

Future Directions for the HRS Harmonized Cognitive Assessment Protocol

Jacqueline M. Torres, M. Maria Glymour*

Department of Epidemiology & Biostatistics, UC San Francisco, San Francisco, CA

*Corresponding Author: Maria.Glymour@ucsf.edu

Funding: JMT and MMG report funding from the National Institutes of Health for the present study. JMT reports funding from the National Institutes of Health outside of the present study. MMG reports funding from the National Institutes of Health and the Robert Wood Johnson Foundation outside of the present study.

Acknowledgments: We thank Kaitlin Swinnerton for expert analyses of HRS data as we were preparing this document.

Conflicts of Interest: None to declare
Date: 18 October, 2021

Manuscript Word Count: 6062

ABSTRACT

In the absence of effective pharmacological treatment to halt or reverse the course of Alzheimer's disease and related dementias (ADRDs), population-level research on the modifiable determinants of dementia risk and outcomes for those living with ADRD is critical. The Harmonized Cognitive Assessment Protocol (HCAP), fielded in 2016 as part of the U.S. Health and Retirement Study (HRS) and multiple international counterparts, has the potential to play an important role in such efforts. The stated goals of the HCAP are to improve both our ability to understand the determinants, prevalence, costs, and consequences of cognitive impairment and dementia in the U.S. and to support cross-national comparisons. The first wave of HCAP demonstrated feasibility and value of the more detailed cognitive assessments. To achieve its full potential, we recommend that the HCAP: **1)** increase the representation of racial/ethnic minority participants disproportionately affected by ADRDs, **2)** administer the HCAP repeatedly and with greater frequency to facilitate longitudinal analyses, cross-national harmonization, and precise characterization of practice and period effects, **3)** field a clinical dementia assessment for at least a subset of HCAP participants, **4)** phase-in a baseline assessment for people under age 60 in order to facilitate research on the earlier stages of cognitive impairment and midlife dementia prevention, **5)** consider dropping lower performing items from the HCAP to increase feasibility of scaling the assessment to more participants at more frequent intervals, **6)** enhance the documentation, multi-lingual options, training activities, and randomized sub-studies of language effects or other tools to promote the use of the HCAP and cross-national comparisons, **7)** incorporate blood-based AD biomarkers, and **8)** improve the capacity of the HCAP to understand the consequences of dementia, including for caregivers and for policy-relevant outcomes such as use of home and community-based services. We believe that the capacity of the HCAP to achieve its stated goals will be greatly enhanced by considering these changes and additions.

INTRODUCTION

As of 2017, an estimated 55 million people were living with Alzheimer's Disease and Alzheimer's Disease Related Dementias (AD/ADRD) worldwide; approximately 10 million individuals are diagnosed with AD/ADRDs each year.¹ In the absence of effective pharmacological treatment to halt or reverse the course of AD/ADRDs,^{2,3} population-level research on the modifiable determinants of ADRD risk and outcomes for those living with the disease is critical.⁴ The U.S. Health and Retirement Study (HRS) as well as its global partner studies have been foundational to this effort to improve population-level knowledge on AD/ADRD prevention and care, with hundreds of manuscripts indexed in PubMed in the past 15 years. As part of the larger HRS, the extensive Harmonized Cognitive Assessment Protocol (HCAP)⁵ was fielded in 2016 to a subset of respondents 65 years and older, supplementing the brief cognitive assessments historically administered to the core study.

The HCAP has the potential to play a critical role in future dementia research, both in the U.S. and globally. In the U.S., the HCAP promises to enable updated classifications of mild cognitive impairment and dementia that may be translated to the broader HRS sample. Current HRS studies rely on dementia classifications generated from the Aging, Demographics and Memory Study (ADAMS),⁶ which similarly fielded a comprehensive cognitive assessment to a subset of HRS participants. However, the ADAMs study was fielded nearly two decades ago, includes less than a third of the participants who were included in the HCAP study, and – most importantly – has too few respondents from historically marginalized backgrounds to produce reliable estimates for many racial/ethnic subgroups.⁷ This final point is critical given both the disproportionate burden of ADRDs faced by Black, Latinx, and Indigenous individuals in the U.S.⁸ and the growing diversity of the U.S. older adult population.⁹

Globally, the HCAP has been replicated across multiple partner studies – in Mexico,¹⁰ China¹¹, India¹², and England¹³ – and there are plans for implementation in other settings. Despite tremendous investment in these global aging studies to support correspondence with HRS, true harmonization of the cognitive measures historically included in the core surveys has been challenging due to the substantial cross-study differences in the cognitive assessment batteries fielded to the core samples. HCAP is thus a valuable initiative to improve the quality of cognitive assessment, permit some degree of domain specificity, and support better harmonization. Despite this potential, researcher use of the HCAP data has been somewhat slow initially, with relatively few citations using the HCAP in empirical studies indexed in PubMed to date.

In light of the urgent need for population-level ADRD research, we have several recommendations to promote the impact of HCAP for its stated goals and provide the most relevant information for research on cognitive aging:

1. Prioritize the addition of racial/ethnic minority participants in the HRS HCAP sample, aiming to maximize statistical power to evaluate between-group differences (i.e., precision of estimates of inequality) and within-group predictors (i.e., precision of estimates of determinants of cognitive aging within racial/ethnic minority groups).
2. Administer HCAP as frequently as possible; for small, representative subsamples incorporate HCAP into the main HRS questionnaire to enhance longitudinal analyses and support precise characterization of practice and period effects.
3. Add a clinical dementia assessment for at least a small sample of HCAP participants to establish sensitivity and specificity of the HCAP predictions.

