

Factors associated with healthy aging in Latin American populations

Received: 10 February 2023

Accepted: 11 July 2023

Published online: 10 August 2023

 Check for updates

Hernando Santamaria-Garcia ^{1,2,3,15} , **Agustín Sainz-Ballesteros**^{4,15}, **Hernán Hernández**^{4,5}, **Sebastian Moguilner**^{1,4,6,7}, **Marcelo Maito**⁴, **Carolina Ochoa-Rosales**⁴, **Michael Corley**⁸, **Victor Valcour**^{1,9}, **J. Jaime Miranda** ^{10,11,12,13}, **Brian Lawlor**^{1,14} & **Agustin Ibanez** ^{1,4,6,14} 

Latin American populations may present patterns of sociodemographic, ethnic and cultural diversity that can defy current universal models of healthy aging. The potential combination of risk factors that influence aging across populations in Latin American and Caribbean (LAC) countries is unknown. Compared to other regions where classical factors such as age and sex drive healthy aging, higher disparity-related factors and between-country variability could influence healthy aging in LAC countries. We investigated the combined impact of social determinants of health (SDH), lifestyle factors, cardiometabolic factors, mental health symptoms and demographics (age, sex) on healthy aging (cognition and functional ability) across LAC countries with different levels of socioeconomic development using cross-sectional and longitudinal machine learning models ($n = 44,394$ participants). Risk factors associated with social and health disparities, including SDH ($\beta > 0.3$), mental health ($\beta > 0.6$) and cardiometabolic risks ($\beta > 0.22$), significantly influenced healthy aging more than age and sex (with null or smaller effects: $\beta < 0.2$). These heterogeneous patterns were more pronounced in low-income to middle-income LAC countries compared to high-income LAC countries (cross-sectional comparisons), and in an upper-income to middle-income LAC country, Costa Rica, compared to China, a non-upper-income to middle-income LAC country (longitudinal comparisons). These inequity-associated and region-specific patterns inform national risk assessments of healthy aging in LAC countries and regionally tailored public health interventions.

Aging is not a uniform process across the world. Most of the research into cognitive and functional aging has traditionally been conducted in high-income settings within the United States and Europe, neglecting diverse populations and the specific combination of risk factors seen in Latin American and Caribbean (LAC) countries. Despite the urgent need to assess regional diversity and deliver tailored evidence for diverse populations^{1–4}, the evidence around healthy aging in LAC countries is lacking. Addressing this knowledge gap is essential because risks stem

from multiple disparity-related cumulative exposures affecting aging and dementia. Latin American populations have unique ethnic admixtures, education and sociodemographic heterogeneity. In addition, the current prevalence of dementia in LAC countries is estimated at 8.5% and is projected to be 19.33% by 2050, representing an increase of 220% approximately⁵. Such prevalence is higher compared to other regions⁵, including Europe (currently 6.9% and projected up to 7.7% by 2050) or North America (currently 6.5% and projected up to 12.1% by 2050)^{5,6}.

A full list of affiliations appears at the end of the paper.  e-mail: hernando.santamaria@gbhi.org; agustin.ibanez@gbhi.org

Previous studies in high-income countries (HICs)^{1,7} that capture risk factors of healthy and pathological brain aging were not accurate for low-income and middle-income countries (LMICs) or LAC countries^{1,6}. Thus, assessing specific risk factors in LAC countries constitutes a critical priority for understanding healthy aging.

Healthy brain aging refers to the functional brain ability that allows someone to live their life to their fullest capacity. Healthy aging is traditionally evaluated using cognitive and functional ability measures, constituting proxy markers of brain health⁸. Cognition involves multiple domains (that is, attention, problem-solving, learning and memory, among others), while functional ability encompasses personal activities of daily living (ADLs) and higher-order instrumental skills⁸.

Several factors have been associated with healthy or unhealthy aging outcomes in previous studies^{3,9}. These include demographic factors such as age and sex; social factors such as educational level, socioeconomic status (SES) and social support, together known as social determinants of health (SDH); health status (including cardiometabolic factors such as hypertension, diabetes and obesity, and falls); mental health symptoms (including depression and anxiety); and lifestyle factors (alcohol consumption, smoking and physical activity). Most of those factors have been described as potentially modifiable risk factors for dementia by the Lancet Commission on Dementia Prevention, Intervention, and Care³. Although multiple factors contribute to healthy and pathological aging in HICs, less-modifiable factors, such as age^{10–12} and sex^{13,14}, are considered top contributors to pathological aging^{10–12}. Healthy aging in HICs is also influenced by modifiable factors such as cardiometabolic factors, mental health symptoms and lifestyle factors^{5,15}. The relative importance of mentioned risk factors may be different and heterogeneous in countries with increased social and health disparities such as LAC countries^{3,16}. Compared to HICs, in LAC countries converging multiple factors were associated with pathological aging, including a substantial contribution of SDH¹⁷, a higher prevalence of cardiometabolic factors¹⁸, mental health symptoms¹⁹ and barriers to healthy lifestyle^{20,21}.

Previous evidence in healthy aging in LAC countries identified multiple gaps^{1,22–25}, including insufficient understanding of the unique determinants and risk factors of aging in the region, and no attempt to simultaneously assess the associations and interactions between the different potential risk factors of healthy aging. Omissions in evaluating such interactions can lead to spurious or improper relationships between risk factors and healthy aging outcomes. Moreover, emerging evidence suggests that models of risk do not always generalize from HICs to LMICs¹. Other research gaps in the region include limited studies combining cross-sectional and longitudinal methods; absence of automated data-driven approaches for assessing multiple risk and protective factors, with no inclusion of techniques to confirm the validity of results (that is, out-of-sample validation procedures), leading to potential biases due to the assumptions of a priori theoretical models; inadequate representation of diverse populations from HICs, upper-income and middle-income countries (UMICs), and LMICs; and lack of region-specific risk factors of cognition and functional ability. Our work aimed to address these gaps using a convergent approach.

Using cross-sectional and longitudinal approaches, we assessed multiple potential risk factors (demographic, SDH, health status, lifestyle and mental health symptoms) of cognition and functional

ability in healthy aging across LAC countries. We hypothesized that disparity-related social and health factors, as opposed to demographic factors such as age and sex, would be more important in healthy aging across LACs compared to data reported in previous studies from HICs. We also expected country-level differences in healthy aging risk factors in LAC countries according to their income categories.

Results

Using a data-driven approach based on machine learning procedures, we assessed multiple potential risk factors (including demographic factors, SDH, health status, mental health symptoms and lifestyle) that affect cognition and functional ability in healthy aging (total $n = 44,394$ participants; Fig. 1a–e). Cross-sectional and longitudinal data from nationally representative survey cohorts included various LACs with different levels of socioeconomic development (Fig. 1b). As a first step, a multimethod approach (Fig. 1f) comprising linear regression, elastic net, least absolute shrinkage and selection operator (LASSO), and ridge regression was implemented to ensure the robustness of our machine learning results. We choose ridge regression based on the convergence observed with this multimethod approach (Extended Data Tables 1 and 2, and Extended Data Fig. 1). We used harmonized databases from national aging surveys from four LACs with different socioeconomic indexes based on the World Data Bank categorization. Two countries ranked as Latin American-LMICs (LA-LMICs, Colombia ($n = 23,694$ participants) and Ecuador ($n = 5,235$ participants)) and two countries as LA-HICs (Chile ($n = 1,301$ participants)) and Uruguay ($n = 1,450$ participants) were incorporated. We ran a second longitudinal analysis ($n = 9,258$ participants) from Costa Rica ($n = 5,694$ participants), a UMIC (LA-UMIC), and compared it with a non-LA-UMIC (China, $n = 3,564$ participants), which has the same socioeconomic category.

Cross-sectional analyses in LAC countries

We first assessed the main risk factors of cognition and functional ability using cross-sectional national surveys from all LAC countries. We then used the same approach focusing on LA-LMICs versus LA-HICs and finally across different countries. The multimethod results revealed consistency across methods and metrics (R^2 , Cohen's F_2 , mean square error (MSE), root mean square error (RMSE)) and a high coherence in the weight and ranking of risk factors (by using β estimates) of healthy aging in all countries (Fig. 1k).

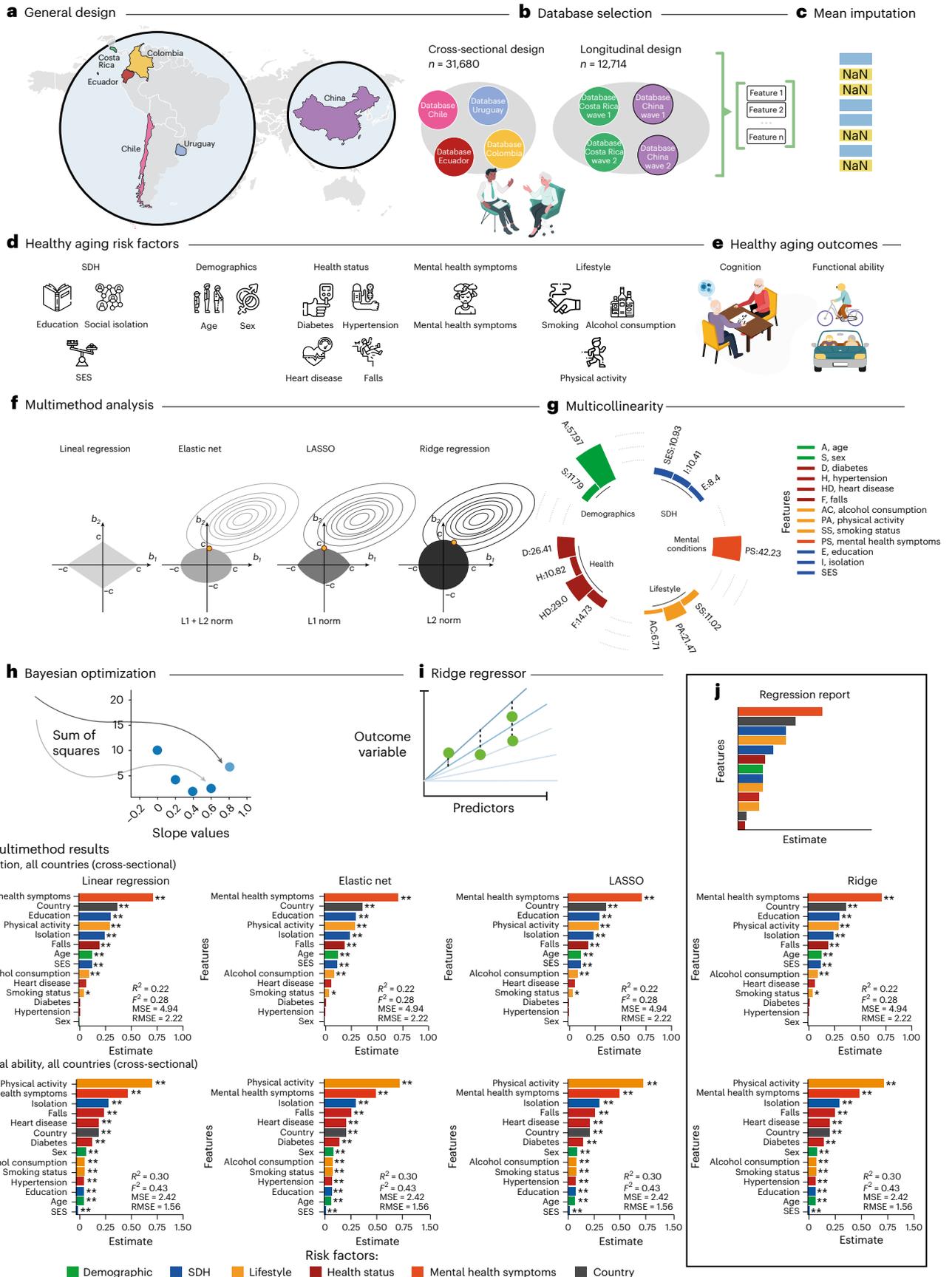
A model assessing the risk factors of cognition from all LAC countries was significant ($F_{1, 28,109} = 565.61$, $P < 0.0001$, $R^2 = 0.22 \pm 0.026$, $F_2 = 0.28$, $MSE = 4.94$, $RMSE = 2.42$) (Fig. 2a). The most important risk factors in order of relevance were: mental health symptoms ($\beta = 0.71$, $P < 0.0001$), country ($\beta = -0.37$, $P < 0.01$), physical activity ($\beta = -0.3$, $P < 0.0001$), education ($\beta = 0.3$, $P < 0.0001$), isolation ($\beta = 0.25$, $P < 0.0001$), age ($\beta = -0.13$, $P < 0.0001$) and SES ($\beta = 0.13$, $P < 0.0001$). Moreover, alcohol consumption ($\beta = -0.11$, $P < 0.01$) and smoking (lifestyle, $\beta = -0.05$, $P < 0.05$) were also significant. No significant effects of cardiometabolic factors were observed for cognition (Table 1 and Fig. 2a).

The model predicting functional ability was significant ($F_{1, 27,991} = 866.71$, $P < 0.00001$; $R^2 = 0.3 \pm 0.026$, $F_2 = 0.43$, $MSE = 2.42$, $RMSE = 1.56$). In order of importance, the most relevant risk factors

Fig. 1 | Methodological framework. **a**, General design of the study describing the countries included in the cross-sectional analyses (Chile, Uruguay, Colombia and Ecuador) and in the longitudinal analyses (Costa Rica and China). **b**, Database selection in the cross-sectional ($n = 31,680$ participants) and longitudinal ($n = 9,258$ participants) studies (total $n = 40,938$ participants). **c**, Imputation procedures. **d**, Risk factors, including demographics, SDH, health status (cardiometabolic factors and falls), mental health symptoms and lifestyle risk factors. **e**, Outcomes: cognition and functional ability. **f**, Multimethod approach, including different regressions (linear regression, elastic net, LASSO, ridge regression). **g**, Multicollinearity between risk factors that justified the selection

of ridge regression as the adequate model. **h**, Bayesian optimization to find the best hyperparameters for ridge regression. **i**, Regression step used with the ridge regression. **j**, Regression report. **k**, The multimethod results revealed high consistency across methods using goodness-of-fit metrics (R^2 , Cohen's F_2 , MSE and RMSE) and a high coherence in the weight and ranking of risk factors of healthy aging (β estimates). In **k**, the upper panel presents the multimethod findings related to cognitive performance. The lower panel displays the functional ability data across all countries in Latin America. Risk factors: demographics, SDH, health, lifestyle, mental health symptoms and country.

were: physical activity ($\beta = -1.06, P < 0.0001$), mental health symptoms ($\beta = 0.73, P < 0.0001$), isolation ($\beta = -0.44, P < 0.0001$), falls ($\beta = 0.38, P < 0.0001$), country ($\beta = 0.33, P < 0.0001$), heart disease ($\beta = 0.29, P < 0.0001$), diabetes ($\beta = 0.22, P < 0.0001$), sex ($\beta = 0.13, P < 0.0001$), education ($\beta = 0.11, P < 0.001$), alcohol consumption ($\beta = 0.11, P < 0.0001$), smoking ($\beta = -0.1, P < 0.001$), hypertension ($\beta = 0.09, P < 0.0001$).



