NOMAS. J Alzheimers Dis. 2021;83(4):1803–1813. [PubMed: 34459397]
11. Vonk JMJ, Gross AL, Zammit AR, et al. Cross-national harmonization of cognitive measures across HRS HCAP (USA) and LASI-DAD (India). PLoS One. 2022;17(2):e0264166. [PubMed: 35213581]
12. Elsevier. https://www.elsevier.com/authors/tools-and-resources/research-data.
13. NIH. https://sharing.nih.gov/data-management-and-sharing-policy/about-data-management-sharing-policy/data-management-and-sharing-policy-overview#after.
14. Prince M, Ferri CP, Acosta D, et al. The protocols for the 10/66 dementia research group population-based research programme. BMC Public Health. 2007;7:165. [PubMed: 17659078]
15. Langa KM, Ryan LH, McCammon RJ, et al. The Health and Retirement Study Harmonized Cognitive Assessment Protocol Project: Study Design and Methods. Neuroepidemiology. 2020;54(1):64–74. [PubMed: 31563909]
16. Ibanez A, Parra MA, Butler C, Latin A, the Caribbean Consortium on D. The Latin America and the Caribbean Consortium on Dementia (LAC-CD): From Networking to Research to Implementation Science. J Alzheimers Dis. 2021;82(s1):S379–S394. [PubMed: 33492297]
17. Fortier I, Burton PR, Robson PJ, et al. Quality, quantity and harmony: the DataSHaPER approach to integrating data across bioclinical studies. Int J Epidemiol. 2010;39(5):1383–1393. [PubMed: 20813861]
18. Doiron D, Burton P, Marcon Y, et al. Data harmonization and federated analysis of population-based studies: the BioSHaRE project. Emerg Themes Epidemiol. 2013;10(1):12. [PubMed: 24257327]
19. Torres JM, Glymour MM. Future Directions for the HRS Harmonized Cognitive Assessment Protocol. Forum Health Econ Policy. 2022.
20. Lezak MD, Howieson DB, LD W, Hannay HJ, Fischer JS. Neuropsychological assessment. 4th ed. New York, NY: Oxford University Press; 2004.
21. Maruta C, Guerreiro M, de Mendonca A, Hort J, Scheltens P. The use of neuropsychological tests across Europe: the need for a consensus in the use of assessment tools for dementia. Eur J Neurol. 2011;18(2):279–285. [PubMed: 20597968]
22. Gross AL, Sherva R, Mukherjee S, et al. Calibrating longitudinal cognition in Alzheimer's disease across diverse test batteries and datasets. Neuroepidemiology. 2014;43(3-4):194–205. [PubMed: 25402421]
23. Griffith LE, van den Heuvel E, Fortier I, et al. Statistical approaches to harmonize data on cognitive measures in systematic reviews are rarely reported. J Clin Epidemiol. 2015;68(2):154–162. [PubMed: 25497980]
24. Abner EL, Schmitt FA, Nelson PT, et al. The Statistical Modeling of Aging and Risk of Transition Project: Data Collection and Harmonization Across 11 Longitudinal Cohort Studies of Aging, Cognition, and Dementia. Obs Stud. 2015;1(2015):56–73. [PubMed: 25984574]
25. Sachdev PS, Lipnicki DM, Kochan NA, et al. The Prevalence of Mild Cognitive Impairment in Diverse Geographical and Ethnocultural Regions: The COSMIC Collaboration. PLoS One. 2015;10(11):e0142388. [PubMed: 26539987]
26. Griffith LE, van den Heuvel E, Raina P, et al. Comparison of Standardization Methods for the Harmonization of Phenotype Data: An Application to Cognitive Measures. Am J Epidemiol. 2016;184(10):770–778. [PubMed: 27769990]
27. Lipnicki DM, Crawford JD, Dutta R, et al. Age-related cognitive decline and associations with sex, education and apolipoprotein E genotype across ethnocultural groups and geographic regions: a collaborative cohort study. PLoS Med. 2017;14(3):e1002261. [PubMed: 28323832]
28. Bock RD, Gibbons RD. Item Response Theory. John Wiley & Sons Inc; 2021.
29. Embretson SE, Reise SP. Item response theory for psychologists. Mahwah, NJ: Erlbaum; 2000.

