


# Cross-National Comparisons of Cognitive and Physical Health in Older Adults Across China, Japan, and Korea: A Systematic Review

INQUIRY: The Journal of Health Care Organization, Provision, and Financing  
Volume 58: 1–10  
© The Author(s) 2021  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/00469580211062451  
[journals.sagepub.com/home/inq](https://journals.sagepub.com/home/inq)  


Sanghun Nam, MS, OT<sup>1</sup> , Kevin T. Pritchard, MS, OTR<sup>2</sup>, Suyeong Bae, MS, OT<sup>1</sup> , and Ickpyo Hong, PhD, OTR<sup>1</sup> 

## Abstract

**Background:** Cross-national studies are an emerging research area in public health. Specifically, cross-national health comparisons are important for understanding the factors driving the success or failure of public health policies. Therefore, this study systematically analyzed studies that compared health status (physical health and cognition) of the older adults using national panel data for three East Asian countries—China, Japan, and Korea.

**Methods:** Google Scholar and PubMed were used for the literature search. The search strategy targeted papers published between 2005 and 2020, yielding a total of 2690 papers, of which seven were selected for the review. The Center for Evidence-Based Medicine (CEBM) criteria was used to assess study design quality. The risk of bias for non-randomized studies (RoBANS) tool, a quality assessment tool developed in Korea to evaluate non-randomized interventional studies, measured risk of bias.

**Results:** Of the seven included papers, two studies performed cognitive comparisons, four studies performed physical health comparisons, and one study compared both cognition and physical health. The studies selected for this study by CEBM criteria included four prospective cohort studies (Level 2B) and three expert opinions without explicit critical appraisal (Level 5). Risk of bias using the RoBANS tool found a risk of confounding variables in four out of seven papers. Finally, measurement items of cognitive and physical health differed across all three countries' panel surveys.

**Conclusion:** These results suggest that in order to compare East Asian health policies according to the aging society, it is necessary to develop consistent cognitive and physical health evaluation tools in the future.

## Keywords

cross-national, cognition, health, health surveys, community surveys

<sup>1</sup>Department of Occupational Therapy, College of Software and Digital Healthcare Convergence, Yonsei University, Wonju-si, Republic of Korea

<sup>2</sup>Division of Rehabilitation Sciences, University of Texas Medical Branch, Galveston, TX, USA

## Corresponding Author:

Ickpyo Hong, PhD, OTR, Department of Occupational Therapy, College of Software and Digital Healthcare Convergence, Yonsei University, 135 Backun Hall, 1, Yeonsedae-gil, Heungeop-myeon, Wonju-si, Gangwon-do, 26493, Republic of Korea.

Email: [ihong@yonsei.ac.kr](mailto:ihong@yonsei.ac.kr)



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and

Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

- **What do we already know about this topic?**
  - We know that the global aging problem requires cross-national health comparison studies of older adults. However, we are also aware that comparative studies in this regard are limited because of inconsistencies in the few survey items used to measure cognitive and physical health across countries. Information bias and measurement error are barriers to studying functional outcomes across countries.
- **How does your research contribute to the field?**
  - We aimed to define limitations and opportunities of using different countries' survey panel data to compare cognitive and physical health across countries. Such examination will contribute to reductions in measurement error and improve research on cognitive and physical health comparisons between countries in the future..
- **What are your research's implications toward theory, practice, or policy?**
  - A systematic review of cognitive and physical health comparisons between China, Japan, and Korea—neighboring East Asian countries—can serve as a foundation for natural experiments on health disparities or help establish health policies between the 3 countries in the future..

## Introduction

In Korea, the proportion of the population aged 60 years and above was 7.2% in 2000, which accelerated to 14% in 2017—the highest growth rate among developed countries.<sup>1</sup> The country is rapidly transforming into an aging society, and the response of different countries to this phenomenon can be examined by comparing health policies across the different Organization for Economic Cooperation and Development (OECD) countries.<sup>2</sup> Accordingly, for Korea, it is important to understand the health policies of neighboring Asian countries to enable it to take the appropriate countermeasures against aging.

Cross-national comparisons of cognitive and physical health in older adults are important for understanding the factors that influence the success or failure of public health policies. They are also used to establish potential policy interventions.<sup>3</sup> However, few studies have conducted cross-country comparisons of people's health status except those comparing North America and Europe.<sup>4</sup> Asia is one of the continents with the largest number of older adults worldwide. Thus, Asian countries need to plan for the future by establishing evidence-based policies to address the cognitive and physical problems caused by aging.<sup>5</sup> Among Asian countries, China, Japan, and Korea in East Asia are countries particularly close in geographic proximity. Therefore, comparing the cognitive and physical health of older adults in these countries could direct public health policies. Each of these three countries of East Asia is independently running its own aging panel surveys and there is a need to harmonize this data to reduce information bias.