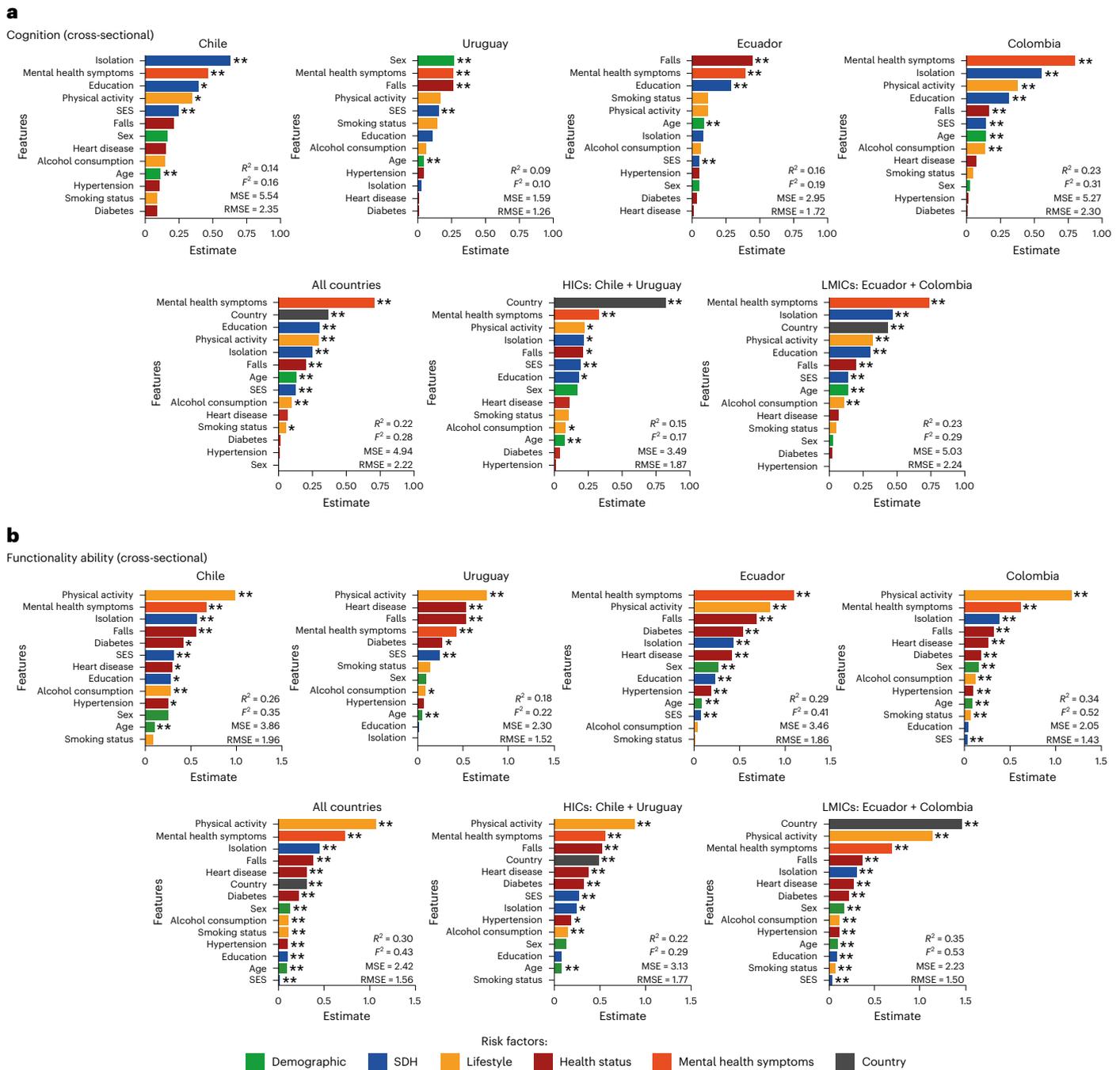


Fig. 2 | Cross-sectional results of cognition and functional ability in Latin America. a, b, Results for cognition (a) and functional ability (b) for each country, all collapsed countries, as well as LA-HICs (Chile and Uruguay) and LA-LMICs (Ecuador and Colombia). The R^2 , Cohen's F_2 , MSE and RMSE are reported for each model. The feature importance and their statistical significance are also provided

for each model. β estimates were used to assess the weight of each feature in the models. The risk factors are: demographics, SDH, lifestyle, health status (cardiometabolic factors and falls), mental health symptoms and country. The two-sided P value of a Student's t -statistic was calculated for the β values of the regression. * $P < 0.05$; ** $P < 0.01$. No asterisk means not significant.

$P < 0.001$), age ($\beta = -0.09, P < 0.0001$) and SES ($\beta = 0.01, P < 0.01$; Table 2 and Fig. 2b).

Stratification according to country income level

For LA-LMICs, the model predicting cognition was significant ($F_{1,25,769} = 535.21, P < 0.00001; R^2 = 0.23 \pm 0.0145, F_2 = 0.29, MSE = 5.03, RMSE = 2.24$). The most relevant risk factors in the model were: mental health symptoms ($\beta = 0.73, P < 0.0001$), isolation ($\beta = 0.46, P < 0.0001$), country ($\beta = 0.43, P < 0.0001$), physical activity ($\beta = -0.32, P < 0.0001$), education ($\beta = 0.30, P < 0.0001$), falls ($\beta = 0.2, P < 0.0001$), SES ($\beta = 0.13,$

$P < 0.0001$), age ($\beta = 0.13, P < 0.0001$) and alcohol consumption ($\beta = 0.08, P < 0.01$).

For LA-HICs, the model predicting cognition was also significant ($F_{1,2,340} = 28.82, P < 0.00001; R^2 = 0.15 \pm 0.0672, F_2 = 0.17, MSE = 3.49, RMSE = 1.87$) (Fig. 2a). The most relevant risk factors included: country ($\beta = 0.81, P < 0.0001$), mental health symptoms ($\beta = 0.34, P < 0.001$), physical activity ($\beta = -0.23, P < 0.05$), isolation ($\beta = 0.22, P < 0.05$), falls ($\beta = 0.21, P < 0.05$), SES ($\beta = 0.19, P < 0.0001$), education ($\beta = 0.18, P < 0.05$), sex ($\beta = 0.18, P < 0.05$), alcohol consumption ($\beta = 0.08, P < 0.05$) and age ($\beta = -0.08, P < 0.0001$).

Table 1 | Demographic and healthy aging information across studies

Country	Sex		Age		Education		Cognition (MMSE) ^a		Functional ability (Barthel)	
	Category	n	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
Demographic and general information of the cross-sectional analyses										
Chile	Female	855	72.13	8.3	1.09	0.38	11.69	3.1	6.93	2.42
	Male	446	70.51	7.4	1.18	0.56	12.11	2.69	7.83	2.11
Uruguay	Female	920	71.07	7.64	1.15	0.47	12.95	1.53	7.75	1.81
	Male	530	70.72	6.81	1.27	0.64	12.98	1.17	8.04	1.54
Colombia	Female	13,582	70.8	8.26	1.08	0.36	11.82	2.74	8.01	1.84
	Male	10,112	70.84	8.12	1.12	0.44	11.94	2.53	8.4	1.65
Ecuador	Female	2,767	71.48	8.72	1.08	0.38	11.44	2.35	6.58	2.3
	Male	2,468	71.3	8.45	1.14	0.51	11.76	2.21	7.31	2.2
Demographic and general information of the longitudinal analyses										
Costa Rica	Female	1109	59.5	3.1	1.5	0.7	10.9	1.2	8.03	1.3
	Male	642	59.5	3.2	1.5	0.7	10.6	1.3	8.33	1.2
China	Female	2194	83.1	10.6	1.0	0.2	10.0	4.4	7.6	2.9
	Male	2093	79.8	9.2	1.2	0.4	11.4	3.7	8.5	2.5

MMSE, Mini-Mental State Examination. ^aFor the longitudinal analyses, the cognition and functional ability relate to wave 2.

We compared the weight of each risk factor of cognition in the models of LA-LMICs and LA-HICs based on the effect sizes (F_2). We observed that SES, isolation, mental health symptoms, physical activity and falls have coefficient estimates higher in LMICs than in HICs. Moreover, age and sex had a more significant role in LA-HICs than in LA-LMICs. Cardiometabolic factors did not reach significant values neither in LA-LMICs nor in LA-HICs (Table 2 and Fig. 2a).

The model predicting functional ability was significant for LA-LMICs ($F_{1,25,807} = 974.02, P < 0.00001; R^2 = 0.35 \pm 0.024, F_2 = 0.53, MSE = 2.23, RMSE = 1.50$). The most relevant risk factors were: country ($\beta = 1.46, P < 0.0001$), physical activity ($\beta = -1.13, P < 0.0001$), mental health symptoms ($\beta = 0.69, P < 0.0001$), falls ($\beta = 0.36, P < 0.0001$), isolation ($\beta = 0.3, P < 0.0001$), SES ($\beta = 0.27, P < 0.0001$), heart disease ($\beta = 0.27, P < 0.0001$), diabetes ($\beta = 0.22, P < 0.01$), sex ($\beta = 0.16, P < 0.0001$), hypertension ($\beta = 0.11, P < 0.0001$), alcohol consumption ($\beta = 0.11, P < 0.0001$), age ($\beta = -0.09, P < 0.0001$), education ($\beta = -0.08, P < 0.001$), smoking status ($\beta = -0.07, P < 0.001$) and SES ($\beta = 0.04, P < 0.0001$).

The model also achieved significance in predicting functional ability for LA-HICs ($F_{1,2,340} = 44.85, P < 0.00001; R^2 = 0.22 \pm 0.061, F_2 = 0.29, MSE = 3.13, RMSE = 1.77$) (Fig. 2b). The most relevant risk factors were physical activity (lifestyle, $\beta = 0.87, P < 0.0001$), mental health symptoms ($\beta = 0.55, P < 0.0001$), falls ($\beta = 0.52, P < 0.0001$), country ($\beta = 0.49, P < 0.0001$), heart disease ($\beta = 0.37, P < 0.0001$), diabetes ($\beta = 0.33, P < 0.01$), SES ($\beta = 0.27, P < 0.0001$), isolation ($\beta = 0.24, P < 0.05$), hypertension ($\beta = 0.18, P < 0.05$), alcohol consumption ($\beta = 0.15, P < 0.001$) and age ($\beta = -0.08, P < 0.0001$).

When comparing the weight of each risk factor of functionality in LA-LMICs and LA-HICs using effect sizes (F_2), all factors were statistically significant for LA-LMICs, while smoking status was not statistically significant for LA-HICs. The most relevant risk factors on LA-LMICs compared to HICs were physical activity, mental health symptoms, country, SES, isolation, falls, heart disease, diabetes and hypertension (Table 2 and Fig. 2b).

We also run a similar group of models to predict cognition and functional ability in each country. Across countries, and consistent with regional analysis, the more relevant risk factors of cognition were mental health symptoms, SDH, physical activity and education. In LA-LMICs (Colombia and Ecuador), age was also a significant risk factor

of cognition although sex did not reach significant values. In LA-HICs (Uruguay and Chile), sex and age were significant risk factors of cognition. The more critical risk factors of functional ability in all LACs were a larger combination of mental health symptoms, SDH, lifestyle and cardiometabolic factors. While in LA-LMICs (Colombia and Ecuador), mental health symptoms, physical activity, SES and isolation had the highest scores to predict functional ability, in LA-HICs cardiometabolic factors, age and sex had a more relevant role (Extended Data Table 3).

Longitudinal comparisons between Costa Rica and China

The most relevant risk factors of cognition and functional ability were investigated using longitudinal data taken from national surveys from Costa Rica; those results were compared with longitudinal data taken from a non-LA-UMIC (China). To this end, we derived the risk factors from the first wave and the outcomes (cognition and functional ability) from the last wave of the longitudinal survey assessments. We tested and compared the independent models for Costa Rica and China.

We tested two independent models to assess the most relevant risk factors of cognition. The model predicting cognition was significant for Costa Rica ($F_{1,5,694} = 3.95, P < 0.00001; R^2 = 0.14 \pm 0.03, F_2 = 0.15, MSE = 1.51$ and $RMSE = 1.22$). The model assessing cognition for China also reached significant values ($F_{1,3,112} = 22.78, P < 0.00001; R^2 = 0.21 \pm 0.03, F_2 = 0.23, MSE = 15.50$ and $RMSE = 3.95$). For Costa Rica, the most relevant risk factors included education ($\beta = 0.51, P < 0.00001$), sex ($\beta = 0.33, P < 0.00001$), mental health symptoms ($\beta = 0.18, P < 0.001$) and age ($\beta = 0.02, P < 0.0001$). For China, the most relevant risk factors were mental health symptoms ($\beta = -0.85, P < 0.00001$), sex ($\beta = 0.78, P < 0.00001$), physical activity ($\beta = 0.77, P < 0.00001$) and age ($\beta = 0.19, P < 0.00001$). We ran an extra group of analyses to compare the weight of each significant risk factor of cognition between Costa Rica and China. Those analyses revealed that education was a stronger risk factor for Costa Rica than China. In contrast, age, sex and mental health symptoms were more relevant risk factors in China than in Costa Rica (Tables 3 and 4, Fig. 3a and Extended Data Fig. 2).

We used an independent model to assess the most relevant risk factors of functional ability in each country. The model predicting functional ability for Costa Rica reached significant values ($F_{1,5,694} = 5.67, P < 0.00001; R^2 = 0.14 \pm 0.0222, F_2 = 0.16, MSE = 2.08$ and $RMSE = 1.44$).