30. Bauer DJ, Hussong AM. Psychometric approaches for developing commensurate measures across independent studies: traditional and new models. Psychol Methods. 2009;14(2):101–125. [PubMed: 19485624]
31. Curran PJ, Hussong AM, Cai L, et al. Pooling data from multiple longitudinal studies: the role of item response theory in integrative data analysis. Dev Psychol. 2008;44(2):365–380. [PubMed: 18331129]
32. Gross AL, Mungas DM, Crane PK, et al. Effects of education and race on cognitive decline: An integrative study of generalizability versus study-specific results. Psychol Aging. 2015;30(4):863–880. [PubMed: 26523693]
33. Salthouse TA. Localizing age-related individual differences in a hierarchical structure. Intelligence. 2004;32(6).
34. Burns RA, Butterworth P, Kiely KM, et al. Multiple imputation was an efficient method for harmonizing the Mini-Mental State Examination with missing item-level data. J Clin Epidemiol. 2011;64(7):787–793. [PubMed: 21292440]
35. Jovicich J, Barkhof F, Babiloni C, et al. Harmonization of neuroimaging biomarkers for neurodegenerative diseases: A survey in the imaging community of perceived barriers and suggested actions. Alzheimers Dement (Amst). 2019;11:69–73. [PubMed: 31673595]
36. Alzheimer's Disease Neuroimaging Initiative. https://adni.loni.usc.edu/. Published 2017.
37. ENIGMA: Enhancing Neuro Imaging Genetics Through Meta Analysis. https://enigma.ini.usc.edu/.
38. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. https://www.hgsc.bcm.edu/human/charge-consortium. Published 2022.
39. Fortin JP, Cullen N, Sheline YI, et al. Harmonization of cortical thickness measurements across scanners and sites. Neuroimage. 2018;167:104–120. [PubMed: 29155184]
40. Roseborough AD, Saad L, Goodman M, Cipriano LE, Hachinski VC, Whitehead SN. White matter hyperintensities and longitudinal cognitive decline in cognitively normal populations and across diagnostic categories: A meta-analysis, systematic review, and recommendations for future study harmonization. Alzheimers Dement. 2022.
41. HARNESS HARmoNising Brain Imaging MEthodS for VaScular Contributions to Neurodegeneration. https://harness-neuroimaging.org/.
42. Frisoni GB, Jack CR Jr., Bocchetta M, et al. The EADC-ADNI Harmonized Protocol for manual hippocampal segmentation on magnetic resonance: evidence of validity. Alzheimers Dement. 2015;11(2):111–125. [PubMed: 25267715]
43. Samann PG, Iglesias JE, Gutman B, et al. FreeSurfer-based segmentation of hippocampal subfields: A review of methods and applications, with a novel quality control procedure for ENIGMA studies and other collaborative efforts. Hum Brain Mapp. 2022;43(1):207–233. [PubMed: 33368865]
44. Waller L, Erk S, Pozzi E, et al. ENIGMA HALFpipe: Interactive, reproducible, and efficient analysis for resting-state and task-based fMRI data. Hum Brain Mapp. 2022;43(9):2727–2742. [PubMed: 35305030]
45. Guo C, Niu K, Luo Y, et al. Intra-Scanner and Inter-Scanner Reproducibility of Automatic White Matter Hyperintensities Quantification. Front Neurosci. 2019;13:679. [PubMed: 31354406]
46. Fortin JP, Parker D, Tunc B, et al. Harmonization of multi-site diffusion tensor imaging data. Neuroimage. 2017;161:149–170. [PubMed: 28826946]
47. Beer JC, Tustison NJ, Cook PA, et al. Longitudinal ComBat: A method for harmonizing longitudinal multi-scanner imaging data. Neuroimage. 2020;220:117129. [PubMed: 32640273]
48. Yu M, Linn KA, Cook PA, et al. Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. Hum Brain Mapp. 2018;39(11):4213–4227. [PubMed: 29962049]
49. Eshaghzadeh Torbati M, Minhas DS, Ahmad G, et al. A multi-scanner neuroimaging data harmonization using RAVEL and ComBat. Neuroimage. 2021;245:118703. [PubMed: 34736996]
50. Garcia-Dias R, Scarpazza C, Baecker L, et al. Neuroharmony: A new tool for harmonizing volumetric MRI data from unseen scanners. Neuroimage. 2020;220:117127. [PubMed: 32634595]