4. Phase in a baseline HCAP assessment for people under age 60.
5. Drop low-performing items from HCAP to increase the feasibility of scaling the HCAP sample size and increasing the frequency of its administration.
6. Enhance documentation, multi-lingual options, training activities, and support randomized sub-studies of language effects, and other tools to promote cross-national comparisons.
7. Incorporate blood-based AD biomarkers into the HCAP and international partner studies as soon as feasible
8. Improve the capacity of the HCAP and/or the broader HRS for capturing the *consequences* of dementia.

The motivation for each of these recommendations is described in more detail below.

1. Prioritize recruitment of racial/ethnic minorities

Our highest priority recommendation is oriented to improving the utility of the HCAP for understanding determinants of AD/ADRD and AD/ADRD disparities in the US. To accomplish this, it is critical that the HCAP increase the representation of Black, Latinx, and other racial/ethnic minority groups disproportionately impacted by AD/ADRD^{8,14} and inequities in dementia diagnosis and care.^{15,16} Racial/ethnic minorities represent a growing share of the US older adult population; by 2040 an expected 34% of older adults will be racial/ethnic minorities, up from 19% in 2008 and 23% in 2018.⁹ The combination of elevated incidence and growing representation as a percent of population implies that from a population health perspective, older people of color and their families will comprise an ever larger fraction of people affected by AD/ADRD.

While the HRS has historically oversampled Black and Latinx adults relative to their share of the community-dwelling population 50 years and older, the 2016 HCAP sample remains 71% non-Latinx white. The absolute numbers of Black and Latinx respondents are very small – only 527 Black respondents and 363 Latinx respondents with non-missing cognitive scores are included. While classifications of mild cognitive impairment (MCI) and dementia are not yet available for the 2016 HCAP, we estimate that approximately 27% of those in the HCAP sample would have been classified as having mild cognitive impairment or dementia (approximately 21% with MCI and 6% with dementia), based on applying a well-established algorithm¹⁷ to the core HRS cognitive measures. This means that the HCAP sample may only include approximately 142 Black respondents and 98 Latinx individuals with probable MCI or dementia, and 32 Black and 22 Latinx respondents with probable dementia only. This compares to 642 NH white respondents who we estimate would have MCI or dementia (143 with probable dementia only). While these numbers are estimates based on the core HRS assessments and may change slightly if using the full HCAP battery, they illustrate that even with oversampling relative to the population, the absolute numbers of Black and Latinx adults with the primary outcome of interest are very small.

The very small numbers of Black and Latinx older adults in the analytic sample of the 2016 HCAP have several important consequences. First, many within-HCAP analyses focused on AD/ADRD disparities (i.e. comparing outcomes for Black or Latinx respondents to those for non-Latinx white respondents) and/or drivers of AD/ADRDs *among* Black or Latinx respondents will be underpowered and imprecise. Second, estimates that rely on a crosswalk between the HRS and the HCAP (i.e. to generate probabilities of MCI/dementia based on shared information across the two studies) could be less reliable for Black and Latinx participants due to the small

sample size. The ADAMS similarly carried out detailed neuropsychological exams with a subset of respondents from the core HRS, and authors have used the diagnoses available in ADAMs to develop thresholds or probabilities for MCI and dementia classification in the core HRS sample.^{7,17,18} In one such study, Wu and authors⁷ reported that the very small number of Latinx participants in the ADAMs (n=84) meant that stable estimates could not be derived for this subgroup in the HRS. Third, research on the social and behavioral drivers of ADRD overall is rendered less efficient because the range of social and behavioral profiles is unnecessarily restricted. Racial/ethnic diversity enhances the variability of risk factors that could be evaluated as determinants of ADRD, and will therefore improve our ability to identify modifiable social or behavioral targets for prevention or improving the well-being of people living with ADRD.

We recommend that both the proportion and the absolute number of Black and Latinx participants be increased substantially, ideally to account for a full two-thirds of the HCAP sample. This may mean increasing the overall HCAP sample size or reducing the number of non-Latinx white participants, as determined by budgetary constraints. Of course, increasing the overall sample size would be far preferable if financially feasible. This shift would mean that the proportion of racial/ethnic subgroups in the HCAP exceed their proportion of community-dwelling older adults more broadly. However, nationally representative estimates could still be easily achieved with crosswalks to the core HRS sample and sample weighting.

2. Prioritize longitudinal assessments

Administering the HCAP repeatedly and at frequent intervals for the same subset of respondents in the HRS and its international counterparts is essential for assessments and comparisons of the rate of within-person cognitive decline for all respondents. This

recommendation is partially driven by the fact that cognitive performance assessments from a single time point – as well as indicators of mild cognitive impairment and dementia based on these single time point assessments -- are often strongly influenced by factors such as length and/or quality of education and test-taking abilities that do not necessarily predict the rate of cognitive decline^{19,20} or correspond with AD/ADRD risk. In addition, measures of longitudinal cognitive change better predict brain aging than cognitive performance assessments at a single time point.^{21,22} Repeated assessments among HCAP participants and their global counterparts with therefore allow for better science on the determinants of brain aging.

This recommendation is also made in service of improving the capacity for cross-national harmonization of HCAP-based outcomes. One of the central goals of the HCAP is to facilitate cross-national comparisons across the HRS and international partner studies in the interest of making “comparable classifications to discriminate normal, cognitive impairment, and dementia status”.⁵ We suggest that this goal will be nearly impossible to achieve without the completion of repeated HCAP assessments within all relevant studies.