Table 2 | Comparisons of the risk factors of aging between LA-HICs and LA-LMICs in cross-sectional analyses using the Ridge regression models

Risk factor	Feature	LA-HICs			LA-LMICs			LA-HICs versus LA-LMICs
		β estimates	t	P	β estimates	t	P	
Comparison of risk factors of cognition between LA-HICs and LA-LMICs								
	Intercept	14.35	16.72	<0.0001	17.59	64.49	<0.0001	–
	Country	0.81	10.23	<0.0001	0.43	8.59	<0.0001	LA-HICs > LA-LMICs $P < 0.0001$
Demographics	Age	–0.08	–61.16	<0.0001	–0.13	–3.74	<0.0001	LA-HICs < LA-LMICs $P < 0.0001$
	Sex	–0.18	–1.99	NS	0.02	0.7	NS	–
SDH	Isolation	0.22	2.12	<0.05	0.46	11.13	<0.0001	LA-HICs < LA-LMICs $P < 0.0001$
	SES	0.19	4.54	<0.0001	0.13	22.53	<0.0001	LA-HICs < LA-LMICs $P < 0.0001$
	Education	0.18	2.32	<0.05	0.3	8.67	<0.0001	LA-HICs < LA-LMICs $P < 0.0001$
Health status	Diabetes	0.04	0.35	NS	0.01	0.4	NS	–
	Hypertension	0.0	0.06	NS	0	0.01	NS	–
	Heart disease	–0.11	–1.22	NS	–0.07	–1.65	NS	–
	Falls	0.21	2.53	<0.05	0.2	6.61	<0.0001	LA-HICs < LA-LMICs $P < 0.001$
Mental health	Mental health symptoms	0.34	3.45	<0.001	0.73	15.46	<0.0001	LA-HICs < LA-LMICs $P < 0.0001$
	Physical activity	–0.23	–2.41	<0.05	–0.32	–9.12	<0.0001	LA-HICs < LA-LMICs $P < 0.0001$
Lifestyle	Alcohol consumption	0.08	2.1	<0.05	0.1	3.09	<0.01	–
	Smoking status	–0.11	–1.34	NS	–0.05	–1.69	NS	–
Comparison of risk factors of functional ability across LA-HICs and LA-LMICs								
	Intercept	7.25	8.59	<0.0001	9.81	54.06	<0.0001	–
	Country	0.49	6.28	<0.0001	1.46	44.13	<0.0001	LA-HICs < LA-LMICs $P < 0.0001$
Demographics	Sex	0.13	1.46	NS	0.16	7.92	<0.0001	NS
	Age	–0.08	–63.59	<0.0001	–0.09	–389.92	<0.0001	NS
SDH	SES	0.27	6.65	<0.0001	0.04	9.0	<0.0001	LA-HICs < LA-LMICs $P < 0.0001$
	Isolation	0.24	2.29	<0.05	0.3	10.86	<0.0001	LA-HICs < LA-LMICs $P < 0.05$
	Education	0.08	1.04	NS	0.08	3.57	<0.001	NS
Health status	Diabetes	0.33	2.95	<0.01	0.22	8.71	<0.0001	LA-HICs < LA-LMICs $P < 0.0001$
	Hypertension	0.18	2.35	<0.01	0.11	5.6	<0.0001	LA-HICs < LA-LMICs $P < 0.0001$
	Heart disease	0.37	4.38	<0.0001	0.27	10.25	<0.0001	LA-HICs < LA-LMICs $P < 0.0001$
	Falls	0.52	6.34	<0.0001	0.36	18.31	<0.0001	LA-HICs < LA-LMICs $P < 0.01$
Mental health	Mental health symptoms	0.55	5.75	<0.0001	0.69	22.01	<0.0001	LA-HICs < LA-LMICs $P < 0.05$
	Physical activity	–0.87	–8.93	<0.0001	–1.13	–48.55	<0.0001	LA-HICs < LA-LMICs $P < 0.0001$
Lifestyle	Alcohol consumption	0.15	3.7	<0.001	0.11	5.19	<0.0001	NS
	Smoking status	0.0	0.02	NS	–0.07	–3.49	<0.001	NS

NS, not significant. The two-sided P value of a Student's t -test was calculated for the β estimates of the regression.

The model predicting functional ability for China also was significant ($F_{1,3,563} = 28.7, P < 0.00001; R^2 = 0.24 \pm 0.032, F_2 = 0.31, MSE = 6.30$ and $RMSE = 2.51$). The most relevant risk factors of functional ability for Costa Rica were heart disease ($\beta = 0.84, P < 0.00001$), mental health symptoms ($\beta = 0.67, P < 0.00001$), hypertension ($\beta = 0.42, P < 0.00001$), diabetes ($\beta = 0.35, P < 0.00001$), education ($\beta = 0.27, P < 0.00001$), physical activity ($\beta = 0.27, P < 0.00001$), sex ($\beta = 0.17, P < 0.0001$), alcohol consumption ($\beta = 0.15, P < 0.0001$) and age ($\beta = 0.01, P < 0.00001$). For China, the critical risk factors were physical activity ($\beta = 0.065, P < 0.0001$), sex ($\beta = 0.52, P < 0.00001$), diabetes ($B = 0.46, P < 0.0001$), mental health symptoms ($\beta = 0.44, P < 0.0001$), heart disease ($\beta = 0.32, P < 0.0001$), hypertension ($\beta = 0.32, P < 0.0001$) and age ($\beta = 0.13, P < 0.00001$; Tables 3 and 4, Fig. 3b and Extended Data Fig. 2).

We compared the countries' significant risk factors of cognition and functional ability by using the β estimates of each significant risk

factor (Fig. 3a,b, bottom). Those analyses showed higher predictive scores for education, hypertension, heart disease and mental health symptoms in Costa Rica than in China. Moreover, education and alcohol consumption were the only significant risk factors in Costa Rica. In contrast, age, sex and physical activity reached higher predictive values for China than for Costa Rica.

Imputation and complementary analyses

We ran models with individuals who had complete values in all variables measured in all countries. However, we ran the same group of analyses using the imputation-by-means method to handle missing values from one variable for the Costa Rica dataset in models of prediction of cognition and functional ability. The imputation method allows for increasing the number of observations when variables are assumed to exhibit a reduced number of outlier values while favoring

Table 3 | Comparison of risk factors of cognition between Costa Rica and China in the longitudinal analyses using the Ridge regression models

Risk factors	Feature (taken in wave 1)	Costa Rica cognition (MMSE) of wave 2			China cognition (MMSE) of wave 2		
		β estimates	t	P	β estimates	t	P
Demographics	Intercept	12.01	7.53	NS	24.07	10.32	NS
	Age	0.024	8.61	<0.0001	0.19	80.10	<0.00001
	Sex	0.33	2.66	<0.01	0.78	2.66	<0.0001
SDH	Education	0.51	7.14	<0.00001	0.26	0.85	NS
	SES	0	0	NS	0.01	0.38	NS
	Isolation	0.19	0.95	NS	0.44	0.90	NS
Health status	Diabetes	0.12	0.85	NS	0.14	0.23	NS
	Hypertension	0.03	0.32	NS	0.16	0.56	NS
	Heart disease	0.20	0.65	NS	0.15	0.38	NS
	Falls	1.85	3.29	NS	0	0	NS
Mental health symptoms	Mental health symptoms	0.16	2.10	<0.05	0.85	2.19	<0.05
Lifestyle	Physical activity	0.03	0.30	NS	0.77	2.73	<0.01
	Alcohol consumption	0.07	0.77	NS	0.10	0.33	NS
	Smoking status	0.07	0.59	NS	0.13	0.42	NS

Ten iterations of results were conducted to obtain ten β estimates for each predictor, providing the minimum variance for group comparisons, which was analyzed with a two-sided Mann–Whitney *U*-test, with Bonferroni correction.

Table 4 | Comparison of risk factors of functional ability between Costa Rica and China in the longitudinal analyses using the Ridge regression models

Risk factors	Feature (taken in wave 1)	Costa Rica functional ability (Barthel) of wave 2			China functional ability (Barthel) of wave 2		
		β estimates	t	P	β estimates	t	P
Demographics	Intercept	11.79	1.92	NS	16.12	2.01	NS
	Age	0.01	4.90	<0.00001	0.13	9.29	<0.00001
	Sex	0.17	1.09	<0.05	0.52	3.02	<0.0001
SDH	Education	0.27	3.01	<0.0001	0.15	0.82	NS
	SES	0	0	NS	0.08	0.45	NS
	Isolation	0.06	0.23	NS	0.13	0.46	NS
Health status	Diabetes	0.34	1.94	<0.0001	0.46	1.31	<0.01
	Hypertension	0.42	3.08	<0.00001	0.32	1.88	<0.0001
	Heart disease	0.84	2.16	<0.00001	0.32	1.36	<0.01
	Falls	0	0	NS	0	0	NS
Mental health symptoms	Mental health symptoms	0.67	3.99	<0.00001	0.44	1.91	<0.00001
Lifestyle	Physical activity	0.28	1.87	<0.0001	0.65	3.96	<0.00001
	Alcohol consumption	0.13	1.12	<0.05	0.17	1.00	NS
	Smoking status	0.04	0.32	NS	0.17	0.95	NS

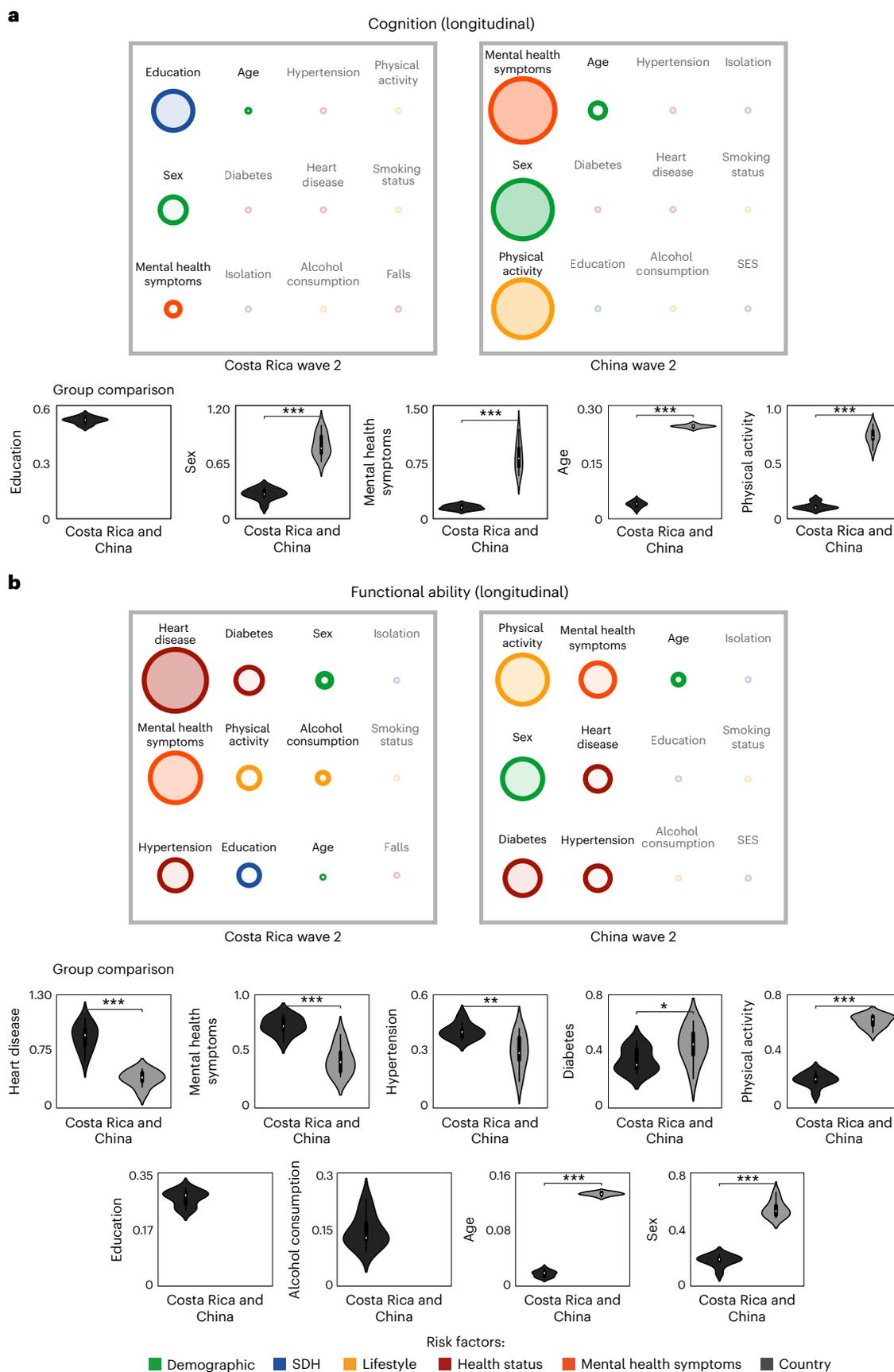
Ten iterations of results were conducted to obtain ten β estimates for each predictor, providing the minimum variance for group comparisons, which was analyzed with a two-sided Mann–Whitney *U*-test, with Bonferroni correction.

data interpretation²⁶. We implemented this procedure to maintain the same number of variables to compare models between Costa Rica and China. The results were consistent after running regression models for both cognition and functional ability with the imputation method (Extended Data Table 4).

We also performed complementary analyses to determine the independent associations between specific risk factors (sex, age, mental health status, lifestyle, health status, including cardiometabolic factors, and SDH) and healthy aging outcomes. These analyses highlighted the significant effects of each risk factor when examined independently (Extended Data Table 5).

Discussion

The present study aimed to assess the most relevant risk factors of healthy aging (cognition and functional ability) across different LAC countries. In line with our hypothesis, the results revealed a heterogeneous and distributed set of social and health disparity-related risk factors of cognition (mental health symptoms, SDH, education and physical activity) and functional ability (mental health symptoms, SDH, education, physical activity and cardiometabolic factors) across LAC countries. Such heterogeneous sets of risk factors were more accentuated in LA-LMICs compared to LA-HICs. The longitudinal study yielded similar results by comparing an LA-UMIC (Costa Rica) and a non-LA-UMIC



(China) as social and health disparity factors reached a higher predictive role in Costa Rica than in China, and a more complex group of features determined functional ability in Costa Rica, in contrast to the classical

demographic risk factors in China. Across all analyses, disparity-related social and health risk factors were more significantly associated with healthy aging than classical factors such as age and sex in LAC countries.