51. Marquand AF, Kia SM, Zabihi M, Wolfers T, Buitelaar JK, Beckmann CF. Conceptualizing mental disorders as deviations from normative functioning. Mol Psychiatry. 2019;24(10):1415–1424. [PubMed: 31201374]
52. Verdi S, Marquand AF, Schott JM, Cole JH. Beyond the average patient: how neuroimaging models can address heterogeneity in dementia. Brain. 2021;144(10):2946–2953. [PubMed: 33892488]
53. Moyer D, Ver Steeg G, Tax CMW, Thompson PM. Scanner invariant representations for diffusion MRI harmonization. Magn Reson Med. 2020;84(4):2174–2189. [PubMed: 32250475]
54. Bayer JMM, Thompson PM, Ching CRK, et al. Site effects how-to & when: an overview of retrospective techniques to accommodate site effects in multi-site neuroimaging analyses. PsyArXiv Preprints. 2022.
55. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: a systematic review and metaanalysis. Alzheimers Dement. 2013;9(1):63–75 e62. [PubMed: 23305823]
56. Erkinjuntti T, Ostbye T, Steenhuis R, Hachinski V. The effect of different diagnostic criteria on the prevalence of dementia. N Engl J Med. 1997;337(23):1667–1674. [PubMed: 9385127]
57. Prince MJ, de Rodriguez JL, Noriega L, et al. The 10/66 Dementia Research Group's fully operationalised DSM-IV dementia computerized diagnostic algorithm, compared with the 10/66 dementia algorithm and a clinician diagnosis: a population validation study. BMC Public Health. 2008;8:219. [PubMed: 18577205]
58. GBD Dementia Collaborators. Use of multidimensional item response theory methods for dementia prevalence prediction: an example using the Health and Retirement Survey and the Aging, Demographics, and Memory Study. BMC Med Inform Decis Mak. 2021;21(1):241. [PubMed: 34380485]
59. Costa A, Bak T, Caffarra P, et al. The need for harmonisation and innovation of neuropsychological assessment in neurodegenerative dementias in Europe: consensus document of the Joint Program for Neurodegenerative Diseases Working Group. Alzheimers Res Ther. 2017;9(1):27. [PubMed: 28412978]
60. Chen D, Jutkowitz E, Iosepovici SL, Lin JC, Gross AL. Pre-statistical harmonization of behavioral instruments across eight surveys and trials. BMC Med Res Methodol. 2021;21(1):227. [PubMed: 34689753]
61. Batterham PJ, Sunderland M, Slade T, Calear AL, Carragher N. Assessing distress in the community: psychometric properties and crosswalk comparison of eight measures of psychological distress. Psychol Med. 2018;48(8):1316–1324. [PubMed: 28967345]
62. Babiloni C, Arakaki X, Azami H, et al. Measures of resting state EEG rhythms for clinical trials in Alzheimer's disease: Recommendations of an expert panel. Alzheimers Dement. 2021;17(9):1528–1553. [PubMed: 33860614]
63. Prado P, Birba A, Cruzat J, et al. Dementia ConnEEGtome: Towards multicentric harmonization of EEG connectivity in neurodegeneration. Int J Psychophysiol. 2022;172:24–38. [PubMed: 34968581]
64. Jessen F, Amariglio RE, van Boxtel M, et al. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. Alzheimers Dement. 2014;10(6):844–852. [PubMed: 24798886]
65. Rohr S, Pabst A, Riedel-Heller SG, et al. Estimating prevalence of subjective cognitive decline in and across international cohort studies of aging: a COSMIC study. Alzheimers Res Ther. 2020;12(1):167. [PubMed: 33339532]
66. Ward A, Arrighi HM, Michels S, Cedarbaum JM. Mild cognitive impairment: disparity of incidence and prevalence estimates. Alzheimers Dement. 2012;8(1):14–21. [PubMed: 22265588]
67. Mukadam N, Sommerlad A, Huntley J, Livingston G. Population attributable fractions for risk factors for dementia in low-income and middle-income countries: an analysis using cross-sectional survey data. Lancet Glob Health. 2019;7(5):e596–e603. [PubMed: 31000129]

