



ARTICLE



<https://doi.org/10.1057/s41599-025-05708-9>

OPEN

# Low well-being among middle-aged people: inherent or external factors

Chao Li<sup>1</sup>, Jie Mi<sup>2</sup>, Jiaxu Zhang<sup>1</sup>, Bo Shi<sup>1</sup>, Alexander Keeley<sup>1</sup> & Shunsuke Managi<sup>1</sup>✉

Well-being is increasingly recognized as a fundamental goal at both individual and societal levels. Significant differences in well-being across age groups have long been detected and noted. However, the primary factors contributing to this disparity remain unknown. Here, leveraging an extensive global survey dataset and advanced machine-learning techniques, this study investigates the reasons underlying notably low levels of well-being among middle-aged individuals. Utilizing an exogenous switching treatment effect model (ESTEM) enhanced by machine learning, we analyze over 1.9 million individual observations from 168 countries collected between 2009 and 2022. Our results empirically confirm a U-shaped relationship between age and subjective well-being, indicating that middle-aged individuals consistently experience the lowest well-being. Further analysis reveals that middle-aged people receive significantly harsher external treatments compared to younger and older age groups, highlighting external societal conditions as critical contributors to the midlife crisis phenomenon. Conversely, elderly populations inherently experience higher subjective well-being. Temporal analyses indicate that external treatments for younger and middle-aged groups are becoming increasingly stringent relative to those for the elderly. By systematically mapping these treatment effects and intrinsic differences among age groups, this study provides critical insights to inform targeted policies and social programs designed to improve quality of life, thereby supporting equitable improvements in human well-being across the lifespan.

<sup>1</sup>Kyushu University, Fukuoka, Japan. <sup>2</sup>Asian Development Bank Institute, Tokyo, Japan. ✉email: [managi@doc.kyushu-u.ac.jp](mailto:managi@doc.kyushu-u.ac.jp)

## Introduction

Well-being has increasingly been viewed as a fundamental objective for both individuals and society, recently influencing public policy decisions (Diener and Chan, 2011; Diener et al. 2018; López Ulloa et al. 2013; Stone et al. 2010). Accurately improving the well-being of specific demographic groups effectively contributes to the welfare of society as a whole. The declining birth rate has exacerbated aging-related issues in most, if not all, major countries (Margolis and Myrskylä, 2011; Nomaguchi and Milkie, 2020). Lower well-being levels within particular age groups can have broader societal implications, underscoring the importance of understanding variations in well-being across age demographics. Indeed, the subject of age-related differences in well-being has consistently attracted extensive attention and debate among researchers, policymakers, and the general public (Diener et al. 2018; Wu et al. 2024). While numerous studies have attempted to characterize the trajectory of well-being throughout the lifespan, empirical results remain inconsistent, frequently yielding contradictory findings (López Ulloa et al. 2013).

Most studies have identified a link between age and well-being, e.g., (Blanchflower, 2021; Blanchflower and Piper, 2022; López Ulloa et al. 2013; Steptoe et al. 2015). Due to limitations in data availability, computing power, computational resources, and interpretability, previous studies have typically represented the age-well-being association through relatively simple patterns such as linear (Frijters and Beaton, 2012; Van Landeghem, 2012), U-shaped (Blanchflower, 2021; Blanchflower and Graham, 2022), and inverted U-shaped curves (Easterlin, 2006). Among these patterns, the U-shaped relationship has consistently been observed and remains the most widely accepted, where human well-being is lowest during middle age. This finding aligns intuitively with the commonly recognized phenomenon of the midlife crisis (Easterlin, 2006; Giuntella et al. 2023). For example, financial stress, a notable contributor to the midlife crisis, typically peaks in middle age, following an inverted U-shaped distribution (García Mata, 2023; Plagnol, 2011). On the other hand, several other studies have demonstrated either an inverted U-shaped or linear link between age and human well-being, which conflicts significantly with the U-shaped perspective. These studies argue that middle-aged people characterized by higher levels of energy, better health, greater experience, and improved skills would not logically experience lower well-being during midlife (Blanchflower and Oswald, 2008; Hansen and Blekesaune, 2022).