Fig. 3 | Longitudinal risk factors of cognition and functional ability in an LA-UMIC (Costa Rica) and a non-LA-UMIC (China). **a, b**, Longitudinal risk factors of cognition (**a**) and functional ability (**b**) were grouped into the following factors: demographics; SDH; lifestyle; health status (cardiometabolic factors and falls); mental health symptoms; and country. Features were ordered from most to least influential in the regression. The feature importance ranks in the regression model for cognition and functional ability are highlighted, accompanied by their statistical significance. Feature importance is represented by the radius of the circles and accentuated by the intensity of the color. The bottom parts of both panels show the countries' comparison analyses (violin plots) used to test

differences in the weight of significant risk factors (β estimates) of cognition and functional ability ($n = 9,258$). Ten iterations of the results were conducted to obtain ten β estimates for each risk factor, providing the minimum variance for performing group comparisons, which was analyzed with a two-sided Mann–Whitney U -test with Bonferroni correction. The specific values of the violin plots (minimum, maximum, center, 25th and 75th quartiles, inferior and superior whiskers) are provided in Extended Data Table 6 (for cognition) and Extended Data Table 7 (for functional ability). * $P < 1.00 \times 10^{-2} \leq 5.00 \times 10^{-2}$; ** $P < 1.00 \times 10^{-3} \leq 1.00 \times 10^{-2}$; *** $P < 1.00 \times 10^{-4} \leq 1.00 \times 10^{-3}$.

Regarding cognition, SES is a relevant risk factor^{1,27} having a crucial role in this study. Cognition and low SES have been linked to specific SDH, such as social exclusion, isolation and reduced social interactions^{28,29}, which are more prevalent in LAC countries³⁰, especially in older adults³¹. Mental health symptoms were identified as a top risk factor, probably due to their increased prevalence across LAC countries compared to other regions^{5,19,32–34}. Cardiometabolic factors did not reach significant scores as reported previously¹⁸. One possible interpretation could be the limited ability of the scale used to measure cognition, the Mini-Mental State Examination (MMSE), to identify executive functioning and reduced capacity to capture the early stages of cognitive decline³⁵. However, this null association could be better explained by complex interactions between risk factors in our models. This interpretation was supported by complementary analyses demonstrating the significant effects of cardiometabolic factors when analyzed separately. Future studies should systematically assess how combining different potential risk factors¹⁵ could modify the model's predictive role. Regardless, our results highlight a heterogeneous combination of risk factors impacting cognition³⁶ associated with regional inequity and specificity.

Regarding functionality, our results confirm the role of physical activity in functional ability³⁷, probably through reducing noncommunicable diseases³⁷. Mental health symptoms, which are highly prevalent in LAC countries^{5,19,32–34}, may reduce autonomy and motivation for daily activities³⁸. Falls also impacted functional ability³⁹, which may be exacerbated in LAC countries by the barriers to accessing health and social support⁴⁰. Although cardiometabolic factors are associated with functional ability^{41,42}, this association was less pronounced than with other factors. Social isolation and SES^{41,43}, usually accentuated in LAC countries⁴¹, were also relevant risk factors. Thus, a heterogeneous combination of risk factors related to regional disparities was associated with cognition and functional ability.

Classical risk factors associated with cognition and functional ability, such as age and sex, were less accentuated than those related to social and health disparities, unlike in other regions^{1,10–14,42}. Although these factors were significant when analyzed independently, their effects diminished or disappeared when combined with other risk factors. In LAC countries, there is an increased presence of SDH^{17,31}, an augmented prevalence of cardiometabolic factors¹⁸ and mental health symptoms^{5,19,32}, and barriers to healthy lifestyles^{20,21}. The combination of these disparity-related risk factors in LAC countries could attenuate the impacts of sex and age as risk factors of healthy aging. In contrast, classical risk factors may have a more substantial role in healthy aging in Europe and the United States. In those regions, there is also a more substantial control of modifiable social and health disparities, which would help diminish their effects and explain the contrasting pattern. Similarly, social and health disparities were stronger risk factors in LA-LMICs than in LA-HICs, which is consistent with the larger inequalities in the former⁹. Also, differences between Costa Rica and China confirmed this pattern despite their similar socioeconomic development (Extended Data Tables 6 and 7). In summary, heterogeneous and disparity-related factors were critically associated with cognition and functional ability in LAC countries, especially in LA-LMICs,

and presented stronger influences than other classic factors such as age and sex.

Our approach based on machine learning methods can address multicollinearity and high-dimensional data⁴⁴ and incorporate sample validation processes, thereby providing a more reliable assessment of the model's performance on unseen data than classical statistical models⁴⁴. Our modeling approach can handle complex interactions between risk factors and outcomes more effectively than standard regressions and other traditional statistical methods to assess associations between variables⁴⁵. Similarly, it can identify the top predictors without assuming a priori theoretical rankings, which are usually required when classical statistical methods are applied⁴⁴ (Extended Data Table 8).

There are limitations to our study. First, as in previous reports of SDH and cardiometabolic factors⁴⁶, data collection is based on self-reports and could be prone to bias. Our study combined self-reported data and standardized objective measures to diminish such potential bias. Still, future assessments with objective measures would be needed to confirm our results. Second, the large population might have impacted statistical significance and reduced prediction accuracy in machine learning algorithms⁴⁷. We handled these concerns using a sample size robust enough for each statistical comparison and effect size estimation; combined statistical indexes confirmed the machine learning accuracy obtained. Our study only used MMSE and the Barthel index as the primary measures of cognitive and functional outcomes in the context of healthy aging, which may not capture all aspects of healthy aging. Finally, some potential risk factors of aging, such as mental health symptoms, were assessed using a single self-reported question. Although previous studies have assessed mental health symptoms using similar procedures^{32,48}, future studies should consider using more comprehensive tools to evaluate healthy aging outcomes, mental health symptoms and other relevant factors.

While several global organizations, including the World Health Organization^{22,49} and the Alzheimer's Association^{1,24}, have called for the improvement of public health actions regarding healthy aging in LAC countries, initiatives are still limited, generic and poorly targeted to this region. The result of our study encourages public health leaders to consider the complex interactions of multiple disparity-related factors^{4,15}, including individual health-related markers and SDH. By understanding region-specific risks, policymakers can develop tailored prevention responses. Our study invites public health actions to prioritize programs to address multimodal disparities and promote mental health across the life span, mainly in older ages. Additionally, our study calls for developing national plans to increase population education and promote public resources⁵⁰ to favor social networks and healthy lifestyles, particularly stimulating physical activity. This comprehensive approach can be articulated with public health programs focused on reducing the risks associated with noncommunicable diseases, which are highly prevalent in LAC countries. Our results also invite us to develop social and health plans to handle several aging risks simultaneously rather than reacting to one factor at a time. This multifaceted strategy provides a robust foundation for informing

policies with synergistic effects across multiple conditions, optimizing resource allocation in public health and fostering healthier societies.

In conclusion, our results highlight a differential, region-specific and granular set of risk factors for cognition and functional ability in LACs. The findings reveal risk factors according to region and country, strongly influenced by the greater inequalities that exist in LAC populations. Income status and SDH, specific disease burden, health practices and ancestry-related factors may defy standard models of healthy brain aging, suggesting that the one-size-fits-all approach does not necessarily work. Tailored models should better inform local and regional public health initiatives grounded in more diverse, region-specific approaches.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-023-02495-1>.

References

- Stephan, B. C. M. et al. Prediction of dementia risk in low-income and middle-income countries (the 10/66 Study): an independent external validation of existing models. *Lancet Glob. Health* **8**, e524–e535 (2020).
- Greene, A. S. et al. Brain-phenotype models fail for individuals who defy sample stereotypes. *Nature* **609**, 109–118 (2022).
- Livingston, G. et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet* **396**, 413–446 (2020).
- Ibáñez, A., Legaz, A. & Ruiz-Adame, M. Addressing the gaps between socioeconomic disparities and biological models of dementia. *Brain* <https://doi.org/10.1093/brain/awad236> (2023).
- Nichols, E. et al. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *Lancet Public Health* **7**, e105–e125 (2022).
- Parra, M. A. et al. Dementia in Latin America: paving the way toward a regional action plan. *Alzheimers Dement.* **17**, 295–313 (2021).
- Nichols, E. et al. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* **18**, 88–106 (2019).
- Chen, Y. et al. Defining brain health: a concept analysis. *Int. J. Geriatr. Psychiatry* <https://doi.org/10.1002/gps.5564> (2021).
- Mukadam, N., Sommerlad, A., Huntley, J. & Livingston, G. Population attributable fractions for risk factors for dementia in low-income and middle-income countries: an analysis using cross-sectional survey data. *Lancet Glob. Health* **7**, e596–e603 (2019).
- Vemuri, P. et al. Vascular and amyloid pathologies are independent predictors of cognitive decline in normal elderly. *Brain* **138**, 761–771 (2015).
- Cremers, L. G. M. et al. Predicting global cognitive decline in the general population using the Disease State Index. *Front. Aging Neurosci.* **11**, 379 (2019).
- Casanova, R. et al. Investigating predictors of cognitive decline using machine learning. *J. Gerontol. B Psychol. Sci. Soc. Sci.* **75**, 733–742 (2020).
- Crimmins, E. M., Kim, J. K. & Solé-Auró, A. Gender differences in health: results from SHARE, ELSA and HRS. *Eur. J. Public Health* **21**, 81–91 (2011).
- Ritchie, K. et al. Designing prevention programmes to reduce incidence of dementia: prospective cohort study of modifiable risk factors. *BMJ* **341**, c3885 (2010).
- Ibanez, A. & Zimmer, E. R. Time to synergize mental health with brain health. *Nat. Mental Health* **1**, 441–443 (2023).
- Ibáñez, A. et al. Dementia caregiving across Latin America and the Caribbean and brain health diplomacy. *Lancet Healthy Longev.* **2**, e222–e231 (2021).
- Palloni, A. & McEniry, M. Aging and health status of elderly in Latin America and the Caribbean: preliminary findings. *J. Cross. Cult. Gerontol.* **22**, 263–285 (2007).
- Miranda, J. J. et al. Understanding the rise of cardiometabolic diseases in low- and middle-income countries. *Nat. Med.* **25**, 1667–1679 (2019).
- Kohn, R. et al. Mental health in the Americas: an overview of the treatment gap. *Rev. Panam. Salud Publica* **42**, e165 (2018).
- González, S., Lozano, Ó., Ramírez, A. & Grijalba, C. Physical activity levels among Colombian adults: inequalities by gender and socioeconomic status. *Biomedica* **34**, 447–459 (2014).
- Curcio, C. L., Vanegas, J. H., Palacio, M. C. & Corchuelo Ojeda, J. Elderly and forced displacement in Colombia. *Colomb. Med.* **50**, 52–66 (2019).
- Optimizing Brain Health Across the Life Course: WHO Position Paper* (World Health Organization, 2022); <https://www.who.int/publications/i/item/9789240054561>
- Kolappa, K., Seeher, K. & Dua, T. Brain health as a global priority. *J. Neurol. Sci.* **439**, 120326 (2022).
- Gilmore-Bykovskiy, A. L. et al. Recruitment and retention of underrepresented populations in Alzheimer's disease research: a systematic review. *Alzheimers Dement.* **5**, 751–770 (2019).
- Parra, M. A. et al. Dementia in Latin America: assessing the present and envisioning the future. *Neurology* **90**, 222–231 (2018).
- Allison, P. D. Multiple imputation for missing data: a cautionary tale. *Sociol. Methods Res.* **28**, 301–309 (2000).
- Sosa, A. L. et al. Prevalence, distribution, and impact of mild cognitive impairment in Latin America, China, and India: a 10/66 population-based study. *PLoS Med.* **9**, e1001170 (2012).
- Migeot, J., Calivar, M., Granchetti, H., Ibáñez, A. & Fittipaldi, S. Socioeconomic status impacts cognitive and socioemotional processes in healthy ageing. *Sci. Rep.* **12**, 6048 (2022).
- Evans, I. E. M., Martyr, A., Collins, R., Brayne, C. & Clare, L. Social isolation and cognitive function in later life: a systematic review and meta-analysis. *J. Alzheimers Dis.* **70**, S119–S144 (2019).
- Gao, Q. et al. Loneliness among older adults in Latin America, China, and India: prevalence, correlates and association with mortality. *Int. J. Public Health* **66**, 604449 (2021).
- Santamaria-Garcia, H. et al. The impacts of social determinants of health and cardiometabolic factors on cognitive and functional aging in Colombian underserved populations. *Geroscience* <https://doi.org/10.1007/s11357-023-00755-z> (2023).
- Syed, S. et al. Global prevalence and risk factors for mental health problems in police personnel: a systematic review and meta-analysis. *Occup. Environ. Med.* **77**, 737–747 (2020).
- Shin, J. & Cho, E. Trajectories of depressive symptoms among community-dwelling Korean older adults: findings from the Korean longitudinal study of aging (2006–2016). *BMC Psychiatry* **22**, 246 (2022).
- Wu, J. J., Wang, H.-X., Yao, W., Yan, Z. & Pei, J.-J. Late-life depression and the risk of dementia in 14 countries: a 10-year follow-up study from the Survey of Health, Ageing and Retirement in Europe. *J. Affect Disord.* **274**, 671–677 (2020).
- Baek, M. J., Kim, K., Park, Y. H. & Kim, S. The validity and reliability of the Mini-Mental State Examination-2 for detecting mild cognitive impairment and Alzheimer's disease in a Korean population. *PLoS ONE* **11**, e0163792 (2016).