In China, the China Health and Retirement Longitudinal Study (CHARLS) is a survey being conducted in 10 provinces aimed at representing residents of the country over the age of 45 years; in Japan, the Japanese Study on Aging and Retirement (JSTAR) is a multidisciplinary panel study aimed at Japanese seniors residing in the community; in Korea, the Korean Longitudinal Study of Aging (KLoSA) is a nationally representative panel survey of over 10,000 community

residents aged 45 years and older. Using data from these three surveys—the CHARLS, JSTAR, and KLoSA—it is possible to compare the cognitive and physical health of people in the three aforementioned East Asian countries. However, inconsistency between the survey items being compared between these countries can lead to inconsistent and biased results.<sup>6</sup> For example, a problem arises when there are only a few common items in the database between the CHARLS, JSTAR, and KLoSA. Assuming that there are few common items after extracting cognitive items from different studies to compare individuals' cognitive function, comparisons between countries thus becomes impossible or limited at best. Furthermore, when there are common items and the rating scales used between items differ, it is difficult to compare findings between countries due to the increased risk for measurement error. Another challenge is that the influence of culture cannot be excluded when research results are interpreted and understood between multiple countries.<sup>7</sup> Given that the panel surveys of each country mentioned above are based on a country-specific database, there is likely inherent confounding related to cross-country comparisons of cognitive and physical health. Since there is already inherent variability between countries, it is imperative to reduce variability between measures of function.

We aimed to examine whether the panel survey databases can support cross-country comparisons of health, as well as to identify the limitations of such comparisons. The main purpose of this study was to examine the usefulness and limitations of each country's panel database through a systematic review of the literature comparing the cognitive and physical health of older adults in three East Asian countries: China, Japan, and Korea. The reviewed literature used data from the CHARLS, JSTAR, and KLoSA. This comparison of cognitive and physical health between China, Japan, and Korea can inform health care services and public health policies. For example, by comparing the cognitive and physical health of people in China and Korea, it is possible to identify the cognitive and physical health that people in Korea lack compared with those in China.

This research is scalable to the Health and Retirement Study panel survey and its seventeen international sister studies. Gaps in health between countries can be used to reinforce health policies tailored to a specific country. We aimed to understand how nationally representative panel survey databases have been used to make cross-country comparisons of people's cognitive and physical health.

This study answers the following research questions: 1) Can information from a national panel database be used to compare health across countries and 2) What are the measurement limitations concerning the comparisons between cognitive and physical health in China, Japan, and Korea using national panel databases?

## Methods

Articles comparing the cognition and physical health of older adults across the target countries using data from the CHARLS, JSTAR, and KLoSA were shortlisted. We undertook a systematic literature review to determine whether it would be viable to compare cognitive and physical health between the three countries. We followed the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) guidelines (online supplement) and registered the review protocol with the International Prospective Register of Systematic Reviews (PROSPERO; CRD42020208200, October 14, 2020). Ethical approval and consent to participate are not considered necessary for systematic review studies; hence, institutional review board approval was not sought.

### Information Sources

On October 20, 2020, we searched the literature using Google Scholar and PubMed with a combination of the keywords "CHARLS," "JSTAR," "KLoSA," "cognition," and "physical health." A final literature search was conducted on November 5, 2020, to update our results with any recent publications.

China's CHARLS is a survey being conducted in 10 provinces of the country aimed at representing residents of China over the age of 45 years. In 2008, a pilot survey of approximately 1500 households was conducted. With a structure and content closely modeled on the KLoSA, the CHARLS interviews respondents over the age of 45 years and their spouses.<sup>8</sup> Meanwhile, Japan's JSTAR is a multi-disciplinary panel study aimed at Japanese seniors residing in the community. Its reference sample included more than 4200 Japanese people aged 50–75 years. The JSTAR is designed to capture the same core concepts of the Health and Retirement Study (HRS) in the United States, especially those like the Study of Health, Aging, and Retirement in Europe (SHARE). The first wave of the study, conducted in 2007, started with a stratified random sample of five municipalities whose medical expense records were considered.<sup>9</sup> The KLoSA is a nationally representative panel survey of over 10,000

community residents aged 45 years and older in Korea. The KLoSA uses a lower age limit to reflect Korea's older adult retirement transition.<sup>10</sup> This age limit has been adopted in surveys in China and India, where the informal employment sector is relatively large. The baseline survey tool has the same core content as the HRS.<sup>11</sup>

### Literature Search

PubMed is a search engine that provides access to the MEDLINE database, which includes references to a wide range of topics related to health, including life sciences, biomedical science, and health psychology. Meanwhile, Google Scholar is a search engine for academic purposes, allowing researchers to search for articles and other similar publications. We limited our search using these two search engines to articles published between 2005 and 2020. The keywords we used were CHARLS, JSTAR, KLoSA, cognition, and physical health. The literature search was conducted by two researchers.