Box 1

Qualitative harmonization of self-experienced decline in cognitive capacity

Subjective cognitive decline is self-experienced decline in cognitive ability from a normal level in the absence of objective impairment, and may be the first sign of Alzheimer's disease.[64] A recent collaborative research project aimed to estimate the prevalence of subjective cognitive decline (SCD) in and across international cohort studies of aging.[65] Each study contributing data to the project asked their participants different sets and numbers of questions relevant to determining self-experienced decline in cognitive capacity, requiring the data to be harmonized for more accurate comparison and pooling.

The project used two approaches to harmonizing self-experienced decline in cognitive capacity: qualitative and quantitative.

Qualitative: Two authors independently compared all items assessing self-experienced decline in cognitive capacity across the studies, and identified one common item from each that broadly addressed problems or difficulties with memory. The original data for these items were transformed to a binary variable indicating the presence or absence of self-experienced decline in cognitive capacity, with any indication of decline in the original responses categorized as “presence”.

Study	Item selected for qualitative harmonization	Original coding
Active Ageing	Do you feel you have more problems with your memory than most?	1 = yes, 2 = no
CFAS	Have you ever had any difficulty with your memory? If yes, is that a problem for you?	0 = no, 1 = yes, moderate, 2 = yes, severe
EAS	Compared with one year ago, do you have trouble remembering things more often, less often or about the same?	1 = more often, 2 = less often, 3 = about the same
SLASII	Overall, how would you rate your memory or other mental abilities as compared to earlier period of your life (more than one year ago)?	1 = much better, 2 = a bit better, 3 = a bit worse, 5 = much worse

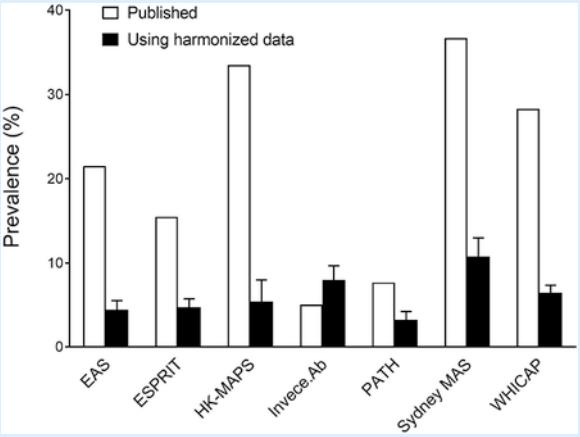
Note: only 4 of the 16 studies included in the project are shown.

Box 2

Using harmonized cognitive impairment and dementia data for international comparisons

When researching MCI and dementia on a global scale, a great benefit of using harmonized data is the capacity for more accurate comparisons across different ethno-regions.

- More accurate comparisons of prevalence and incidence of dementia and related conditions. While the high variation in reported rates of mild cognitive impairment (MCI) across different countries are partially explained by differences in location and demographics, there is a significant contribution from differences in definition and methodology.[66] These differences can be reduced by harmonizing cognitive test, functional and subjective cognitive complaint data and applying a uniform approach to classifying MCI. This approach has yielded much more similar rates of MCI than previously reported.[25] The figure shows the prevalence of MCI previously reported for seven cohort studies representing five different countries alongside more uniform rates produced using harmonized data.[25]



- Better understanding of risk factors for dementia and related conditions as universal, or as differing between races/ethnicities and regions, including strength of association between risk factors and outcome. Not only are there ethno-regional differences in the prevalence of risk factors for dementia, such as more diabetes and hypertension in developing countries like India,[2] but analysis of harmonized data on an international scale suggests that the strength of association between particular risk factors and CI and dementia can also differ.[8] A risk factor's prevalence and strength of association with dementia determine the proportion of dementia in a population that can be attributed to the risk factor. This proportion was able to be estimated for various dementia risk factors and more accurately compared across eight countries where identical 10/66 protocols had been used.[67]

Box 3

Quantitative harmonization of self-experienced decline in cognitive capacity

The quantitative harmonization approach used both common and unique items to model the latent construct of self-experienced cognitive decline that is equivalent in meaning and metrics across studies.[64] The common item serves as an anchor to link the unique items, for example, item 2 in Study 2 can be linked to item 6 in Study 3 via the common item. The 2-Parameter Logistic (2-PL) Item Response Theory (IRT) model[30,31] was used to evaluate measurement equivalence of the items (item difficulty and item discrimination) across studies, and based on the model, latent scores for each participant were estimated.