36. Ibanez, A. The mind's golden cage and cognition in the wild. *Trends Cogn. Sci.* **26**, 1031–1034 (2022).
37. DiPietro, L. Physical activity in aging: changes in patterns and their relationship to health and function. *J. Gerontol. A Biol. Sci. Med. Sci.* **56 Spec No 2**, 13–22 (2001).
38. Kazama, M. et al. Early impact of depression symptoms on the decline in activities of daily living among older Japanese: Y-HALE cohort study. *Environ. Health Prev. Med.* **16**, 196–201 (2011).
39. Nagarkar, A. & Kulkarni, S. Association between daily activities and fall in older adults: an analysis of longitudinal ageing study in India (2017-18). *BMC Geriatr.* **22**, 203 (2022).
40. Kinder, F., Mehmood, S., Hodgson, H., Giannoudis, P. & Howard, A. Barriers to trauma care in south and central America: a systematic review. *Eur. J. Orthop. Surg. Traumatol.* **32**, 1163–1177 (2022).
41. Kivimäki, M. et al. Physical inactivity, cardiometabolic disease, and risk of dementia: an individual-participant meta-analysis. *BMJ* **365**, 11495 (2019).
42. Borelli, W. V., Leotti, V. B., Strelow, M. Z., Chaves, M. L. F. & Castilhos, R. M. Preventable risk factors of dementia: population attributable fractions in a Brazilian population-based study. *Lancet Reg. Health Am.* **11**, 100256 (2022).
43. Stalsberg, R. & Pedersen, A. V. Are differences in physical activity across socioeconomic groups associated with choice of physical activity variables to report? *Int. J. Environ. Res. Public Health* **15**, 922 (2018).
44. Rajula, H. S. R., Verlato, G., Manchia, M., Antonucci, N. & Fanos, V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. *Medicina* **56**, 455 (2020).
45. Obermeyer, Z. & Emanuel, E. J. Predicting the future—big data, machine learning, and clinical medicine. *N. Engl. J. Med.* **375**, 1216–1219 (2016).
46. Garcia-Cifuentes, E. et al. The role of gait speed in dementia: a secondary analysis from the SABE Colombia study. *Dement. Geriatr. Cogn. Disord.* **49**, 565–572 (2020).
47. Goldenholz, D. M., Sun, H., Ganglberger, W. & Westover, M. B. Sample size analysis for machine learning clinical validation studies. *Biomedicines* **11**, 685 (2023).
48. Santamaría-García, H. et al. The role of social cognition skills and social determinants of health in predicting symptoms of mental illness. *Transl. Psychiatry* **10**, 165 (2020).
49. Lock, S. L., Chura, L. R., Dilworth-Anderson, P. & Peterson, J. Equity across the life course matters for brain health. *Nat. Aging* **3**, 466–468 (2023).
50. Ibanez, A. et al. The power of knowledge about dementia in Latin America across health professionals working on aging. *Alzheimers Dement.* **12**, e12117 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

¹Global Brain Health Institute, University of California San Francisco, San Francisco, CA, USA. ²Center of Memory and Cognition Intellectus, Hospital Universitario San Ignacio Bogotá, San Ignacio, Colombia. ³Pontificia Universidad Javeriana (PhD Program in Neuroscience) Bogotá, San Ignacio, Colombia. ⁴Latin American Brain Health Institute, Universidad Adolfo Ibañez, Santiago de Chile, Chile. ⁵Faculty of Engineering, University of Concepción, Concepción, Chile. ⁶Cognitive Neuroscience Center, Universidad de San Andrés and Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina. ⁷Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. ⁸Department of Medicine, Division of Infectious Diseases, Weill Cornell Medicine, New York, NY, USA. ⁹Memory and Aging Center, University California San Francisco, San Francisco, CA, USA. ¹⁰Centro de Excelencia en Enfermedades Crónicas, Universidad Peruana Cayetano Heredia, Lima, Peru. ¹¹Department of Medicine, School of Medicine, Universidad Peruana Cayetano Heredia, Lima, Peru. ¹²Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK. ¹³The George Institute for Global Health, University of New South Wales, Sydney, New South Wales, Australia. ¹⁴Trinity College Dublin, The University of Dublin, Dublin, Ireland. ¹⁵These authors contributed equally: Hernando Santamaría-García, Agustín Sainz-Ballesteros.

✉ e-mail: hernando.santamaria@gbhi.org; agustin.ibanez@gbhi.org

Methods

Cross-sectional analyses of participants

Participants ($n = 31,680$; female = 16,074; mean age = 70.81 years, s.d. = 8.22) were recruited from the national surveys on health, well-being and aging performed in four LAC countries: Chile (2016, $n = 1,301$; females = 689; mean age = 71.83 years, s.d. = 8.29)⁵¹; Uruguay (2011, $n = 1,450$; females = 798; mean age = 70.75 years, s.d. = 7.40)⁵¹; Ecuador (2012, $n = 5,235$; females = 1,530; mean age = 70.09 years, s.d. = 7.82)⁵²; and Colombia (2015, $n = 23,694$; females = 1,357; mean age = 70.79 years, s.d. = 8.26)⁵³. The interviews from all national surveys were performed face to face and participants were selected following a probabilistic, clustered, stratified and multistage design in each country. We only included individuals with no diagnosis of dementia, as assessed during an initial screening (Fig. 1a–e). The databases of national surveys from all countries are open and were obtained according to the established procedures for each country.

Longitudinal analyses of participants

We analyzed data from the national aging study of an LA-UMIC: Costa Rica (Costa Rican Study on Longevity and Healthy Aging (CRELES), $n = 5,694$; ref. 54). This survey includes two waves of information on the risk factors of aging (wave 1 = 2012, wave 2 = 2015–2016). Longitudinal information from a non-LA-UMIC in China (Chinese Longitudinal Healthy Longevity Survey Series, $n = 3,564$; ref. 55) was used as control information. This survey includes two waves of information (wave 1 = 2011, wave 2 = 2014–2015). In this longitudinal approach, we assessed the role of factors taken in the first wave in predicting the scores of measures of cognition and functional ability of wave 2. The final population from both the Costa Rica and China datasets totaled 9,258 participants (Fig. 1b).

The interviews from all national surveys were performed face to face and participants were selected following a probabilistic, clustered, stratified, multistage design in each country. Information collected in all countries, in the context of national aging surveys, was representative of each population (Colombia, Ecuador⁵², Uruguay⁵¹, Chile⁵¹, Costa Rica⁵⁴ and China⁵⁵). The instruments, harmonization and computational approaches of this study followed the same parameters as described in the cross-sectional analyses.

Instruments

Risk factors. The final datasets from the surveys in all LAC countries consisted of five harmonized risk factors (demographics, SDH, health, lifestyle and mental health symptoms).

Demographics. Demographics included age (years), sex (women and men only) and years of education attained. No questions were asked regarding gender identity aspects in any country.

SDH. To assess this risk factor, we included information previously associated with brain health⁵⁰, including measures of isolation (whether the participant lives alone or not), SES (which was measured with a composite index that included income and housing conditions) and educational level (which has three levels: low (elementary school and middle school), medium (high school) and high (bachelor's and postgraduate degrees)).

Health status. This included health factors previously associated with brain health such as cardiometabolic factors (hypertension, diabetes and cardiovascular risks) and falls.

Lifestyle. This included smoking behavior, alcohol consumption, and physical activity.

Mental health symptoms. We collected information on the presence of depression, anxiety and other mental health symptoms. Those symptoms were categorized as binomial variables.

Outcome variables. Cognition. Cognition was assessed in national surveys from LAC countries with the shortened version of the Folstein MMSE. Following previous procedures^{56,57}, we used the abbreviated MMSE, which has 19 questions and items. A cutoff of 13 points was used in each survey to determine if the participant had cognitive decline⁵⁷.

Functional ability. The Barthel index⁵⁸ assesses difficulties in ADLs, including basic activities of daily life such as self-maintenance skills (dressing, bathing, grooming, toilet use, and bowel and bladder continence) and mobility skills⁵⁶. We analyzed information from a group of activities collected via participants' interviews and harmonized across countries referring to daily living functioning. The scores were built in each database by computing a ratio of positive responses according to previous studies⁵⁸. For a further description of specific questions to assess the risk factors and the healthy aging outcomes used in each country, see the Supplementary Information.

Data harmonization of variables across countries

We ran harmonization procedures⁵⁹ for some variables because the answer options of some measures were different between countries. Similarly, missing values were observed in some datasets. All variables used in the analyses of this study had complete values in at least 80% of cases. Thus, we collected 28,109 participants with complete values in the variables of the analyses to assess risk factors of cognition and a population of 27,991 participants for functional ability (Fig. 1a–e).

Harmonization procedures of variables across countries in the cross-sectional analyses. MMSE. The test contained the following questions: (1) participants should provide the date. This question scores one point for each of the following correct answers: day of the week, day of the month, month and year; (2) the interviewer names three objects and the participants must remember them. This question scores one point for each object recalled; (3) the interviewer gives the individual a piece of paper to perform three actions. Performing each action correctly scores one point; the actions are: taking a piece of paper with the right hand, folding the paper in half with both hands and placing the paper on the lap; (4) the participant remembers the name of the previous three objects again. This question is scored the same as question 2; (5) participants copy a drawing. One point is scored if the drawing is done correctly. The final scale is achieved using standardized scores of the abbreviated MMSE except a question about remembering a sequence of five numbers, which scored five points. This question was removed because it was evaluated in different ways across the databases and no complete information was recruited to harmonize the scores across sites.

SDH harmonization. Educational level was set to three levels: low (elementary and middle school); middle (high school); and high (bachelor's and postgraduate degrees). SES was calculated on the proportion of available housing resources and services.

Lifestyle harmonization. Alcohol consumption was set to four levels: never (participant never consumes alcohol); normal (participant consumes alcohol less than one day per week); overdrinking (participant consumes alcohol two to six times per week); severe (participant consumes alcohol daily). All other variables were dichotomous (yes/no), except for age, which was collected by all countries as a discrete variable. Non-common features were dropped.

Harmonization procedures of variables across countries in the longitudinal analyses. The MMSE information was computed on 13 points in Costa Rica and China because the drawing test was not conducted in Costa Rica. We also ran imputation by means to impute the missing values of one feature (falls) in the Costa Rica

dataset. For imputation, we used the mean scores in the aforementioned variables from other HICs (Chile and Uruguay). To control for the effects of imputation, additional analyses without imputation were also performed (Extended Data Table 4).

Statistical analysis

No statistical method was used to predetermine sample size because we used the complete database of the national aging surveys from different LAC countries. All datasets from the national aging surveys from LAC countries in cross-sectional and longitudinal analyses included representative samples from each country.

Cross-validation for hyperparameter tuning. First, we conducted the best search for the most appropriate set of hyperparameters using cross-validation to obtain the best possible generalization results^{60,61}. For each model, we implemented a Bayesian optimization approach for hyperparameter tuning, with cross-validation = 3 and ten iterations, on 75% of the data (training dataset), and evaluated the results on a validation dataset (25% of the data). Then, we randomly divided the data on a new training sample (75%) and testing sets (25%) into $k = 10$ folds; each subset was used for training $k - 1$ times and validation, using the best hyperparameters obtained from the Bayesian optimization in the previous step. Finally, we obtained the margins of errors, the β estimates and t -tests by assessing the mean of $k = 10$ iterations. This facilitated the identification of optimal regularization strength and additional hyperparameters for the ridge regression model.

Solver selection. We optimized the regularization strength, the maximum number of iterations and the solver. We followed scikit-learn's implementation for ridge, which allows testing and comparing different solvers, including: (1) auto: it chooses the solver automatically based on the data. This solver reached 1,000 maximum iterations and an alpha of 0.0001; (2) singular value decomposition involving features vectors⁶². This approach had 10,000 maximum iterations and an alpha of 0.001; (3) Cholesky, a standard linear function that allows obtaining a closed-form solution (Extended Data Table 8). The maximum number of iterations with this approach was 100,000 and the alpha was 0.01; (4) *sqr*, a dedicated regularized least-square routine, and *sag*, a stochastic average gradient descent⁶³, which uses a maximum number of iterations of 1,000,000; and (5) *sparse_cg*, a conjugate gradient solver involving *sag* and *saga* that follows the stochastic average gradient descent process and an optimized version of this approach, respectively⁶³ (Extended Data Table 2).

Moreover, we used different regularization and solver selection processes for elastic net and LASSO. For elastic net, we used two solvers including (1) *cyclic*, which repeats features sequentially by default with a maximum number of iterations of 1,000 and an alpha of 0.0001; and (2) *random*, which updates a random coefficient in every iteration, with a maximum number of iterations of 10,000 and an alpha of 0.001. Moreover, different L1 ratios (ranging from 0.5 to 0.9) were used, and the iterations and alpha were fitted according to this ratio. For LASSO, we implemented two solvers, including: (1) *cyclic* with a maximum number of iterations of 1,000 and an alpha of 0.0001; and (2) *random* with a maximum number of iterations of 10,000 and an alpha of 0.001 (Extended Data Tables 2 and 8). The score we used for optimization was the square loss, which is used in ridge regressions as default⁶⁴. This method imposes greater penalties on larger errors compared to smaller ones⁶⁴.

The closed-form solution for ridge regression, which includes a regularization parameter a to control the amount of shrinkage, can be expressed as follows:

$$\min_w \|Xw - y\|^{(2)} + a\|w\|^{(2)}$$

$$\min_{|w|} \|Xw - y\|_2^2 + a\|w\|_2^2$$

where w represents the coefficients of the regression model; X is the design matrix, where each row represents an observation and each column represents a risk factor variable; y is the target variable; and a is the regularization parameter (also known as the ridge parameter), which controls the amount of shrinkage: the larger the value of a , the greater the amount of shrinkage.

The parameters of a used for our analyses are provided in Extended Data Table 8.

Ridge regression method. We assessed the correlation and collinearity between risk factors (as revealed by the variance inflation factor, scores above 5). Those analyses showed multicollinearity (Fig. 1g and Extended Data Table 1) and correlations among variables (Extended Data Fig. 1). Thus, we chose ridge regression models to assess the more relevant risk factors of healthy aging. This method helps to (1) handle the risk factors' multicollinearity and correlations; (2) reduce overfitting because it introduces a regularization term to penalize large coefficients, thus improving the model's generalization⁶⁴; (3) improve model stability because it shrinks the coefficients toward zero, effectively reducing the impact of noise or irrelevant features on the model's predictions⁶⁵; (4) improves model interpretability because it shrinks the coefficients, which helps interpret the importance of different risk factor variables selected based on previous theoretical accounts⁶⁶; and (5) addresses variance and interactions because it is a recommended process to tackle regression challenges in the presence of multidimensionality and complex interactions between risk factors⁶⁷.

The ridge regression model using cross-sectional data can be expressed as:

$$\min(\beta) \|Y - X\beta\|^2 + \lambda\|\beta\|^2$$

where Y is the outcome variable, X is the matrix of the risk factor variables, β is the vector of the coefficients and λ is the regularization strength.

The ridge regression model for longitudinal data can be expressed as:

$$\hat{y} = X(\beta) + \varepsilon, \text{ subject to } \|\beta\|_2^2 \leq t$$

where \hat{y} represents the predicted outcome, X is the matrix of the risk factors, β is the vector of the coefficients, ε is the error term, $\|\beta\|_2^2$ is the L2-norm of β (sum of squared coefficients) and t is the regularization parameter.