### Eligibility Criteria

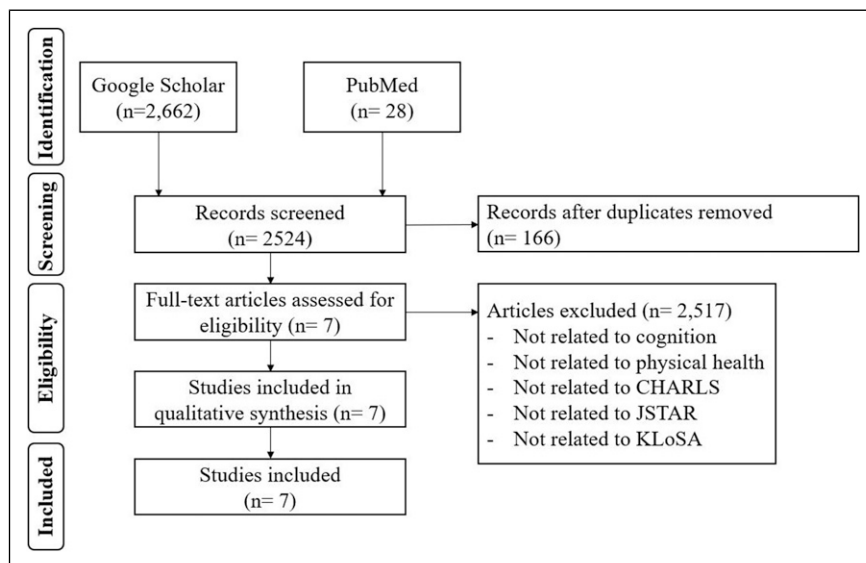
Two researchers determined the inclusion and exclusion criteria. Studies that met the following criteria were eligible for inclusion: compared cognition and physical health between China, Japan, and Korea; compared two or more countries among China, Japan, and Korea; comparative study using the CHARLS, JSTAR, or KLoSA database; written in Korean or English; published from 2005 to 2020.

### Literature Collection and Screening

Two researchers participated in the literature search and screening. After collecting the data and excluding duplicate entries, relevant literature was selected based on the inclusion criteria. We ensured review consistency between the reviewers and performed calibration exercises before starting the review.

### Data Extraction and Analysis

Review authors individually screened the titles and abstracts generated from the collected literature. Titles and abstracts that met the inclusion criteria or that were unclear were included. After inclusion based on title and abstract, two reviewers screened the full texts and determined whether they met the inclusion criteria. To resolve any discrepancies regarding the eligibility of a study, additional input was sought from the author(s) of the study. Discussions were held and external advice was sought to resolve disagreements between the two reviewers. As mentioned in the Information Sources section, 2690 articles were searched. Among these, 166 duplicate studies were removed, and 2517 studies that did not include the CHARLS, JSTAR, KLoSA, and cognitive and



**Figure 1.** The PRISMA flow diagram.

physical health were excluded. The procedure used to select the literature is shown in [Figure 1](#).

The final selected studies were classified into literature comparing cognitive function, literature comparing physical health, and literature comparing cognitive function and physical health. Next, the variables used in the classified literature were organized and investigated. Finally, an analysis method to compare cognitive function and physical health was investigated.

### Quality of the Literature

The Centre for Evidence-Based Medicine (CEBM) Levels of Evidence were used to evaluate the study quality based on the level of evidence, as discussed by Phillips et al.<sup>12</sup> The CEBM's levels of evidence are as follows: Level 1 includes systematic reviews, meta-analysis, and randomized controlled trial studies; Level 2A includes well-designed individual randomized controlled trials; Level 2B includes individual prospective cohort studies; Level 3A includes systematic reviews of case-control studies; Level 3B includes individual retrospective case-control studies; Level 4 includes case series; Level 5 includes expert opinions without explicit critical appraisal. To assess the risk of bias, we used Risk of Bias for Non-randomized Studies (RoBANS) version 2.0.<sup>13,14</sup> The RoBANS is a document quality assessment tool developed in Korea to evaluate non-randomized interventional research. We anticipated that the included studies would not be randomized studies because the CHARLS, JSTAR, and KLoSA are longitudinal panel surveys. The RoBANS covers six areas: the selection of participants, confounding variables, the measurement of exposure, the blinding of outcome assessments, incomplete outcome data, and selective outcome

reporting. It is structured such that the risk of bias can be assessed. Each area of the RoBANS tool can be judged as “low”, “high”, and “unclear” risk of bias. Two authors independently evaluated the quality of the selected studies based on the RoBANS tool, and items inconsistent with each other were determined by consulting the corresponding author.