Despite significant investments in creating comparable measurement protocols across studies, there remain important cross-study differences that make interpretation of comparisons of single time point assessments ambiguous. These differences partly stem from the fact that translated versions of the HCAP within or across country settings may have yielded more or less difficult tasks. For example, when the words included in the word lists administered as part of verbal memory tasks are translated verbatim across language(s) spoken in international partner countries, the comparability is not maintained. Translated words could be shorter or longer or more or less common across contexts, thereby altering the level of difficulty regardless of a participant’s cognitive status. Variation in the level of difficulty across cognitive tasks based on

their translation and their familiarity across study contexts may contribute to substantial differences in baseline scores for reasons that have nothing to do with cognitive aging.

For many components of the HCAP, international partner studies did alter their tasks in their counterpart cognitive batteries in order to fit their specific contexts.^{10,23} For example, instead of the original MMSE administered in the HCAP, studies administered modified, contextually sensitive versions of the MMSE. In the original MMSE, participants are asked to spell “World” backwards. However, the Hindi Mental State Examination,²⁴ included in the LASI-DAD study, accounts for the lower rates of literacy among the current birth cohorts of Indian older adults by asking participants to instead list the days of the week in reverse order from Sunday to Monday. In another example, as part of the Telephone Interview for Cognitive Status (TICs), respondents in the HRS HCAP were asked to correctly name “cactus” in response to being asked “what is that prickly plant that grows in the desert?”. In the LASI-DAD, respondents were instead asked to correctly name a coconut; in the Mex-Cog, respondents were asked to correctly name a pencil. These and many other context-specific modifications are critical. However, with a single wave of data for each study it would be challenging to convincingly attribute mean cross-study differences in scores on these items to true underlying cognitive differences rather than measurement differences.

Challenges to comparing single time-point cognitive performance scores may also result from the fact that current older adult cohorts within and across countries have had dramatically different educational and occupational opportunities, resulting in systematic differences in exposure to test-taking, mental math practice, or other activities that might prepare them for the format of an extensive cognitive assessment like the HCAP. These differences may prepare participants to achieve better scores across and even within settings for reasons that may be

largely or entirely driven by earlier-life socio-economic conditions. Once again, it is unclear whether these same factors that influence cognitive test-taking performance will also be influential for cognitive aging. This means that it may be impossible to tell whether cross-national differences in baseline cognitive scores are driven by true differences in cognitive aging vs. differences in protocols across studies and/or cross-country contextual differences (e.g. in average levels of formal education).

Item response theory and other modern psychometric tools to estimate latent variables make the most efficient use of available data and can help solve the problem of meaningfully comparing across versions of a scale.^{25,26} These methods are not a panacea, however. These approaches identify off of either assuming that at least some items are truly comparable between versions (anchor items) or making assumptions about the distribution of the latent variable (e.g., that it is the same across settings). The latter approach, based on distributional assumptions, is circular if the goal is to compare prevalence of cognitive impairment across settings. The former approach, based on identifying anchor items, is fragile if there are only a small number of anchoring items (fragile in the sense that small violations of the assumptions may have large implications for the overall findings). The assumption that an item is truly equivalent across languages (and cultures) is not easily testable. Within-person change, in contrast, has intrinsic relevance that is more plausibly comparable across settings.

With cross-sectional data only, the best insights for cross-national comparisons are more likely to arise from contrasting the age-slope (i.e., between-person comparisons of people of different ages) between different settings than directly comparing cognitive scores. This is similar conceptually to comparing older adults in each setting to younger individuals in the same setting and using the deviation from within-setting healthy norms as the cognitive assessment,

instead of the raw cognitive score. For this, we need to field the HCAP in younger adults who are unlikely to have had substantial age-related cognitive deterioration. A final approach would involve randomly assigning bilingual individuals to take the test in different languages. This would allow us to estimate an average version effect.

The above methods are ad hoc and each introduce their own new challenges. Our main hope for cross-national harmonization going forward rests on the ability to compare within-person slopes in cognitive performance scores derived from repeated assessments using the same protocol for the same set of individuals. Longitudinal analyses allow within-person comparison which will substantially control for stable national differences that affect test-taking skills but have no effect on AD/ADRD.

With a small number of repeated assessments, the HCAP may also be able to help disentangle “practice” or retest effects from cognitive aging: practice effects are a perennial methodological challenge for longitudinal cognitive aging research.^{27,28} Such disentangling would be enabled by randomly selecting some individuals for more frequent repeated measures and others for delayed introduction to the HCAP protocol. In particular, we suggest that if the HCAP enrolls a new group of participants at follow-up, average practice effects could be calculated as the difference in scores for respondents who are completing the HCAP for the second time and respondents matched on age and education who are completing the HCAP for the first time. This would be an improvement over the typical way of quantifying the magnitude of practice effects, which entails administering the same cognitive assessment to a group of respondents within a period of time that is so short that changes in performance would most plausibly be due to practice effects.²⁹ This typical approach generates practice effect estimates

that may be of limited relevance for the typical cognitive aging cohort study, which usually collects repeated assessments two or three years apart.

Finally, the COVID-19 pandemic has illustrated the plausibility of major period effects in cognitive assessments. As long-time users of HRS, we suspect there have been substantial period effects in prior waves for one reason or another (e.g. protocol or interview changes or happenstance of timing). Period effects compromise the ability to evaluate cognitive change or trends in the population. If a small subsample of the HCAP participants receives repeated assessments more frequently than the main repeat iteration cycle, it offers important advantages for modeling and accounting for period effects when evaluating rates of age-related change or population trends.