Multimethod analyses. To confirm the robustness of our approach, we implemented multimethod confirmatory analyses (Fig. 1f,g,k). This comparative assessment involved examining the outcomes of ridge regression in conjunction with linear regression, elastic net and LASSO techniques. Linear regression was used to fit a predictive model to the observed data. It calculates the strength of the relationship between risk factors and outcome variables and helps to determine whether some explanatory variables may have no linear relationship with the outcomes. Elastic net combines the advantages of ridge and LASSO regression, incorporating both L1 and L2 regularization⁶⁸. Furthermore, it balances the benefits of both techniques, providing a compromise between the sparsity of LASSO and the stability of ridge regression. This can result in better prediction accuracy and improved model interpretability⁶⁸. LASSO regressions encourage sparsity in the model by driving some coefficients to zero, leading to more straightforward and more interpretable models, and allowing for automatic feature selection to deal with many risk factors⁶⁹.

Goodness-of-fit and weight of risk factor parameters. We used different parameters to assess the goodness of fit in the studies, including (1) F -statistic (F) for the regression coefficients of the models; (2) R^2 ,

a measure of how close the data points correspond to the fitted line and as the coefficient of determination for the regression models; (3) Cohen's F_2 to assess the effect sizes of the regression models and risk factors; (4) MSE, an estimator of the average of the squared difference between estimated and actual values and to assess the goodness of fit of the regression models; (5) the RMSE as a measure of the standard deviation of the residuals (prediction errors) and to assess the distance between the regression line data points; and (5) β estimates to assess the weight of a factor in a regression model. Different linear regression models were used to identify the best goodness-of-fit parameters across the analyses. The multimethod procedure confirmed the consistency of the results and suggested that linear approaches were the best choice. Other additional nonlinear methods are out of the scope of this study.

Cross-sectional analyses. We ran two independent ridge regression models to assess the risk factors of healthy aging: one model was run to assess the risk factors of cognition (MMSE scores as the outcome), and the other was implemented to assess functional ability (Barthel scores as the outcome) risk factors across all LAC countries. Risk factors including demographics, SDH, health status, lifestyle and mental health symptoms were included as risk factors in each model. Ridge regression models were first run for cognition and functionality across all LAC countries (Colombia, Ecuador, Chile and Uruguay). Second, the regression models were run to group LAC countries according to their income level: HICs (Chile and Uruguay) and LMICs (Colombia and Ecuador). Third, independent regression models were run for each country. We evaluated the regression models by reporting R^2 , 99% of confidence intervals, feature significance and β estimates. Each model considers the F value and the F_2 values to assess the effect sizes ($F_2 \geq 0.02$, $F_2 \geq 0.15$ and $F_2 \geq 0.35$ representing small, medium and large effect sizes, respectively⁷⁰).

Longitudinal analyses. The longitudinal evolution of outcomes (cognition and functional ability in two different moments for Costa Rica: wave 1 (2012) and wave 2 (2016); two for China: wave 1 (2011) and wave 2 (2014)). We computed the most relevant risk factors of cognition and functional ability in the last wave for Costa Rica and China using the risk factors of wave 1. Independent ridge regression models comprised the best hyperparameters using the scheme described in the cross-sectional analyses. All models and statistical analyses were run using Python v.3.9.13.

Ethics and inclusion statement

This work involved a collaboration between scientists in multiple countries including Argentina, Chile, Colombia, Ireland, Peru and the United States. Contributors from all sites are included as coauthors or in acknowledgements according to their contributions. Researchers residing in Latin American countries have been involved in study design, study implementation, methodological procedure, and writing and reviewing processes. The current research is locally relevant due to the high prevalence of cognitive decline in LAC countries. Roles and responsibilities were agreed among collaborators ahead of the research. Local ethics committees approved all research involving human participants. To prevent any stigmatization, all identifying information has been removed to preserve the privacy of individuals. Each country included in this study have retained ownership of all human material shared for research purposes.

We endorse the Nature Portfolio guidance on LMIC authorship and inclusion. Authorship was based on the intellectual contribution, commitment and involvement of each researcher in this study. We included authors born in LMICs and other underrepresented countries in this study. This study holds local relevance for each investigated country by presenting disaggregated findings, thereby offering country-specific risk factors of healthy aging. The selection of variables was informed

by previous research and in accordance with established guidelines for global aging studies.

Ethical approval

All methods were carried out in accordance with relevant guidelines and regulations provided by the Declaration of Helsinki (2013). Data for all countries including LMICs were collected via in-person interviews implemented in the context of national aging surveys taken in each country.

In each country, participants gave informed consent, which was approved by the respective ethics committees. Data collection and analysis posed no risks concerning stigmatization, incrimination, discrimination, animal welfare, environmental, health, safety, security or personal concerns. No transfer of biological materials, cultural artifacts or traditional knowledge occurred. The authors reviewed pertinent studies from all seven countries while preparing the manuscript. The Pontificia Universidad Javeriana's ethical institutional committee in Bogotá, Colombia, approved all experimental protocols used in the analyses of this study (no. FM-773-2021).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data included in this study in the cross-sectional and longitudinal analyses were collected in the context of the national aging surveys of five LAC countries (Ecuador, Colombia, Chile, Uruguay and Costa Rica) and from China. Data from the national aging survey from Ecuador included in the study are publicly available and can be accessed after providing researchers information at <https://www.ecuadorencifras.gob.ec/encuesta-de-salud-bienestar-del-adulto-mayor/>. Data from the national aging survey of Colombia (National Study of Health, Well-being, and Aging (SABE Colombia) 2015) can be accessed after filling a registration form on the Web page of the Ministry of Health and Social Protection (el Ministerio de Salud y de Protección Social) in Colombia. This procedure lasts around 10 days and can be made at this <https://www.datos.gov.co>. Data from Chile and Uruguay included in this study were part of the SABE-Survey on Health, Well-being, and Aging in Latin America and the Caribbean, 2000 (Inter-university Consortium for Political and Social Research 3546). Data from this study are publicly available after providing researcher information at <https://www.icpsr.umich.edu/web/NACDA/studies/3546/versions/V1>. Data from Costa Rica included in this study were taken from CRELES. Data from this study are publicly available after providing researcher information at <http://creles-download.demog.berkeley.edu/CRdata.pl>. Data from the national aging survey of China (China Health and Retirement Longitudinal Study) can be accessed after registration at <https://charls.pku.edu.cn/en/>. Access to the raw data from all databases described in this section lasts around 10 days. All authors had access to the raw data. All software used in this study and its versions are specified on the `conda_env.yml` file in the GitHub repository.

Code availability

All code for the data analysis associated with the manuscript is available for download from GitHub at <https://github.com/Al-BrainLat-team/Heterogeneous-risk-factors-LAC>.

References

- Albala, C. et al. The health, well-being, and aging ("SABE") survey: methodology applied and profile of the study population [article in Spanish]. *Rev. Panam. Salud Publica* **17**, 307–322 (2005).
- Freire, W. et al. *Encuesta Nacional de Salud, Bienestar y Envejecimiento SABE I Ecuador 2009–2010* (Ministerio de Salud Pública/Universidad San Francisco de Quito, 2010).

53. Gomez, F., Corchuelo, J., Curcio, C.-L., Calzada, M.-T. & Mendez, F. SABE Colombia: Survey on Health, Well-Being, and Aging in Colombia—study design and protocol. *Curr. Gerontol. Geriatr. Res.* **2016**, 7910205 (2016).
54. Rosero-Bixby, L., Dow, W. H., Brenes, G., Gu, D. & Dupre, M. Costa Rican longevity and healthy aging study. In *Encyclopedia of Gerontology and Population Aging* (eds Gu, D. & Dupre, M.) (Springer, 2019).
55. Yi, Z., Gu, D., Poston, D. L., & Vlosky, D. A. (Eds.) *Healthy Longevity in China: Demographic, Socioeconomic, and Psychological Dimensions* (Springer Science & Business Media, 2008).
56. Garcia-Cifuentes, E. et al. Muscular function as an alternative to identify cognitive impairment: a secondary analysis from SABE Colombia. *Front. Neurol.* **13**, 695253 (2022).
57. Icaza, M. G. & Albala, C. *Minimal State Examinations (MMSE) del Estudio de Demencia en Chile: Análisis Estadístico* (Organización Panamericana de la Salud, 1999).
58. Mahoney, F. & Barthel, D. W. Functional evaluation: the Barthel index. *Md. State Med. J.* **14**, 61–65 (1965).
59. Maito, M. A. et al. Classification of Alzheimer's disease and frontotemporal dementia using routine clinical and cognitive measures across multicentric underrepresented samples: a cross-sectional study. *Lancet Reg. Health Am.* **17**, 100387 (2023).
60. Santamaría-García, H. et al. Uncovering social-contextual and individual mental health factors associated with violence via computational inference. *Patterns* **2**, 100176 (2021).
61. Donnelly-Kehoe, P. A. et al. Robust automated computational approach for classifying frontotemporal neurodegeneration: multimodal/multicenter neuroimaging. *Alzheimers Dement* **11**, 588–598 (2019).
62. Banerjee, S. & Roy, A. *Linear Algebra and Matrix Analysis for Statistics* Vol. 181 (CRC, 2014).
63. Bottou, L. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010: 19th International Conference on Computational Statistics* (eds Lechevallier, Y., Saporta, G.) (Physica-Verlag HD, 2010).
64. Tsigler, A. & Bartlett, P. L. Benign overfitting in ridge regression. *J. Mach. Learn. Res.* **24**, 1–76 (2023).
65. Moreno-Salinas, D., Moreno, R., Pereira, A., Aranda, J. & de la Cruz, J. M. Modelling of a surface marine vehicle with kernel ridge regression confidence machine. *Appl. Soft. Comput.* **76**, 237–250 (2019).
66. Abonazel, M. R. & Taha, I. M. Beta ridge regression estimators: simulation and application. *Commun. Stat. Simul. Comput.* <https://doi.org/10.1080/03610918.2021.1960373> (2021).
67. Hoerl, A. E. & Kennard, R. W. Ridge regression—biased estimation for nonorthogonal problems. *Technometrics* **12**, 55 (1970).
68. De Mol, C., De Vito, E. & Rosasco, L. Elastic-net regularization in learning theory. *J. Complex.* **25**, 201–230 (2009).
69. Li, Y., Nan, B. & Zhu, J. Multivariate sparse group lasso for the multivariate multiple linear regression with an arbitrary group structure. *Biometrics* **71**, 354–363 (2015).
70. Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D. & Mermelstein, R. J. A practical guide to calculating Cohen's f^2 , a measure of local effect size, from PROC MIXED. *Front. Psychol.* **3**, 111 (2012).

Acknowledgements

We thank the national aging survey teams from Chile, Uruguay, Colombia, Ecuador, Costa Rica and China for providing the data for this study. The manuscript appropriately cites lead authors and teams when necessary. A.I. is partially supported by grants from the Agencia Nacional de Investigación y Desarrollo (ANID)/Fondo Nacional de Desarrollo Científico y Tecnológico Regular (nos. 1210195, 1210176 and 1220995); ANID/Fondo de Financiamiento de Centros de Investigación en Áreas Prioritarias (FONDAP) (no. 15150012); ANID/Programa de Investigación Asociativa/Anillos de Investigación en Ciencia y Tecnología (ACT210096); Fondo de Fomento al Desarrollo Científico y Tecnológico (nos. ID2010152 and ID2210029); ANID/FONDAP (no. 15150012); Takeda Pharmaceuticals (no. CW2680521); and the Multi-partner Consortium to Expand Dementia Research in Latin America (ReDLat), which is supported by the Fogarty International Center and the National Institutes of Health, the National Institutes of Aging (nos. R01 AG057234, R01 AG075775, R01 AG21051 and CARDS-NIH), Alzheimer's Association (no. SG-20-725707), Rainwater Charitable Foundation's Tau Consortium, the Bluefield Project to Cure Frontotemporal Dementia and the Global Brain Health Institute. The contents of this publication are solely the responsibility of the authors and do not represent the official views of these institutions. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

A.I. and H.S.-G. developed the study concept and the study design. H.S.-G., H.H., S.M. and M.M. performed the testing and curated the data. H.H., S.M., M.M. and H.S.-G. performed the data analysis under the supervision of A.I. H.S.-G., A.S.-B. and A.I. interpreted the results and drafted the manuscript. C.O.-R., M.C., V.V., J.J.M., and B.L. provided critical revisions on analyses and discussion of the manuscript. All authors provided critical revisions, participated sufficiently in the work and approved the final version of the manuscript for submission.

Competing interests

The authors declare no competing interests.

Additional information

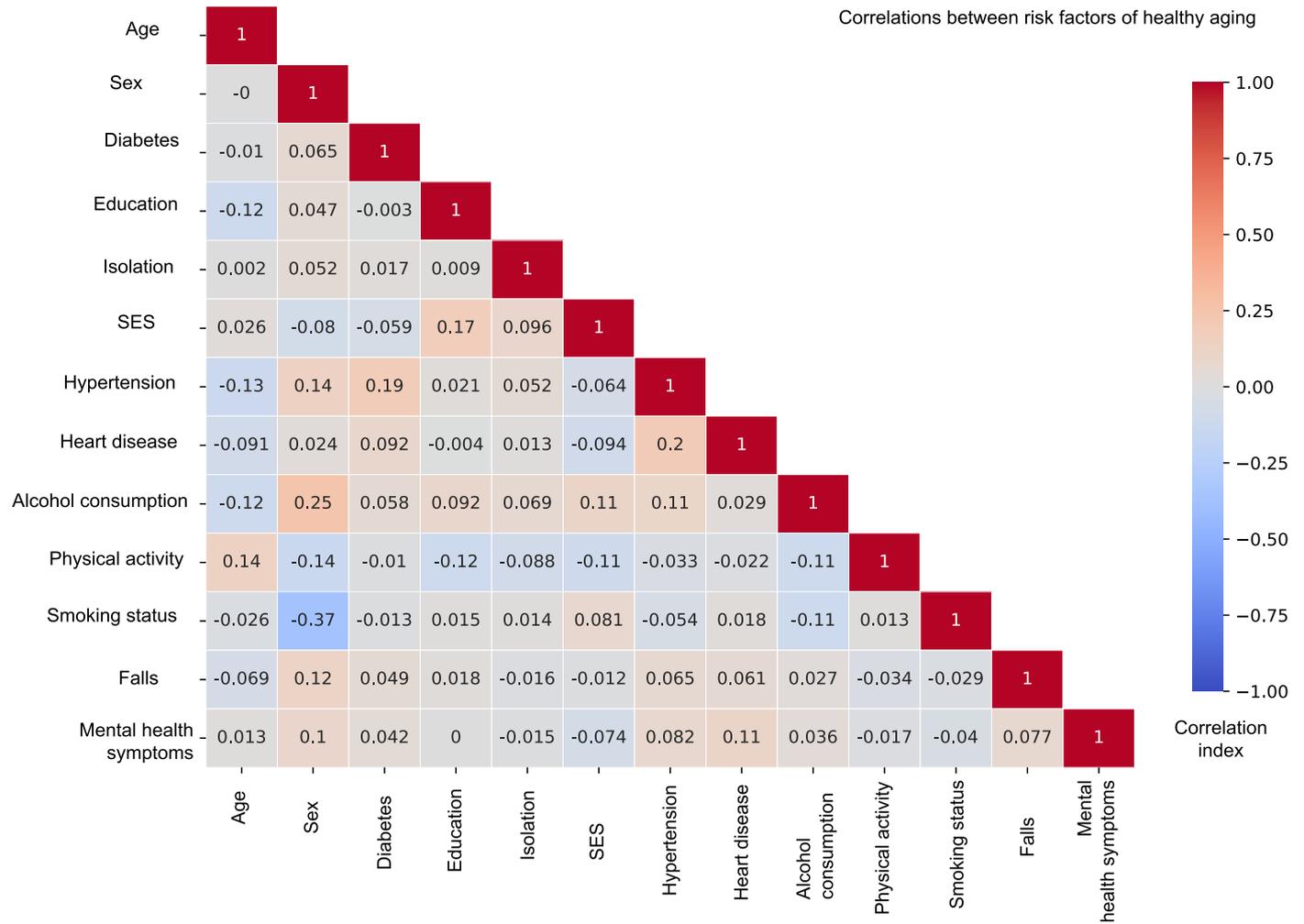
Extended data is available for this paper at <https://doi.org/10.1038/s41591-023-02495-1>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41591-023-02495-1>.