## Results

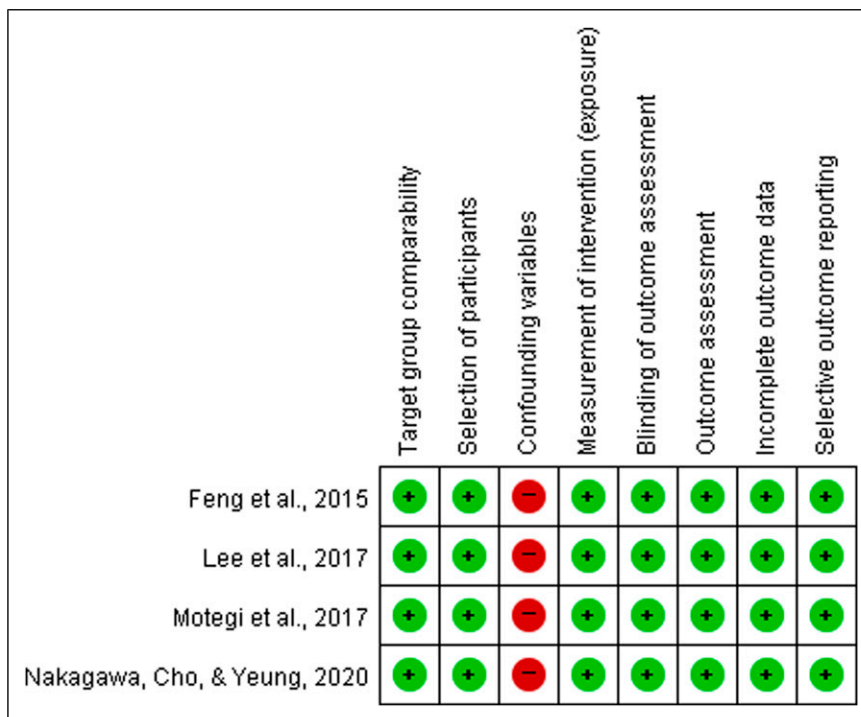
### Quality of the Literature

We extracted seven studies. Regarding their levels of evidence using the CEBM's criteria, four were revealed to be Level 2B studies,<sup>3,4,15,16</sup> and three were Level 5 studies.<sup>17-19</sup> Level 2B encompasses individual prospective cohort studies, and Level 5 refers to expert opinions without explicit critical appraisal. The results regarding the extracted studies' levels of evidence are shown in [Table 1](#). In addition, the risk of bias was evaluated using the RoBANS tool, and it was found that the overall risk of bias was low in all studies. However, four of the seven studies showed risk from confounding variables.<sup>3,4,15,16</sup> Although studies used similar items in the data of CHARLS, JSTAR, and KLoSA, the uniformity of items between each country's panel survey was inconsistent. In addition, the timing of the investigations differed for the CHARLS, JSTAR, and KLoSA panel investigations. Three of the seven studies were determined to be Level 5 studies, which simply compared panel survey questionnaire items by country, such as those of the CHARLS, JSTAR, and KLoSA.<sup>17-19</sup> Therefore, these three studies did not qualify for the RoBANS tool. The results of our analysis are summarized in [Figures 2 and 3](#).

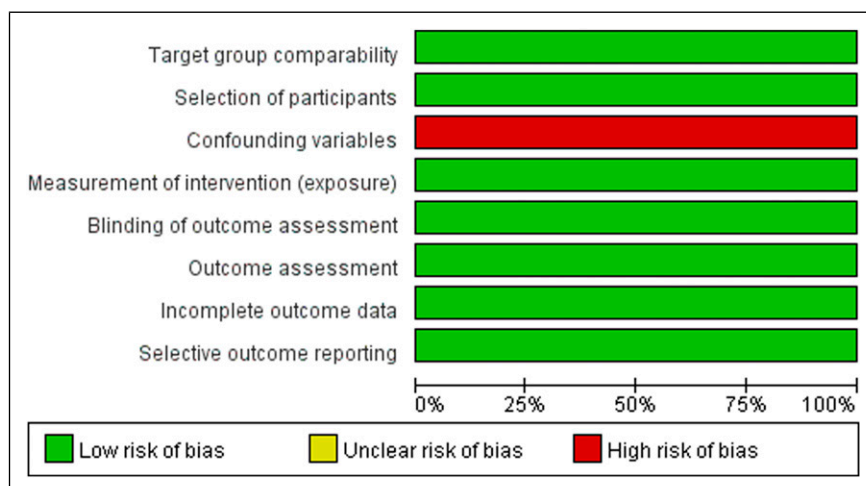
**Table 1.** Centre for Evidence-Based Medicine (CEBM) Levels of Evidence.

Evidence Level	n
Level 1A Systematic review of homogeneous RCTs (similar population, intervention, etc.) with or without meta-analysis	
Level 1B Well-designed individual RCT (not a pilot or feasibility study with a small sample size)	
Level 2A Systematic review of cohort studies	
Level 2B Individual prospective cohort study, low-quality RCT (e.g., <80% follow-up or low number of participants; pilot and feasibility studies); ecological studies; and two-group, non-randomized studies	Motegi et al, 2017 Feng et al, 2015
Level 3A Systematic review of case-control studies	Lee et al, 2018 Nakagawa et al, 2020
Level 3B Individual retrospective case-control study; one-group, non-randomized pre-posttest study; cohort studies	
Level 4 Case series (and low-quality cohort and case-control study)	
Level 5 Expert opinion without explicit critical appraisal	Hu and Lee, 2012 Shih et al, 2012 Wang et al, 2014

Note. RCT = randomized controlled trial.



**Figure 2.** Risk of bias summary for non-randomized studies (RoBANS).



**Figure 3.** Risk of bias graph for non-randomized studies (RoBANS).

**Table 2.** Cognitive and Physical Health Comparison Study Classification.

Study	Cross-National	Compared Item
Motegi et al, 2017	China, Japan, Korea	Cognition
Shih et al, 2012	China, Japan, Korea	Cognition
Feng et al, 2015	China, Korea	Physical health
Hu and Lee, 2012	China, Japan, Korea	Physical health
Wang et al, 2014	China, Japan, Korea	Physical health
Lee et al, 2018	China, Japan, Korea	Physical health
Nakagawa et al, 2020	China, Japan, Korea	Cognition and physical health

Note. Each study includes only content from China, Japan, and Korea.