3. Add a clinical dementia assessment

While the HCAP includes a rigorous and extensive set of cognitive performance assessments, one of the challenges to generating dementia classifications and to understanding the utility of any given component of the HCAP battery is the lack of an available gold standard. While clinical dementia assessments – included in the ADAMs study as well as in select international HCAP studies – are not a gold standard, they are likely the best standard available for benchmarking against. Clinical assessments are therefore important for evaluating the sensitivity and specificity of cognitive assessments, including within and across global studies. Clinical dementia assessments could be used alongside other imperfect benchmarks such as age or mortality to evaluate the performance of specific tasks within the HCAP battery as well as the accuracy of dementia classifications. We therefore recommend that the HCAP incorporate a

clinical dementia assessment for at least a subset of participants that may serve as a benchmark against which to evaluate cognitive performance-based classifications.

The HCAP and global partner studies may also consider a unified approach to clinical dementia assessments in order to improve our understanding of cross-country differences in the sensitivity and specificity of the HCAP and its subcomponents. Such an effort could capitalize on the innovation undertaken for clinical assessments in the LASI-DAD study, in which an online platform for consensus-based dementia diagnosis was validated against in-person diagnoses.³⁰ This online platform allowed for broad geographic coverage of its dementia assessments rather than restricting LASI-DAD respondents to those living in urban centers and in close proximity to multiple clinical experts required for consensus-based in-person diagnosis. Such a platform could allow for broader geographic coverage within each international HCAP study, and also may have the potential to facilitate more unified, cross-country diagnostic assessments.

4. Consider phasing in a HCAP baseline assessment for people under age 65

The HCAP sample includes respondents 65 years and older. The sample age range was intentionally lowered to include slightly younger participants as compared to ADAMS, which included participants age 70 and older. The HCAP based its age criterion on the fact that Medicare begins at age 65; HCAP aims to shed light on the costs of dementia, and many of the costs incurred at the federal level would be driven by Medicare expenditures.⁵ The HCAP authors also rightly indicated that lowering the starting age also allows for understanding earlier stages of cognitive impairment. We suggest that lowering the starting age even further, potentially to include those 50 years and older, as with the broader HRS, would have many

advantages. Select international partner studies already include respondents younger than 65 years of age.

Including younger participants could also have important benefits for understanding midlife determinants of dementia and cognitive decline, particularly if the same respondents are followed as they age. It is now well understood that the biology of dementia unfolds over multiple decades. For example, by the early 50s, genetic profiles associated with ADRD predict lower body weight (Brenowitz, AJE 2021) and worse performance on specific cognitive tests (Zimmerman, unpublished analyses in the UK Biobank). Many population-level prevention and intervention efforts should be focused on the decades spanning mid-life.⁴ It is challenging to identify causes of dementia risk that are relevant for mid-life prevention and intervention efforts with an analytic sample that begins just as more advanced stages of dementia are more commonly presenting themselves – well after earlier opportunities for prevention have passed.

Starting the analytic sample at younger ages may also help ameliorate the selection bias present in many cognitive aging studies that begin in late life.³¹ Concerns about selection bias may not be as relevant if the HCAP is simply used as a tool to generate dementia classifications that are imposed on the broader HRS analytic sample. However, the empirical studies produced with the HCAP have yet to use the study in this way and have instead used the HCAP sample only to evaluate within-sample associations; without statistical correction, these analyses are likely subject to greater selection bias because of the age restriction than parallel analyses using the broader HRS sample. Lowering the starting age for HCAP may have the ancillary benefit of including a more diverse group of participants (aligned with our first recommendation) given that younger birth cohorts in the US are more diverse than their older counterparts.⁹

One potential concern is that variability in cognitive performance scores will be limited among those under 65 years of age, thereby reducing the utility of supplementing the HCAP sample with younger participants. These younger participants may not contribute meaningfully to the HCAP goals of providing dementia classifications, given that relatively few participants younger than 65 will be classified as having mild cognitive impairment or dementia. Including these participants may therefore come with added costs with little apparent benefit for the narrow goal of creating dementia classifications. However, in the context of repeated pharmacological failures for dementia treatment and a growing emphasis on midlife prevention, we argue that today's investments in population-level research that begins in midlife are critical down payments on future knowledge about dementia prevention, regardless of immediate yield.

5. Drop low-performing items from HCAP to increase feasibility

While the comprehensive nature of the HCAP is a strength and there have been tremendous investments in mirroring the HCAP measures across global partner studies, there are multiple reasons to consider modifications that might make the HCAP easier to scale. As we have argued, there is particular urgency around including a more diverse subset of HRS respondents and in administering repeated, more frequent assessments across the same set of respondents in order to be able to compare within-person slopes in cognitive performance cross-nationally. It may be necessary to consider dropping items from the HCAP in order to increase the feasibility of scaling the study in these ways.

The factor structure of HCAP allows separate estimation of episodic memory, executive functioning, processing speed, language, and visuo-construction domains. Items load strongly on these domains, as expected given the selection process (Table below reproduced from

Zahodne et al 2021).³² However, some of these domains include multiple items that may be considered for removal. Reliability and validity of the episodic memory assessment is probably not strongly improved by inclusion of the “Brave man” story for example, in the context of 8 other measures of episodic memory, all of which have factor loadings above 0.6. Dropping the lowest two performing language items would also likely have limited impact on reliability.