Correspondence and requests for materials should be addressed to Hernando Santamaria-Garcia or Agustin Ibanez.

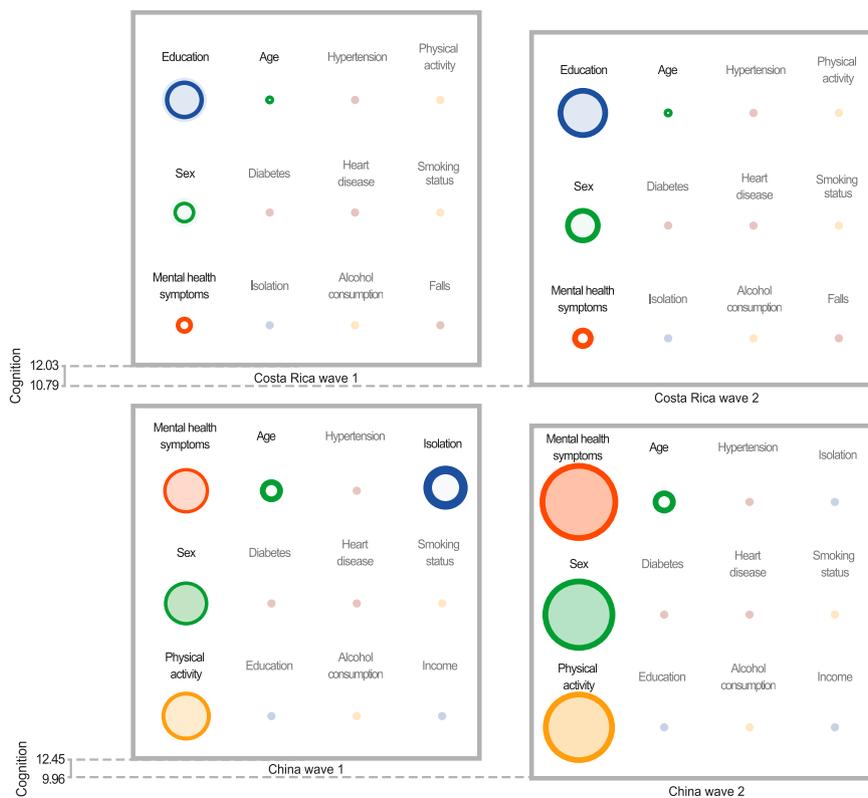
Peer review information *Nature Medicine* thanks Fei Wang, Raphael Machado Castilhos and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Ming Yang, in collaboration with the *Nature Medicine* team.

Reprints and permissions information is available at www.nature.com/reprints.

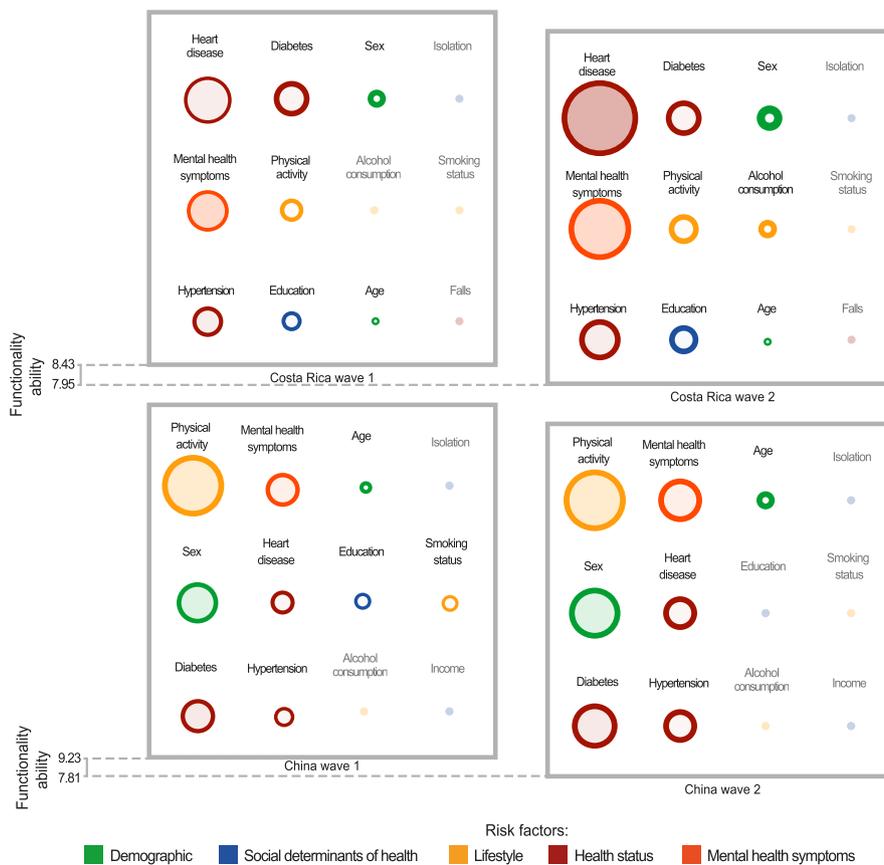


Extended Data Fig. 1 | Correlations between factors and outcomes associated with healthy aging. Matrix of correlation between factors and outcomes (cognition and functional ability) of healthy aging.

A. Cognition (longitudinal)



B. Functionality ability (longitudinal)



Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Longitudinal results of factors associated with cognition and functional ability. Longitudinal predictors of cognition and functional ability. Predictors of each wave of cognition (MMSE scores, panel A) and functional ability (Barthel scores, panel B) are shown for the LA-UMIC (Costa Rica) and the Asian UMIC (China). All panels demonstrated the significance of the predictors assessed in the initial wave for the outcomes evaluated in each

wave (wave 1 and wave 2). The risk factors were grouped into fivefold categories: demographics, social determinants of health, lifestyle, health status, mental health symptoms, and country. The features were ordered from most to least influential in the regression. The feature importance is represented by the radius of the circles and accentuated by the intensity of the color.

Extended Data Table 1 | Collinearity test between factors associated with healthy aging

Feature	Variance Inflation Factor (VIF)
Age	56.04
Sex	11.59
Education	8.17
Social isolation	10.16
Socioeconomic status	10.44
Diabetes	25.74
Hypertension	10.77
Heart disease	2.78
Falls	14.18
Mental health symptoms	38.43
Physical activity	20.53
Smoking status	10.86
Alcohol consumption	6.54

Extended Data Table 2 | Ridge and elastic net regularization parameters in models used to assess factors associated with healthy aging

Ridge			
Solver Selection		Maximum iterations	Alpha
Auto: chooses the solver automatically based on the type of data		1000	0.0001
SVD: uses a Singular Value Decomposition of data to compute the Ridge coefficients.		10000	0.001
Cholesky: uses a standard solver for a linear equation to obtain a closed-form solution.		100000	0.01
LSQR: uses a dedicated regularized least-squares routine		1000000	
Sparse_cg: uses a conjugate gradient solver			
sag: uses a Stochastic Average Gradient descent			
Saga: uses an improved unbiased version of Stochastic Average Gradient descent			
The score we used for optimization was square loss (Ridge default).			
Elastic net			
Solver Selection	l1 ratio	Maximum iterations	Alpha
Cyclic: repeats features sequentially by default	0.5	1000	0.0001
random: updates a random coefficient in every iteration.	0.6	10000	0.001
	0.7	100000	0.01
	0.8	1000000	
	0.9		
Lasso			
Solver Selection		Maximum iterations	Alpha
cyclic: repeats features sequentially by default		1000	0.0001
random: updates a random coefficient in every iteration.		10000	0.001
		100000	0.01
		1000000	

Extended Data Table 3 | Predictors of healthy aging outcomes in each LAC. Ridge regression models of factors associated with healthy aging in each LAC

Cognition	Colombia			Ecuador			Uruguay			Chile		
	(β) Estimates	t value	p value	(β) Estimates	t value	p value	(β) Estimates	t value	p value	(β) Estimates	t value	p value
intercept	18.74	74.78	P < 0.0001	16.24	30.35	P < 0.0001	14.28	19.31	P < 0.0001	16.38	9.79	P < 0.0001
Age	-0.14	-377.52	P < 0.0001	-0.08	-98.52	P < 0.0001	-0.04	-36.19	P < 0.0001	-0.11	-48.59	P < 0.0001
Sex	0.02	0.74	ns	-0.05	-0.65	ns	-0.26	-3.09	0.001	-0.16	-1.02	ns
Diabetes	0.01	0.23	ns	0.03	0.37	ns	0	0.02	ns	0.08	0.4	ns
Education	0.31	7.85	P < 0.0001	0.29	4.5	P < 0.0001	0.11	1.56	ns	0.39	2.45	P < 0.01
Social isolation	0.55	10.68	P < 0.0001	-0.08	-0.77	ns	-0.02	-0.28	ns	0.63	2.94	P < 0.01
SES	0.14	21.64	P < 0.0001	0.05	3.73	P < 0.001	0.15	3.78	0.001	0.24	3	P < 0.01
Hypertension	-0.01	-0.39	ns	0.05	0.76	ns	0.04	0.5	ns	-0.11	-0.72	ns
Heart disease	-0.07	-1.7	ns	-0.01	-0.1	ns	-0.01	-0.06	ns	-0.15	-0.96	ns
Alcohol consumption	0.14	3.78	P < 0.001	-0.06	-0.93	ns	0.06	1.84	ns	0.14	1.57	ns
Physical activity	-0.38	-9.69	P < 0.0001	-0.12	-1.71	ns	-0.17	-1.77	ns	-0.34	-1.99	P < 0.05
Smoking status	-0.05	-1.57	ns	-0.12	-1.71	ns	-0.14	-1.76	ns	-0.09	-0.57	ns
Falls	0.17	5.18	P < 0.0001	0.45	6.82	P < 0.0001	0.26	3.28	0.01	0.21	1.36	ns
Mental symptoms	0.8	15.25	P < 0.0001	0.4	4.23	P < 0.0001	0.26	2.66	0.01	0.46	2.69	P < 0.01
Functional Ability	(β) Estimates	t value	p value	(β) Estimates	t value	p value	(β) Estimates	t value	p value	(β) Estimates	t value	p value
intercept	13.19	84.48	P < 0.00001	9.06	15.94	P < 0.00001	7.49	8.05	0.00001	7.75	5.4	P < 0.0001
Age	-0.09	-397.63	P < 0.00001	-0.09	-101.48	P < 0.00001	-0.05	-33.07	0.00001	-0.1	-53.08	P < 0.0001
Sex	0.15	7.36	P < 0.00001	0.28	3.53	P < 0.0001	-0.09	-0.85	ns	0.24	1.76	ns
Diabetes	0.18	7.13	P < 0.00001	0.54	5.8	P < 0.00001	0.27	2	P < 0.05	0.42	2.38	P < 0.05
Education	0.05	1.87	ns	0.23	3.31	P < 0.0001	-0.02	-0.18	ns	0.28	2.13	P < 0.05
Isolation	0.39	12.11	P < 0.00001	-0.44	-3.91	P < 0.0001	0	-0.01	ns	0.58	3.21	P < 0.001
SES	0.03	7.28	P < 0.00001	0.08	5.36	P < 0.00001	0.24	4.79	0.00001	0.31	4.57	P < 0.0001
Hypertension	0.09	4.82	P < 0.00001	0.19	2.72	P < 0.01	0.07	0.71	ns	0.24	1.98	P < 0.05
Heart disease	0.26	9.52	P < 0.00001	0.43	4.29	P < 0.00001	0.53	4.93	0.00001	0.29	2.26	P < 0.05
Alcohol consumption	0.12	5.28	P < 0.00001	0.04	0.56	ns	0.09	2.05	P < 0.05	0.28	3.68	P < 0.001
Physical activity	-1.19	-49.03	P < 0.00001	-0.84	-11.5	P < 0.00001	-0.75	-5.92	0.00001	-1	-6.86	P < 0.0001
Smoking status	-0.07	-3.4	P < 0.0001	0	-0.06	ns	-0.13	-1.34	ns	0.08	0.67	ns
Falls	0.32	15.6	P < 0.00001	0.7	9.83	P < 0.00001	0.53	5.28	0.00001	0.56	4.37	P < 0.0001
Mental health symptoms	0.62	19.1	P < 0.00001	1.1	11.28	P < 0.00001	0.43	3.43	0.0001	0.67	4.68	P < 0.0001

ns: non-significant. Abbreviations: SES: Socioeconomic status. No multiple correction methods were implemented.

Extended Data Table 4 | Ridge regression models of risk factors of healthy aging in longitudinal analyses after implementing imputation methods

Panel A. Predictors of cognition in wave 2 in Costa Rica. Models using imputation by means method.
Goodness to fit parameters of the model: $F(1, 5694) = 3.80$, $P < 0.00001$; $R^2 = 0.11 \pm 0.04$, $F^2 = 0.10$].