### Classification of Literature According to Cognitive Comparison and Physical Health Comparison

Table 2 identifies the studies that measured cognition, physical health, or both. Two (28.57%) studies assessed only cognition,<sup>15,17,20</sup> four (57.14%) studies assessed only physical health,<sup>3,16,18,19</sup> and one (14.29%) study compared both.<sup>4</sup> Regarding comparisons between the three East Asian countries, six studies compared between China, Japan, and Korea, and one study compared between China and Korea. No studies compared between China and Japan.

### Comparing Cognitive Items Between Three East Asian Countries

To answer the first research question, Table 3 summarizes the variables used in the reviewed studies. Motegi, Nishimura, and Oikawa<sup>15</sup> used verbal memory and serial subtraction in panel surveys in each country to measure cognitive function. The verbal memory item was word recall. To evaluate word recall, the CHARLS and JSTAR use 10 words, but the KLoSA uses only three words. Serial subtraction uses an evaluation item that subtracts seven from 100 five

consecutive times. All items were coded as 0 and 1 and compared using the summed value. Due to the differences in verbal memory measurement, the KLoSA was excluded from the cognitive function comparison item because the KLoSA evaluates only three words in the word recall item. Cognitive function after retirement by gender, educational background, and wealth level was compared with the summed values of each item.

Shih, Lee, and Das<sup>17</sup> conducted a study comparing the panel survey of the HRS with the panel survey questionnaire of each international country. The study compared different cognitive domains through a survey to encourage rigorous cross-national and international comparative studies of aging populations. The items of verbal memory, orientation, and serial subtraction are the same for the CHARLS, JSTAR, and KLoSA. However, the CHARLS and the KLoSA also have visuoconstruction items, which the JSTAR does not. In addition, the KLoSA has language and executive functioning items that are not in the CHARLS and JSTAR. Furthermore, as mentioned earlier, the number of words used in the verbal memory items differs between countries. In the case of the KLoSA, the current year, month, and day are collected as one

**Table 3.** Variables used by Panel Survey.

Study	Variables Used in the Literature
Moteji et al, 2017	CHARLS (2010) Verbal memory (ten words), serial subtraction
	JSTAR (2009) Verbal memory (ten words), serial subtraction
	KLoSA (2010) Verbal memory (three words), serial subtraction
Shih, et al, 2012	CHARLS (2006) Verbal memory (ten words), orientation, visuoconstruction, serial subtraction
	JSTAR (2007) Verbal memory (ten words), orientation, serial subtraction
	KLoSA (2008) Verbal memory (three words), orientation, visuoconstruction, serial subtraction (100-7), language, executive functioning
Feng et al, 2015	CHARLS (2005) Heart disease, stroke, chronic lung disease, and cancer, activities of daily living, instrumental activities of daily living
	KLoSA (2006) Heart disease, stroke, chronic lung disease, and cancer, activities of daily living, instrumental activities of daily living
Hu and Lee, 2012	CHARLS (2008) Hypertension, diabetes, hypercholesterolemia, cancers excluding minor skin cancer, lung disease, heart problems, stroke, liver disease, stomach or other digestive disease, kidney disease, emotional nervous or psychiatric problems, memory-related disease, arthritis or rheumatism
	JSTAR (2007) Hypertension, diabetes, hypercholesterolemia, cancers excluding minor skin cancer, COPD excluding asthma, heart problems, stroke, liver disease, peptic ulcer disease, emotional nervous or psychiatric problems, memory-related disease, Parkinson's disease, arthritis or rheumatism, osteoporosis, hip or femoral fracture
	KLoSA (2006) Hypertension, diabetes, cancers excluding minor skin cancer, lung disease, heart problems, stroke, liver disease, emotional nervous or psychiatric problems, arthritis or rheumatism
Wang et al, 2014	CHARLS (2008) Smoking, drinking, vigorous, moderate, and mild exercise
	JSTAR (2007) Smoking, drinking
	KLoSA (2006) Smoking, drinking
Lee et al, 2018	CHARLS (2012) Activities of daily living, diabetes, heart conditions, stroke, hypertension, and arthritis
	JSTAR (2011) Activities of daily living, diabetes, heart conditions, stroke, hypertension, and arthritis
	KLoSA (2012) Activities of daily living, diabetes, heart conditions, stroke, hypertension, and arthritis
Nakagawa, et al, 2020	CHARLS (2008) Cancer, chronic lung disease, diabetes, heart disease, and stroke, activities of daily living, time orientation and serial subtraction
	JSTAR (2009) Cancer, chronic lung disease, diabetes, heart disease, and stroke, activities of daily living, time orientation and serial subtraction
	KLoSA (2008) Cancer, chronic lung disease, diabetes, heart disease, and stroke, activities of daily living, time orientation and serial subtraction

item, but in the case of the CHARLS and the JSTAR, these are collected as three items.