TABLE 1 ABOUT HERE

We propose that the HCAP consider a framework for continual evaluation of item performance. We have already acknowledged the lack of a gold standard against which to definitively assess item performance. However, items could be assessed against several metrics that collectively serve as a benchmark against which to evaluate their utility: burden, sensitivity to age, contribution to reliability of domain measure (including across racial/ethnic groups), and association with subsequent adverse outcomes, such as death. Top among these is burden of assessing the item. Items that are easy or brief may be retained even if less reliable. Detailed data on administration time from the 2016 implementation can be used to compare time demand versus contribution to measurement reliability. This should be considered within the context of specific subgroups (i.e., if performance of some items varies by subgroups). If an HCAP item does not vary or demonstrates limited variation with respondents’ age, it may be an indication that the item is instead measuring an aspect of cognitive performance that is established earlier in life and does not decline over time, and is therefore not a useful measure of cognitive aging. Items could also be evaluated against clinical dementia diagnoses, if those are incorporated into the HCAP going forward, and/or in relationship to subsequent mortality, institutionalization, or incidence of ADL limitations. Items that are overly burdensome and/or not associated with age, clinical diagnosis, or patient-centered outcomes could be considered for removal.

We acknowledge that removing items needs to be weighed against an interest in including well-known cognitive assessment batteries, which were carefully considered in designing the HCAP and meant to enable harmonization across a number of studies both in the U.S. and globally. These include the Mini-Mental State Examination – included in numerous studies of cognitive aging worldwide –, the Telephone Interview for Cognitive Status – included in the core HRS--, and the assessment battery included in the 10/66 global studies. Removing individual items from these batteries could challenge efforts to harmonize HCAP data with data from other sources and to create necessary crosswalks to the core HRS. These needs for harmonization need to be balanced against the needs for an assessment battery that may be more feasibly scaled to a larger, more diverse sample and administered more frequently.

6. Enhance documentation and other resources to promote cross-national comparisons

NIA has invested substantially in providing outstanding support to data users of HRS and to some extent for international sister studies. HCAP and its global counterparts need similar support. Investing in enhanced documentation that presents detailed protocols for each study and highlights cross-study differences as well as the rationale for any differences would enhance usability. The detailed item-by-item codebook provided by the LASI-DAD study could serve as a template that could be expanded to compare specific cognitive assessment tasks and the rationale for study differences across international partner studies.

Harmonized code in common statistical languages (including in R or other software that is free or low-cost) should similarly be made available in a way that is modeled after the code available for the core studies on the Gateway to Global Aging. The Mex-Cog study team already provides publicly available code in SAS and STATA that can be used to clean their HCAP

equivalent data. This code both facilitates the use of the Mex-Cog data by reducing the arduous task of cleaning raw data across the many, many variables yielded from the extensive cognitive assessment battery and helps reduce the chance that coding errors made during data cleaning contribute to incorrect and/or inconsistent results. The code provided by the Mex-Cog team could be used as a template for the HCAP and other partner studies.

Training short courses or applied data workshops using the HCAP and international partner study data could also enhance uptake. There is a longstanding and successful tradition of such training opportunities related to HRS. These could be fielded remotely to promote international participation and collaboration.

Small embedded substudies to evaluate language or version effects would be valuable if these are feasible. HRS has a tradition of embedding randomized substudies, for example to evaluate mode effects for phone versus in-person interviews. Likewise, bilingual participants in HRS or partner studies could be randomly assigned to alternative language versions of the test. These randomized studies do not necessarily have to be completed in HRS or HRS sister study participants; age-appropriate volunteers who do not participate in the studies would likely be adequate for the purposes of creating the international cross-walks.

7. Add blood AD biomarkers.

The landscape of blood-based AD biomarkers is expanding rapidly, with multiple studies showing that blood-based markers can predict future progression to AD.³³⁻³⁵ These developments are exciting, given that AD biomarkers have long been dependent on invasive and/or costly procedures – namely, brain imaging, lumbar punctures, or autopsies. The logistic complications and high costs of these invasive biomarkers has contributed to highly selected samples, with no

population-representative research on AD biomarkers available to-date. Racial/ethnic minorities, those with low socio-economic status, and those living far from academic medical centers have historically been excluded from biomarker-based AD/ADRD research,³⁶ despite the fact that these are the very populations that face a disproportionate burden of AD/ADRD. The severe sample selection bias present in highly select studies that include neuroimaging data has been shown to contribute to misleading conclusions.³⁷ ³⁶ Less invasive and costly means of assessing biological indicators of AD may allow for population-representative research that includes historically excluded groups and reduces selection bias. Studies have already shown these measures to be feasibly collected in multi-ethnic, community-based samples. ³⁴

Nevertheless, studies exploring these novel biomarkers to-date have largely been based on the same highly select set of respondents included in existing AD biomarker studies, potentially replicating the same issues described above. Specifically, they are vulnerable to strong selection biases, largely do not reflect the diversity of the US older population, and would not necessarily give a clear benchmark for performance of the biomarkers in different groups of people, including across racial/ethnic groups.

As a population-based study with oversamples of Black and Latinx community-dwelling older adults, the HCAP has a potentially important role to play in advancing AD/ADRD research via the inclusion of blood-based AD biomarkers. Adding these to HRS and specifically to the HCAP, ideally with an enhanced, more diverse sample, would be extremely valuable for the field as a whole. The HCAP team already plans to collect venous blood in its second round of data collection. In addition, numerous international partner studies have demonstrated the feasibility of collecting blood as part of their HCAP equivalent studies or their core studies, suggesting that

there may be an opportunity to expand biomarker-based AD research across harmonized, global samples.

We realize that the science around blood-based biomarkers for dementia is changing rapidly; we cannot speak to the preferred biomarkers or the technical feasibility. Given the early state of the science, the HCAP has the opportunity to provide rigorous population-based evidence before biases based on highly select samples (e.g. clinical populations, community-based populations near academic medical centers) are baked into foundational knowledge on blood-based AD biomarkers. If the HCAP waits until the science is more fully established, critical opportunities to evaluate these biomarkers within a prospective, population-based cohort study in association with longitudinal AD/ADRDs outcomes may be missed.