Type of predictor		Beta (β) Estimate	t value mean	p value mean
	_intercept	11.842.516	8.612.929	0
Demographics	Age	0.022109	11.598.449	0
	Sex	0.358246	276.059	0.010652
SDH	Education	0.500323	6.701.905	0
	Isolation	0.01	0.65485	0.552358
Health status	Diabetes	0.01	1.038.866	0.34156
	Hypertension	0.01	0.367683	0.740714
	Heart disease	0.01	0.571074	0.590581
	Falls	0.01	0.616598	0.577839
Mental health symptoms	Mental symptoms	0.155269	1.164.443	0.278789
Lifestyle	Alcohol consumption	0.01	0.525905	0.616515
	Physical activity	0.01	0.335521	0.740782
	Smoking status	0.01	0.422123	0.683545

Panel B. Predictors of functional capacity in wave 2 in Costa Rica. Models using imputation by means method.

Goodness to fit parameters of the model: $F(1, 5694) = 5.61$, $P < 0.00001$; $R^2 = 0.13 \pm 0.06$, $F^2 = 0.15$].

Type of factor		Beta (β) Estimate	t value mean	p value mean
	_intercept	4.621.263	2.609.299	0.020261
Demographics	Age	0.010958	4.510.959	0.040294
	Sex	0.01	0.737945	0.496519
SDH	Education	0.258568	2.825.756	0.007043
	Isolation	0.01	0.393355	0.706279
Health status	Diabetes	0.312372	1.674.278	0.118464
	Hypertension	0.407844	2.903.439	0.007207
	Heart disease	0.649884	1.600.779	0.164817
	Falls	2.723.801	2.876.403	0.017296
Mental health symptoms	Mental symptoms	0.61052	356.636	0.003034
Lifestyle	Alcohol consumption	0.160573	1.345.031	0.205568
	Physical activity	0.303345	2.000.146	0.07057
	Smoking status	0.01	0.519333	0.624439

Abbreviations. SDH=Social determinants of health. SES=Socioeconomic status No multiple correction methods were implemented.

Extended Data Table 5 | Ridge regression models of individual impact of each type of risk factor on healthy aging

Model predicting cognition including only health conditions					
	β Estimate	t value	p value	value	
_intercept	5.347.556	660.575	P< 0.0001		
Hypertension	0.308815	147.602	P< 0.0001		
Heart disease	0.516509	200.139	P< 0.0001		
Diabetes	0.168224	67.355	P< 0.0001		
Falls	0.57434	279.248	P< 0.0001		Differences of R ² and F regarding the model including all factors
R-squared				0.046482	-0.25604
Adjusted R-squared				0.046345	-0.25582
F-squared				0.048747	-0.38498
Model predicting cognition including only lifestyle factors					
	β Estimate	t value	p value	value	
_intercept	10.399.319	1.666.672	P< 0.05		
Alcohol consumption	0.242764	136.119	P< 0.0001		
Physical activity	-1.351.893	-695.218	P< 0.0001		
Smoking status	-0.144269	-80.801	P< 0.0001		Differences of R ² and F regarding the model including all factors
R-squared				0.094376	0.02855034439
Adjusted R-squared				0.094279	0.02848809687
F-squared				0.104211	0.04519922861
Model predicting cognition including only SDH					
	β Estimate	t value	p value	value	
_intercept	8.075.324	1.603.332	P< 0.05		
Education	0.46045	182.413	P< 0.0001		
Social isolation	-0.342752	-13.384	P< 0.001		
SES	-0.019599	-8.049	P< 0.001		Differences of R ² and F regarding the model including all factors
R-squared				0.015463	-0.28705
Adjusted R-squared				0.015357	-0.28681
F-squared				0.015706	-0.41802
Model predicting cognition including only Mental health conditions					
	β Estimate	t value	p value	value	
_intercept	6.310.501	98.443	P< 0.0001		
Mental health symptoms	0.909668	437.158	P< 0.0001		Differences of R ² and F regarding the model including all factors
R-squared				0.021274	-0.28124
Adjusted R-squared				0.021239	-0.28093
F-squared				0.021736	-0.41199
Model predicting cognition including only demographics					
	(β) Estimates	t value	p value	value	
_intercept	14.578.693	1.594.367	P< 0.05		
Age	-0.100823	-6.233.024	P< 0.0001		
Sex	0.418467	216.127	P< 0.0001		Differences of R ² and F regarding the model including all factors
R-squared				0.206078	-0.09644
Adjusted R-squared				0.206021	-0.09615
F-squared				0.259569	-0.17416

ns: non-significant. Abbreviations: SES: Socioeconomic status. No multiple correction methods were implemented.

Extended Data Table 6 | Group comparisons of the magnitude of risk factors of cognition between Costa Rica and China

Risk Factors	Costa Rica						
	minimum	maximum	center	Quartile 25%	Quartile 75%	whisker inferior	whisker superior
Age	0.011435	0.034005	0.024112	0.021003	0.026924	0.012122299	0.034004722
Sex	0.224981	0.40343	0.339154	0.325984	0.375184	0.252184676	0.403429877
Education	0.486365	0.549871	0.519667	0.513217	0.528321	0.490560099	0.549870867
Mental health symptoms	0.12552	0.218525	0.169749	0.143059	0.193514	0.125519723	0.218524558
Physical activity	0.003901	0.109597	0.039027	0.019777	0.037324	0.003900506	0.06364352
Risk Factors	China						
	minimum	maximum	center	Quartile 25%	Quartile 75%	whisker inferior	whisker superior
Age	0.187699	0.199974	0.193611	0.191812	0.194927	0.187699	0.199598
Sex	0.636338	0.96909	0.789591	0.710629	0.870065	0.636338	0.96909
Education	0.048111	0.429426	0.265508	0.173115	0.376382	0.048111	0.429426
Mental health symptoms	0.602924	1.232689	0.859683	0.723746	0.978574	0.602924	1.232689
Physical activity	0.627052	0.908322	0.77537	0.744155	0.835761	0.627052	0.908322

Extended Data Table 7 | Comparisons of the magnitude of risk factors of functional ability between Costa Rica and China

Risk Factors	Costa Rica						
	minimum	maximum	center	Quartile 25%	Quartile 75%	whisker inferior	whisker superior
Age	0.005936	0.019442	0.011788	0.007411	0.014566	0.005936265	0.019442243
Sex	0.078929	0.214377	0.171382	0.16797	0.189558	0.13558882	0.214377459
Diabetes	0.238686	0.490439	0.345114	0.279554	0.42674	0.238686071	0.490439383
Education	0.230654	0.304218	0.271284	0.251972	0.287755	0.230653867	0.304218494
Hypertension	0.376281	0.501206	0.422866	0.39516	0.431775	0.376281187	0.486697743
Heart disease	0.546176	1.077583	0.844483	0.73693	0.943549	0.546176333	1.077583167
Alcohol consumption	0.080332	0.223812	0.13611	0.112481	0.159252	0.080332476	0.223811948
Physical activity	0.186716	0.339602	0.281794	0.266187	0.308306	0.203008776	0.339601792
Mental health symptoms	0.563489	0.754958	0.676098	0.644889	0.726047	0.563489257	0.754958317
Risk Factors	China						
	minimum	maximum	center	Quartile 25%	Quartile 75%	whisker inferior	whisker superior
Age	0.127298	0.136238	0.132465	0.13111	0.134656	0.127298029	0.13623775
Sex	0.466699	0.642352	0.526995	0.484698	0.551429	0.466698572	0.642352079
Diabetes	0.204456	0.644642	0.462946	0.385519	0.551254	0.204455905	0.644641847
Education	0.010034	0.315218	0.151049	0.0899	0.20463	0.010034094	0.315218195
Hypertension	0.180356	0.436049	0.322079	0.28427	0.393464	0.180356253	0.436049379
Heart disease	0.119354	0.446786	0.322644	0.287639	0.398931	0.120702068	0.446785929
Alcohol consumption	0.082116	0.225934	0.177046	0.164362	0.204828	0.103661624	0.225933662
Physical activity	0.581949	0.69775	0.655955	0.630027	0.686631	0.581948955	0.697750298
Mental health symptoms	0.330473	0.616553	0.443535	0.364047	0.497958	0.330473161	0.616552896

Extended Data Table 8 | Closed-form solution for ridge regressions

Cognition			
	Alpha	Max_iter	Solver
LA-HICs	0.0001	1000000	saga
LA-LMICs	0.001	1000	auto
All	0.01	1000000	saga
Functional ability			
	Alpha	Max_iter	Solver
LA-HICs	0.001	100000	saga
LA-LMICs	0.001	1000000	sga
All	0.001	10000	sga

Abbreviations: LA-HICs: Latin American High-Income countries.
LA-LMICs: Latin American Low-and-Middle Income countries.
Max_iter: Maximum iterations.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#): Data collection Data were collected using paper-pencil formats and REDCAP platforms following specific procedures in each country. The researchers were blinded to experimental condition and/or the study hypothesis during data collection.

Data analysis All analyses were run using Python version 3.9.13. All code for data analysis of this manuscript is available for download at GitHub at <https://github.com/AI-BrainLat-team/Heterogeneous-risk-factors-LAC>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Data availability: Data included in this study in the cross-sectional and longitudinal analyses were collected in the context of the National Aging surveys of five LACs countries (Ecuador, Colombia, Chile, Uruguay, and Costa Rica) and from China. Data from the National Aging Survey from Ecuador included in the study are publicly available and can be accessed after providing researchers information at this link: <https://www.ecuadorencifras.gob.ec/encuesta-de-salud-bienestar-del-adulto-mayor/>. Data from the National Aging Survey of Colombia (National Study of Health, Wellbeing, and Aging SABA Colombia 2015) can be accessed after fulfilling a registration form on the webpage of the Ministry of Health and Social Protection (Spanish: el Ministerio de Salud y de Protección Social) in Colombia. This procedure lasts around ten days and can be made at this link: <https://www.datos.gov.co>. Data from Chile and Uruguay included in this study was part of the SABA - Survey on Health, Wellbeing, and Aging in Latin America and the Caribbean, 2000 (ICPSR 3546). Data from this study is publicly available after including researchers' information in this link: <https://www.icpsr.umich.edu/web/NACDA/studies/3546/versions/V1>. Data from Costa Rica included in this study was taken from the Costa Rican Longevity and Healthy Aging Study (CRELES). Data from this study is publicly available after introducing researchers' information in this link: <http://creles-download.demog.berkeley.edu/CRdata.pl>. Data from the National Aging Survey of China (China Health and Retirement Longitudinal Study (CHARLS)) can be accessed after researchers' registration at this link: <https://charls.pku.edu.cn/en/>. Access to raw data from all databases described in this section can last around ten days. The raw data generated in this study are also available for download at <https://github.com/AI-BrainLat-team/Heterogeneous-risk-factors-LAC>. All authors had access to the raw data. All software used in this study and its versions are specified on the condaenv.yml file on the GitHub repository.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	In this study we have included sex and gender of participants in our plan of analyses.
Population characteristics	We ran two types of analyses. First, we implemented cross-sectional analyses in N=31,680 older adults (above of 60 years-old; female= 16,074; mean age=70.81, SD=8.22). Participants were recruited from the National Surveys on Health, Well-Being, and Aging performed in four LACs: Chile (2016, N=1,301; females=689; mean age=71.83 SD= 8,29), Uruguay (2011, N=1,450; females=798; mean age=70.75 SD= 7,40), Ecuador (2012, N=5,235; females= 1,530; mean age= 70.09, SD= 7.82), and Colombia (2015, N=23,694; females=1,357; mean age=70.79 SD= 8.26). In a second group of analyses, we implemented longitudinal analyses using data from the Longitudinal National aging study of Costa Rica, (Costa Rican Study on Longevity and Healthy Aging (CRELES), N=5,694, mean age=59.5, SD=3.2). Moreover, as control information, we also analyzed National Aging longitudinal information from China (assessing data of the China Health and Retirement Longitudinal Study (CHARLS), N=3,546, mean age=81,2, SD=10,1).
Recruitment	Participants were selected following a probabilistic, clustered, stratified, and multistage design in each country. We only included individuals with no diagnosis of dementia, as assessed during an initial screening. Databases of national surveys from all countries are open and were obtained following the established procedures for each country.
Ethics oversight	The Institutional Review Boards of Universidad Javeriana, Bogotá, Colombia (code FM773-2021) reviewed and approved the current study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study combine cross-sectional and longitudinal approaches and machine learning techniques on quantitative data taken from National Aging Surveys from four South American countries (Colombia, Ecuador, Chile, Uruguay), one Center American country (Costa Rica) and one Asian Country (China)
Research sample	The study included older individuals (above 60 years old, n=44,394 participants, females= 19,074; mean age=70.81, SD=8.22) randomly recruited from four South American countries (Colombia, Ecuador, Chile, Uruguay), one Center American country (Costa Rica), and one Asian Country (China). All datasets from the National Aging surveys from Latin American countries, in cross-sectional and longitudinal analyses, included representative samples from each country.
Sampling strategy	Individuals were included in this study following a probabilistic, clustered, stratified, and multistage approach implemented in the National Survey of Aging study in each country. No statistical method was used to predetermine sample size, considering we used the complete database of the National Aging surveys from different Latin American countries.
Data collection	Data were collected using paper-pencil formats and REDCAP platforms following specific procedures in each country. The researchers were blinded to experimental condition and/or the study hypothesis during data collection.
Timing	In general the start date of data collection was between 1999 and 2000 and the stop date was 2016. Moreover, data were collected differently in each country. In Colombia, data were recruited between 2014 to 2015, and in Ecuador, Chile, and Uruguay, data were collected between 1999 and 2000. Data from Costa Rica were collected in 2012, and 2016. Data from China was collected in 2014, and 2016.
Data exclusions	All variables used in the analyses of this study had complete values in at least 80% of cases. Thus, we ran our analyses in a sample of 37,259 participants from 44,394 individuals.
Non-participation	We only assessed information of individuals who participated and finished all questions and requirements from the National Surveys of each country.
Randomization	Participants were not allocated in experimental groups. As part of the machine learning techniques we implemented, the models used in this work follow standard random training-test procedures.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- n/a Involved in the study
- Antibodies
- Eukaryotic cell lines
- Palaeontology and archaeology
- Animals and other organisms
- Clinical data
- Dual use research of concern

Methods

- n/a Involved in the study
- ChIP-seq
- Flow cytometry
- MRI-based neuroimaging

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.

Study protocol Note where the full trial protocol can be accessed OR if not available, explain why.

Data collection Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.

Outcomes Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.