Finally, Nakagawa, Cho, and Yeung<sup>4</sup> compared successful aging in China, Japan, and Korea. The study compared individuals' cognitive and physical functions. As a study method, the dependent variables of no disease, no impairment in daily living activities, normal cognitive function, and active participation were used for comparison. No disease was defined as an individual who did not have any of the following: cancer, chronic lung disease, diabetes, heart disease, or stroke. The cognitive function measurements were the time orientation and serial subtraction. In the case of cognitive function, the current year, month, day, and day of the time orientation were each assigned 1 point, giving a total of 0 to 4 points, and each of the five subtractions was given 1 point, for a total of 5 points. The median value of 9 points, which was the sum of the time orientation and subtraction

scores, was 6 points, and scores above 6 points were coded as 1, while points below 5 points were coded as 0. Active participation was defined as participation in social activities, such as club visits, community activities, volunteer work, leisure activities, and engaging in paid work. Comparing these variables shows that successful aging was the highest in Japan and lowest in rural China.

### *Comparing Physical Health Items Between Three East Asian Countries*

Regarding the panel survey data from the three countries, on average, items of chronic disease, lifestyle (drinking and smoking), and activities of daily living (ADL) were used to compare individuals' physical health. The CHARLS, JSTAR, and KLoSA data variables used to compare the physical health of each study are summarized in Table 3. Feng, Son,

and Zeng<sup>16</sup> aimed to compare the prevalence and correlation of successful aging in Korea and China, and the databases they used were the CHARLS and the KLoSA. As dependent variables, heart disease, stroke, chronic lung disease, and cancer were used to assess subjects without comorbidities. ADL and instrumental ADL (IADL) were used to classify persons without disabilities. In addition, variables of those without depression, those who participated in society, and life satisfaction were used. Finally, those without comorbid diseases, those without disabilities, those without depression, those participating in society, and those with high life satisfaction were defined as having successful aging. Correlations of successful aging included variables of demographics (gender, age, and rural/urban residence), socioeconomic characteristics (financial status, education, and spouse), and health behaviors (smoking, drinking, and exercise). A total of 18.6% in China and 25.2% in Korea showed successful aging. When demographic variables were adjusted, the successful aging of older people in China was 51% lower than that in Korea. In addition, before socioeconomic variables were controlled, rural residence in China showed a negative correlation with successful aging but a positive correlation in Korea.

Hu and Lee<sup>18</sup> compared the chronic disease item of the HRS and the chronic disease item of each country's panel survey. The study compared different chronic diseases through a survey to encourage rigorous cross-national and international comparative studies of aging populations. The results of the study showed that hypertension, diabetes, cancers excluding minor skin cancer, heart problems, stroke, liver disease, emotional nervous or psychiatric problems, arthritis, or rheumatism items are the same across the CHARLS, JSTAR, and KLoSA. However, only the CHARLS and JSTAR were listed for hypercholesterolemia and memory-related diseases. In addition, only the CHARLS and KLoSA have lung disease items. Lastly, chronic diseases covered only by the CHARLS are stomach, other digestive, and kidney diseases. In addition, items belonging only to the JSTAR are chronic obstructive pulmonary disease, peptic ulcer disease, Parkinson's disease, osteoporosis, hip joint, and femur fracture. Despite this wide array of measured disease items, only 8 chronic disease items are similar in the CHARLS, JSTAR, and KLoSA databases.

Wang, Min, and Lee<sup>19</sup> compared HRS health behavior-related items with panel survey data from each country. The study also compared different lifestyles through a survey to encourage rigorous cross-national and international comparative studies of aging populations. The results of the study showed that the CHARLS, JSTAR, and KLoSA have the same categories of smoking and drinking. However, although the CHARLS has vigorous, moderate, and mild exercise items, the JSTAR and KLoSA do not.

The study of Lee et al.<sup>3</sup> compared how disability and morbidity rates differ by using data from the CHARLS, JSTAR, and KLoSA panel surveys, among others. In this study, the same chronic disease items for diabetes, heart

conditions, stroke, hypertension, and arthritis, and the same ADL items for bathing, dressing, eating, getting in and out of bed, and using the toilet were used. The results of the study showed that when the country with the highest disability and morbidity rates was compared with the country with the lowest rates, the hypertension rate was twice as high, the stroke rate was three times higher, and the arthritis rate was more than five times higher. In addition, it was reported that a higher gross domestic product and higher life expectancy were associated with diabetes, heart disease, and improved function in women. Men reported that national indicators of economic status were not significantly associated with the reported disease prevalence. In the study of Lee et al.,<sup>3</sup> although there are various chronic diseases, ADL, and IADL items in each database, only common items were extracted and compared; thus, most of the measured items could not be used.

## Discussion

In this systematic review, studies using national databases in China, Japan, and Korea were searched to investigate whether the cognitive and physical functions of older adults can be compared by country. Seven studies met our inclusion criteria and used the CHARLS, JSTAR, and KLoSA panel surveys. Overall, four of the seven studies were level 2B and three studies were level 5, and the risk of methodological bias was low. The surveys are sister studies of the United States HRS and have been harmonized by the National Institute on Aging (<https://hrs.isr.umich.edu/about/international-sister-studies>); therefore, the studies included similar survey items on cognition and physical health that enable cross-national health status studies. Although the current panel databases are promising for cross-national studies, there are still technical limitations for each panel survey database.