In addition, the potential inclusion of respondents 50 years and older represents a particularly important opportunity for studying how values on AD biomarkers in mid-life might predict later-life dementia risk – as well as how multi-level environmental and individual-level factors may modify the association between AD biomarker values and future dementia risk. Long-term prospective data that is inclusive of younger age groups may be important for informing prevention and early intervention efforts, including early pharmacological intervention should an effective treatment become available.

8. Improve ability to assess consequences of dementia, and specifically measures of long-term services and supports utilization and caregiver experiences.

As already noted, a core goal of the HCAP is to understand the determinants, prevalence, consequences, and costs of dementia in the US. However, the ability of the HRS – as well as the HCAP – to understand the *consequences* of dementia is limited. In particular, the HRS has historically not included sufficient questions in its core surveys on respondents' utilization of

long-term services and supports (LTSS), the specific nature of these LTSS (formal, informal, or both), whether respondents have a need for LTSS that is not being met, or the experiences and wellbeing of respondents' caregivers. Measures that allow for understanding respondents' utilization of and unmet needs for home and community-based services (HCBS) are particularly underrepresented in the HRS. HCBS utilization questions were the subject of an experimental module within the 2012 core HRS as well as the 2011 Health Care Mail Survey,^{38,39} although no questions to our knowledge have addressed whether respondents have an unmet need for these services and/or the consequences of going without any or adequate LTSS. While there are also opportunities to understand LTSS utilization patterns via linked Medicaid data, these analyses are limited to Medicaid beneficiaries who gave permission for data linkage. Furthermore, Medicaid claims data only reflects utilization patterns, rather than a comprehensive understanding of respondents' needs for care and the consequences of not receiving adequate services and supports to meet these needs.

We suggest that modifications to future iterations of the HCAP consider enhancing the ability of the HCAP and/or the core HRS to understand specific details about LTSS access and utilization, including both formal and informal home and community-based services. In particular, questions about unmet or inadequately met needs and their consequences are critical for informing policies and programs that seek to address dementia care.

The HCAP represents a particularly compelling opportunity for research on long-term services and supports generally and caregiving in particular. In the core HRS, interviews are primarily completed with respondents themselves. However, respondent-reported information on need for and utilization of LTSS may be of limited reliability if respondents are already experiencing dementia. On the other hand, informants may also not be able to provide fully

accurate information on respondents' needs for care and utilization patterns, depending on the nature of their relationship to the respondent. Unlike the core HRS, the HCAP includes respondent-informant dyads for all participants. This represents an important opportunity to triangulate information on the LTSS utilization between respondents and informants, potentially increasing the reliability of respondent or informant-only reports.

The inclusion of informants throughout the HCAP may also represent an important opportunity to ask directly about experiences of caregiving for the subset of informants who provide care to HRS respondents, including HRS respondents with dementia. Currently, the HRS relies on respondents (in the case of direct respondent interviews) to report on who, if anyone, provides care for their activities of daily living. However, family informants may have different perspectives on the extent to which they are providing in-home assistance and could be asked directly such that respondent and informant-reported information can be compared. In addition, these family members could report on their experiences of providing care, including their own needs and unmet needs (e.g. for mental health care, financial compensation). These questions would be of critical relevance to informing policies around support for family caregivers. For example, some states made an exception to allow for the family caregivers of Medicaid beneficiaries to be paid during the COVID-19 pandemic. Decision-making about whether this option is maintained or even more widely adopted will require rigorous population-based science.

Before the potential to measure and triangulate information on LTSS can be fully realized, the HCAP should consider additional measures of need for and utilization of LTSS, including HCBS. These additional measures might be patterned after existing studies, such as the National Health and Aging Trends Study (NHATS) and the California Health Interview Survey

LTSS follow-along module, so that multiple data sources may be compared and/or pooled in the future. Measures of caregiving experiences and needs may similarly come from the National Study of Caregiving, which is linked to the NHATS. While incorporating questions about LTSS, including formal and informal caregiving, into the HCAP would likely represent a substantial investment, information gained from these questions will contribute to improved population-based science necessary for informing current policy debates about where and how we should invest resources to support older adults – including those living with AD/ADRDs -- and their family members. In the absence of effective pharmacological interventions, investments in dementia care may be our best chance at improving the quality of life of those living with dementia and their caregivers.

CONCLUSION

The HCAP, along with its international equivalents, represents a critical effort in the landscape of population-based dementia research, both in the U.S. and globally. While the 2016 HCAP represents a tremendous step forward in this overall effort, the core goals of the project have yet to be fully realized and adoption of the study for many of its intended purposes (i.e. dementia classification, cross-national research) could be accelerated. We offer a series of recommendations with the hope of supporting the HCAP in realizing its core goals. Some of these recommendations simply build on the original plans of the HCAP team (e.g. to field longitudinal HCAP assessments, to collect venous blood). We believe that these recommendations may help extend the influence of the HCAP even further by, for example, providing rigorous population-based evidence on blood-based AD biomarkers or on the unmet needs of individuals with dementia and their loved ones. Most critically, prior to making these additional investments, we suggest that the HCAP set a path to ensuring that the analytic sample

is more inclusive of minoritized racial and ethnic groups that have been both disproportionately burdened by AD/ADRDs and excluded from AD/ADRDs research.