	Study			
Items	**1**	**2**	**3**	**4**
1. Common item (see Box 1)	■	■	■	■
2. Have you tended to forget things recently?		■		
3. Difficulty remembering names/things of close people		■		■
4. Difficulty remembering where you kept/put things		■		
5. More effort to remember things than used to?		■		
6. In the past year, how often did you have trouble remembering things?			■	
7. Memory worse than 10 years ago			■	

Key Points:

- Data from multiple sources often represent heterogeneous methodology that includes different assessment instruments and classification criteria.
- Harmonization is the process by which data for similar measures or constructs from different sources are made more comparable.
- Harmonization enables data from multiple sources to be analyzed simultaneously, with techniques such as mega-analysis and individual participant data meta-analyses.
- Statistical harmonization is needed for neuropsychological test data, with methods including standardization, latent variable modelling, and the use of multiple imputation.
- The most popular approach for harmonizing neuroimaging data is ComBaT, with other applications to dementia research including normative modelling and machine learning approaches to statistical harmonization.

Synopsis:

Understanding dementia and cognitive impairment is a global effort needing data from multiple sources across diverse ethno-regional groups. Methodological heterogeneity means that these data often require harmonization to make them comparable before analysis. We discuss the benefits and challenges of harmonization, both retrospective and prospective, broadly and with a focus on data types that require particular sorts of approaches, including neuropsychological test scores and neuroimaging data. Throughout our discussion we illustrate general principles and give examples of specific approaches in the context of contemporary research in dementia and cognitive impairment from around the world.

Clinical care points

- With the increasing digitalization of medical care, data from diverse sources must be harmonized for efficient clinical care and facilitation of clinical research.
- Barriers and facilitators of harmonization should be identified at the national and international levels, so that global clinical research and practice can inform clinical care and prevention of dementia in all jurisdictions.
- Policies and frameworks should be put into place to facilitate harmonization of clinical and research data at both national and international levels.

Table 1.

Example of self-rated health scale harmonization

Study	Coding of original response options to Very good = 1, Good = 2, Poor = 3
Bambui	Very Good, Good = 1; Reasonable = 2; Fair = 3
CFAS	Excellent = 1; Good = 2; Fair, Poor = 3
EAS	Excellent, Very Good = 1; Good = 2; Fair, Poor = 3
HK-MAPS	Cumulative Illness Rating Scale sum of various organ system severity ratings: 0,1=1; 2-4=2; 5-13=3
Invece.Ab	Visual analogue scale: 0-6 = 1; 7-8 = 2; 9-10 = 1
KLOSCAD	Excellent, Good = 1; Fair = 2; Poor = 3
LEILA75+	Very Good/Excellent = 1; Good = 1; Fair = 2; Poor = 3; Very Poor = 3
MoVIES	Excellent = 1; Good = 2; Fair, Poor = 3
PATH	Excellent, Very Good = 1; Good = 2; Fair, Poor = 3
SALSA	Excellent, Very Good = 1; Good = 2; Fair, Poor = 3
SGS	Very Good = 1; Good = 2; Fair, Poor = 3
SLASI	Excellent, Very Good = 1; Good = 2; Fair, Poor = 3
Sydney MAS	Excellent, Very Good = 1; Good = 2; Fair, Poor = 3

Note. Taken from S9 Table in Lipnicki et al. showing how the original self-reported health data from 13 international cohort studies of aging were harmonized to a 3-category variable representing response options very good, good, and poor. *(Adapted from Lipnicki DM, Makkar SR, Crawford JD, et al. Determinants of cognitive performance and decline in 20 diverse ethno-regional groups: A COSMIC collaboration cohort study. PLoS Med. 2019;16(7):e1002853; under CC BY 4.0)*