First, there were limitations related to cognitive measures. Shih, Lee, and Das<sup>17</sup> analyzed the cognitive items used in each country-specific survey and found that the cognitive items do not match. The CHARLS, JSTAR, and KLoSA have only three common cognitive evaluation items. This can be a problem leading to measurement error due to inconsistent comparison results due to discrepancies in the items being compared.<sup>6</sup> In addition, Motegi, Nishimura, and Oikawa<sup>15</sup> used verbal memory and serial subtraction of the CHARLS, JSTAR, and KLoSA. However, when measuring verbal memory, the KLoSA uses only three words, whereas the CHARLS and JSTAR use 10 words.

Second, we found limitations in measuring physical health. In this study, four studies compared physical health in three countries using the items of chronic disease, lifestyle, and ADL in older adults from CHARLS, JSTAR, and KLoSA data. Feng, Son, and Zeng<sup>16</sup> and Lee et al.<sup>3</sup> extracted and compared only similar items from the CHARLS, JSTAR, and KLoSA. However, according to the study by Feng, Son, and Zeng,<sup>16</sup> the items were difficult to unify. This is problematic because this reduces the number of possible physical health



measures for researchers and remains a source for measurement error. Mathematically, the number of test items had an inverse relationship with a measurement precision.<sup>21</sup>

Third, it was found that language and cultural differences in using data from the CHARLS, JSTAR, and KLoSA, and differences in the year of the panel survey may have limitations during comparative study. Problems with language include different nuances and meanings depending on the linguistic interpretation of a given item in each panel survey. In the case of translating the languages between countries, if the meaning is not conveyed correctly, even if the same item is interpreted, measurement errors may occur.<sup>22</sup> In addition, some items could be interpreted differently depending on language and culture.<sup>23</sup> Singer et al,<sup>24</sup> suggested that cultural differences can lead to different health outcomes. For example, individuals living in rural cultures have higher mortality rates from poverty, adult smoking, lack of physical activity, and ischemic heart disease than individuals living in urban cultures.<sup>25</sup> Finally, the CHARLS, JSTAR, and KLoSA have different years of investigation. In the study by Nakagawa, Cho, and Yeung,<sup>4</sup> the duration of the survey was different, and the measurements were potentially influenced by the year in which the survey was conducted (2011 for the CHARLS, 2009 for the JSTAR, and 2008 for the KLoSA). Therefore, if the years differ, the accuracy of the comparison may decrease.

We concluded that there are important differences in the questions comparing cognition and physical health in China, Korea, and Japan. In addition, because the samples of each country participating in the surveys varied by year of self-report, language, and culture, any comparison could have a substantial measurement error. Considering that the survey period in each country also differed, it is difficult to generalize comparisons between countries. Future studies should develop a questionnaire that can facilitate comparisons among the three countries. For example, it may be possible to conduct a survey using the Delphi method and the International Classification of Functioning, Disability and Health (ICF) model by extracting the most similar cognitive and physical health items from the CHARLS, JSTAR, and KLoSA. Subsequently, the common item linking method of the Rasch model can be performed using the Delphi final result.<sup>26,27</sup> Using the cognitive and physical health items of the three countries with the Rasch model, future national policies can be compared. Furthermore, more accurate measures could allow for natural research design methodologies to compare policy and economic changes between countries without the high risk of information bias or measurement error. This research is feasible and has promising scalability as it could be extrapolated to all the HRS sister studies in order to make comparisons across the World.

### Limitations

Our findings should be interpreted with consideration of a few limitations. First, we did not rely on Medical Subject

Headings (MeSH) in the literature search. However, the terms CHARLS, JSTAR, KLoSA, and physical health that correspond to the panel survey in each country did not fit any MeSH terms. Therefore, we cast a wide net in our search strategy, looking for each country's panel surveys (i.e., the CHARLS, JSTAR, and KLoSA) and manually filtered articles using our inclusion/exclusion criteria. Second, few studies compared cognitive and physical health between countries using the CHARLS, JSTAR, and KLoSA. Third, we only included the CHARLS, JSTAR, and KLoSA, which are panel surveys from China, Japan, and Korea, respectively; other panel surveys from the three countries were not included. This allows us to scale our research to the remaining HRS sister studies in an effort to make global comparisons. Fourth, among the studies reviewed in this study, there were panel surveys from various countries except for the CHARLS, JSTAR, and KLoSA; however, we only dealt with data from the CHARLS, JSTAR, and KLoSA. As described above, many studies are being conducted in Europe and the United States to compare cognitive and physical health items caused by aging. However, comparisons of cognitive and physical health panel survey items due to aging in East Asia are understudied. Therefore, in this study, only the CHARLS of China, the JSTAR of Japan, and the KLoSA of Korea were analyzed. Finally, this study collected only manuscripts written in Korean and English. The authors of this study could read in Korean and English. Thus, to reduce errors that could result from incorrect interpretation of information due to language misunderstandings, we searched for documents in Korean and English only. Future research will benefit from developing international collaboration.