REFERENCES

1. World Health Organization. *Global action plan on the public health response to dementia, 2017-2025*. 2017.
2. Ackley SF, Zimmerman SC, Brenowitz WD, et al. Effect of reductions in amyloid levels on cognitive change in randomized trials: instrumental variable meta-analysis. *BMJ*. 02 25 2021;372:n156. doi:10.1136/bmj.n156
3. 2020 Alzheimer's disease facts and figures. *Alzheimers Dement*. Mar 10 2020;doi:10.1002/alz.12068
4. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet*. 08 2020;396(10248):413-446. doi:10.1016/S0140-6736(20)30367-6
5. 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. doi:10.1159/000503004
6. Langa KM, Plassman BL, Wallace RB, et al. The Aging, Demographics, and Memory Study: study design and methods. *Neuroepidemiology*. 2005;25(4):181-91. doi:10.1159/000087448
7. Wu Q, Tchetgen Tchetgen EJ, Osypuk TL, White K, Mujahid M, Maria Glymour M. Combining direct and proxy assessments to reduce attrition bias in a longitudinal study. *Alzheimer Dis Assoc Disord*. 2013 Jul-Sep 2013;27(3):207-12. doi:10.1097/WAD.0b013e31826cfe90
8. Mayeda ER, Glymour MM, Quesenberry CP, Whitmer RA. Inequalities in dementia incidence between six racial and ethnic groups over 14 years. *Alzheimers Dement*. Mar 2016;12(3):216-24. doi:10.1016/j.jalz.2015.12.007
9. Administration on Aging. *2019 Profile of Older Americans*. 2020.
10. Mejia-Arango S, Nevarez R, Michaels-Obregon A, et al. The Mexican Cognitive Aging Ancillary Study (Mex-Cog): Study Design and Methods. *Arch Gerontol Geriatr*. Jul 27 2020;91:104210. doi:10.1016/j.archger.2020.104210
11. Meng Q, Wang H, Strauss J, et al. Validation of neuropsychological tests for the China Health and Retirement Longitudinal Study Harmonized Cognitive Assessment Protocol. *Int Psychogeriatr*. 12 2019;31(12):1709-1719. doi:10.1017/S1041610219000693
12. Lee J, Banerjee J, Khobragade PY, Angrisani M, Dey AB. LASI-DAD study: a protocol for a prospective cohort study of late-life cognition and dementia in India. *BMJ Open*. 07 2019;9(7):e030300. doi:10.1136/bmjopen-2019-030300
13. Cadar D, Abell J, Matthews FE, et al. Cohort Profile Update: The Harmonised Cognitive Assessment Protocol Sub-study of the English Longitudinal Study of Ageing (ELSA-HCAP). *Int J Epidemiol*. 07 09 2021;50(3):725-726i. doi:10.1093/ije/dyaa227
14. Power MC, Bennett EE, Turner RW, et al. Trends in Relative Incidence and Prevalence of Dementia Across Non-Hispanic Black and White Individuals in the United States, 2000-2016. *JAMA Neurol*. Mar 01 2021;78(3):275-284. doi:10.1001/jamaneurol.2020.4471
15. Lin PJ, Daly AT, Olchanski N, et al. Dementia Diagnosis Disparities by Race and Ethnicity. *Med Care*. Aug 01 2021;59(8):679-686. doi:10.1097/MLR.0000000000001577

16. Tsoy E, Kiekhofe RE, Guterman EL, et al. Assessment of Racial/Ethnic Disparities in Timeliness and Comprehensiveness of Dementia Diagnosis in California. *JAMA Neurol.* Jun 01 2021;78(6):657-665. doi:10.1001/jamaneurol.2021.0399
17. Crimmins EM, Kim JK, Langa KM, Weir DR. Assessment of cognition using surveys and neuropsychological assessment: the Health and Retirement Study and the Aging, Demographics, and Memory Study. *J Gerontol B Psychol Sci Soc Sci.* Jul 2011;66 Suppl 1:i162-71. doi:10.1093/geronb/gbr048
18. Collaborators GD. 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.* 08 11 2021;21(1):241. doi:10.1186/s12911-021-01590-y
19. Zahodne LB, Glymour MM, Sparks C, et al. Education does not slow cognitive decline with aging: 12-year evidence from the victoria longitudinal study. *J Int Neuropsychol Soc.* Nov 2011;17(6):1039-46. doi:10.1017/S1355617711001044
20. Seblova D, Berggren R, Lövdén M. Education and age-related decline in cognitive performance: Systematic review and meta-analysis of longitudinal cohort studies. *Ageing Res Rev.* 03 2020;58:101005. doi:10.1016/j.arr.2019.101005
21. Walter S, Dufouil C, Gross AL, et al. Neuropsychological Test Performance and MRI Markers of Dementia Risk: Reducing Education Bias. *Alzheimer Dis Assoc Disord.* 2019 Jul-Sep 2019;33(3):179-185. doi:10.1097/WAD.0000000000000321
22. Mungas D, Beckett L, Harvey D, et al. Heterogeneity of cognitive trajectories in diverse older persons. *Psychol Aging.* Sep 2010;25(3):606-19. doi:10.1037/a0019502
23. Banerjee J, Jain U, Khobragade P, et al. Methodological considerations in designing and implementing the harmonized diagnostic assessment of dementia for longitudinal aging study in India (LASI-DAD). *Biodemography Soc Biol.* 2020 Jul-Sep 2020;65(3):189-213. doi:10.1080/19485565.2020.1730156
24. Ganguli MR, G., Chandra V, Sharma S, et al. A Hindi version of the MMSE: The development of a cognitive screening instrument for a largely illiterate rural elderly population in India. *International Journal of Geriatric Psychiatry.* 1995;10(5):367-377.
25. Kobayashi LC, Gross AL, Gibbons LE, et al. You say tomato, I say radish: can brief cognitive assessments in the US Health Retirement Study be harmonized with its International Partner Studies? *J Gerontol B Psychol Sci Soc Sci.* Nov 2020;doi:10.1093/geronb/gbaa205
26. Chan KS, Gross AL, Pezzin LE, Brandt J, Kasper JD. Harmonizing Measures of Cognitive Performance Across International Surveys of Aging Using Item Response Theory. *J Aging Health.* Dec 2015;27(8):1392-414. doi:10.1177/0898264315583054
27. Weuve J, Proust-Lima C, Power MC, et al. Guidelines for reporting methodological challenges and evaluating potential bias in dementia research. *Alzheimers Dement.* Sep 2015;11(9):1098-109. doi:10.1016/j.jalz.2015.06.1885
28. Vivot A, Power MC, Glymour MM, et al. Jump, Hop, or Skip: Modeling Practice Effects in Studies of Determinants of Cognitive Change in Older Adults. *Am J Epidemiol.* Feb 2016;183(4):302-14. doi:10.1093/aje/kwv212
29. Goldberg TE, Harvey PD, Wesnes KA, Snyder PJ, Schneider LS. Practice effects due to serial cognitive assessment: Implications for preclinical Alzheimer's disease randomized