### Conclusion

We searched for studies comparing cognition and physical health in three East Asian countries and conducted a systematic review. We found that panel survey data could be used to compare people's cognitive and physical health between countries. However, few studies have compared cognitive and physical health between East Asian countries using such data. There are numerous limitations in the existing research comparing these cross-national outcomes due to desynchronized survey years, cultural and language variation, and overall heightened risk for measurement error. This systematic review reveals a need for robust linking methodologies to facilitate comparisons between the East Asian countries. As a follow-up study, the authors will be conducting Delphi research by extracting cognitive and physical health items from databases in China, Japan, and Korea using the ICF model. Accordingly, public health professionals, policymakers, and health care providers will benefit from tools to compare a client's cognitive and physical health regardless of their country of origin.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported in part by National Research Foundation of Korea (NRF) (2020R1GA1100265). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NRF.

## Data Accessibility Statement

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

## ORCID iD

Sanghun Nam  <https://orcid.org/0000-0002-0842-8955>

## References

- Kim T, Ahn J-h., Park B. Impact of regional characteristics on traffic safety by aging stage. *Journal of Korean Society of Transportation*. 2020;38(2):121-133.
- Chen C, Goldman DP, Zissimopoulos J, Rowe JW. Multi-dimensional comparison of countries' adaptation to societal aging. *Proc Natl Acad Sci Unit States Am*. 2018;115(37):9169-9174.
- Lee J, Phillips D, Wilkens J, et al. Cross-country comparisons of disability and morbidity: Evidence from the gateway to global aging data. *J Gerontol: Series A*. 2018;73(11):1519-1524.
- Nakagawa T, Cho J, Yeung DY. Successful aging in East Asia: Comparison among China, Korea, and Japan. *J Gerontol: Ser Bibliogr*. 2020;76(suppl 1):S17-S26.
- Pin NT, Muthuswamy B, Michael F, Richard H, Maria Sofia A, Seng Cheong L. Research and policy to achieve healthy aging in Asia: Recommendations from an expert workshop. *Open Longev Sci*. 2013;7(1).
- Brønnum-Hansen H. Ranking health between countries in international comparisons. *Scand J Publ Health*. 2014;42(3):242-244.
- Mauro R, Sato K, Tucker J. The role of appraisal in human emotions: A cross-cultural study. *J Pers Soc Psychol*. 1992;62(2):301-317.
- Yaohui Z, John S, Albert P, Yan S. *Chinese Health and Retirement Longitudinal Study-Pilot*. Beijing, China: Users Guide National School of Development, Peking University; 2009.
- Ichimura H, Hashimoto H, Shimizutani S. *Japanese Study of Aging and Retirement*. JSTAR First Results; 2009.
- Lee J. *Aging in South Korea: Findings from Korean Longitudinal Study*. Conference paper; 2007.
- Boo K-C, Chang J-Y. Korean longitudinal study of ageing: Research design for international comparative studies. *Survey Research*. 2006;7(2):97-122.
- Phillips B, Ball C, Badenoch D, Straus S, Haynes B, Dawes M. Oxford centre for evidence-based medicine levels of evidence. *BJU International*. 2011;107(5).
- Kim SY, Park JE, Lee YJ, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *J Clin Epidemiol*. 2013;66(4):408-414.
- Kim S. *The Tools of Study Design Algorithm for Medical Literature of Intervention and Risk of Bias for Non Randomized Studies*. Gangwon: Health Insurance Review & Assessment Service; 2013:31-51.
- Motegi H, Nishimura Y, Oikawa M. *Retirement and Cognitive Decline: Evidence from Global Aging Data*. Germany: University Library of Munich; 2017.
- Feng Q, Son J, Zeng Y. Prevalence and correlates of successful ageing: A comparative study between China and South Korea. *Eur J Ageing*. 2015;12(2):83-94.
- Shih RA, Lee J, Das L. *Harmonization of Cross-National Studies of Aging to the Health and Retirement Study Cognition*. WR-861/7. RAND Corporation; 2012.
- Hu P, Lee J. *Harmonization of Cross-National Studies of Aging to the Health and Retirement Study Chronic Medical Conditions*. WR-861/1. RAND Corporation; 2012.
- Wang S, Min J, Lee J. *Harmonization of Cross-National Studies of Aging to the Health and Retirement Study*. WR-861/4. RAND Corporation; 2014.
- Nishimura Y, Oikawa M. *Mental retirement: Evidence from global aging data*. MPRA Paper, 76431, Germany: University Library of Munich; 2017.
- Wright BD, Stone MH. *Best Test Design*. Chicago, IL: Mesa Press; 1979.
- Hongbo Y. The cultural differences and translation thinkings on Chinese and Western translation metaphors. *Journal of Chifeng University (Philosophy and Social Science Chinese Edition)*. 2017;2017(4):30.
- Torop P. Translation as translating as culture. *Sign Syst Stud*. 2002;30(2):593-605.
- Kagawa Singer M, Dressler W, George S, et al. Culture: The missing link in health research. *Soc Sci Med*. 2016;170:237-246.
- Hartley D. Rural health disparities, population health, and rural culture. *Am J Publ Health*. 2004;94(10):1675-1678.
- Andrich D. A rating formulation for ordered response categories. *Psychometrika*. 1978;43(4):561-573.
- Bond T, Yan Z, Heene M. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. England, UK: Routledge; 2015.