controlled trials. *Alzheimers Dement (Amst)*. Mar 2015;1(1):103-11.

doi:10.1016/j.dadm.2014.11.003

30. Lee J, Ganguli M, Weerman A, et al. Online Clinical Consensus Diagnosis of Dementia: Development and Validation. *J Am Geriatr Soc*. 08 2020;68 Suppl 3:S54-S59.

doi:10.1111/jgs.16736

31. Mayeda ER, Filshtein TJ, Tripodis Y, Glymour MM, Gross AL. Does selective survival before study enrolment attenuate estimated effects of education on rate of cognitive decline in older adults? A simulation approach for quantifying survival bias in life course epidemiology. *Int J Epidemiol*. 10 01 2018;47(5):1507-1517. doi:10.1093/ije/dyy124

32. Zahodne LB, Morris EP, Sharifian N, Zaheed AB, Kraal AZ, Sol K. Everyday discrimination and subsequent cognitive abilities across five domains. *Neuropsychology*. Aug 03 2020;doi:10.1037/neu0000693

33. Palmqvist S, Tideman P, Cullen N, et al. Prediction of future Alzheimer's disease dementia using plasma phospho-tau combined with other accessible measures. *Nat Med*. 06 2021;27(6):1034-1042. doi:10.1038/s41591-021-01348-z

34. Brickman AM, Manly JJ, Honig LS, et al. Plasma p-tau181, p-tau217, and other blood-based Alzheimer's disease biomarkers in a multi-ethnic, community study. *Alzheimers Dement*. 08 2021;17(8):1353-1364. doi:10.1002/alz.12301

35. Schindler SE, Bateman RJ. Combining blood-based biomarkers to predict risk for Alzheimer's disease dementia. *Nature Aging*. 2021;1:26-28.

36. Barnes LL. Biomarkers for Alzheimer Dementia in Diverse Racial and Ethnic Minorities-A Public Health Priority. *JAMA Neurol*. 03 01 2019;76(3):251-253.

doi:10.1001/jamaneurol.2018.3444

37. LeWinn KZ, Sheridan MA, Keyes KM, Hamilton A, McLaughlin KA. Sample composition alters associations between age and brain structure. *Nat Commun*. 10 12 2017;8(1):874.

doi:10.1038/s41467-017-00908-7

38. Robinson KN, Menne HL, Gaeta R. Use of Informal Support as a Predictor of Home- and Community-Based Services Utilization. *J Gerontol B Psychol Sci Soc Sci*. 01 01 2021;76(1):133-140. doi:10.1093/geronb/gbaa046

39. Pepin R, Leggett A, Sonnega A, Assari S. Depressive Symptoms in Recipients of Home- and Community-Based Services in the United States: Are Older Adults Receiving the Care They Need? *Am J Geriatr Psychiatry*. Dec 2017;25(12):1351-1360. doi:10.1016/j.jagp.2017.05.021

Table 1. Standardized Factor Loadings from the HCAP Measurement Model, reproduced from Table 2 in Zahodne et al (2021).³²

Cognitive Measure	Estimate	Standard Error	p
Episodic Memory			
Word list immediate	0.79	0.01	<.001
Logical memory immediate	0.73	0.01	<.001
Brave man immediate	0.61	0.01	<.001
Word list delayed	0.76	0.01	<.001
Logical memory delayed	0.71	0.01	<.001
Brave man delayed	0.6	0.01	<.001
MMSE word list delayed	0.61	0.01	<.001
Word list recognition	0.63	0.01	<.001
Logical memory recognition	0.62	0.01	<.001
Constructional praxis-delay	0.73	0.01	<.001
Executive functioning			
Raven's	0.86	0.01	<.001
Number series	0.67	0.01	<.001
Trails B	-0.67	0.01	<.001
Processing speed			
Symbol digit modalities test	0.90	0.01	<.001
Trails A	-0.74	0.01	<.001
Backward counting	0.69	0.01	<.001
Letter cancellation	0.70	0.01	<.001
Language			
Animal fluency	0.77	0.01	<.001
TICS naming	0.75	0.01	<.001
MMSE naming	0.62	0.01	<.001
MMSE writing	0.56	0.01	<.001
Visuoconstruction			
Constructional praxis-copy	0.75	0.01	<.001
MMSE polygons	0.64	0.01	<.001

Note: MMSE - Mini-Mental State Exam; TICS - Telephone Interview for Cognitive Status