For instance, an analysis of the United States General Social Surveys conducted from 1973 to 1994 indicates that well-being peaks during midlife due to improved financial stability, career achievement, and health status, subsequently declining in older age—thus supporting an inverted U-shaped relationship (Easterlin, 2006). The reasons underlying these discrepancies in findings partially stem from differing psychological, economic, and sociocultural theoretical frameworks. Psychological theories, including midlife crisis theory and adaptation theory, propose that inherent psychological transitions during midlife negatively impact well-being, leading to a U-shaped relationship (Giuntella et al. 2023; Lachman, 2015). In contrast, economic theories, particularly the life-cycle hypothesis, anticipate a peak in midlife well-being resulting from economic stability and optimal productivity, thus demonstrating an inverted U-shape (Easterlin, 2006). Furthermore, sociocultural variations, such as differing societal expectations and cohort effects, complicate these patterns and occasionally yield linear relationships (Hansen and Blekesaune, 2022). Methodological differences and variations in datasets employed also contribute significantly to the inconsistencies observed. When well-being is treated as a continuous

variable and analyzed using ordinary least squares regression, a U-shaped association is frequently observed (Blanchflower, 2021; Diener et al. 2013). Conversely, when well-being is approached as a discrete variable in combination with generalized linear regression methods, relationships tend to appear inverted U-shaped (Easterlin, 2006). Cross-sectional analyses, often associated with U-shaped patterns, measure happiness across age groups simultaneously and may confound age and cohort effects (Blanchflower, 2021; Diener et al. 2018). Longitudinal and synthetic panel studies, however, attempt to track cohorts over time, typically utilizing smaller datasets (Easterlin, 2006). However, irrespective of statistical methods employed, previous studies generally exhibit relatively poor goodness-of-fit (Blanchflower and Graham, 2022; Blanchflower and Piper, 2022). In this way, the primary objective of the present study is to verify the empirical relationship between age and well-being with greater accuracy.

Most previous studies have primarily focused on the empirical relationship between age and human well-being itself, with relatively little attention paid to whether this relationship originates from inherent or external factors. Inherent factors refer to qualities or attributes that are intrinsic, natural, and fundamental to individuals, characterizing or influencing their well-being independent of external circumstances. Earlier studies have demonstrated, for instance, that people born in certain generations tend to exhibit higher levels of well-being compared to those born in other decades (Gerstorff et al. 2019; Shu et al. 2023; Sutin et al. 2013). Additionally, health status has been shown to significantly affect human well-being, with elderly populations typically experiencing poorer health outcomes (Bamidis et al. 2014; Li and Managi, 2023; Luu and Palczewski, 2018). These inherent characteristics, common within particular age cohorts, may contribute notably to differences in well-being across age groups (Diener et al. 2018; López Ulloa et al. 2013).

Conversely, external factors, including societal and familial influences, may also substantially shape individual well-being and can be collectively considered as external treatments. In this context, we specifically define “external treatment” as societal, institutional, and environmental conditions externally imposed upon individuals based on their age group, encompassing attitudes, policies, and societal expectations (Lachman, 2015; Nomaguchi and Milkie, 2020). For instance, financial stress affects human well-being and varies considerably by age (García Mata, 2023; Plagnol, 2011). Social role theory further elaborates this idea by positing that individuals fulfill age-specific roles involving caregiving, employment capacity, economic productivity, and social responsibilities, which cumulatively influence their overall well-being (Lachman, 2015; Nomaguchi and Milkie, 2020). Obviously, when societal and familial expectations exceed individuals’ capacities or available resources, this leads to increased stress and diminished well-being (Malik et al. 2022). Moreover, institutional discrimination represents another critical external treatment, potentially exacerbating the pressures arising from social role expectations. Policies and societal attention frequently prioritize support for relatively disadvantaged groups, for example, through education and pension systems, while other groups receive comparatively less support, potentially resulting in disparities and reduced well-being (García Mata, 2023). Such unfairness may lead to the deprivation of well-being. Clarifying whether inherent factors, external treatments, or a combination thereof underpin the age-well-being relationship can facilitate the development of targeted, effective, and efficient policy interventions.

The exogenous switching treatment effect model (ESTEM) is a causal inference method based on the counterfactual prediction to

detect the inherent and external effects of age on human well-being. ESTEM enables researchers to investigate how variations in an exogenous variable affect outcomes under different hypothetical scenarios, thereby isolating effects specifically attributable to these exogenous variables from broader environmental or contextual factors (Balcilar et al. 2015; Kassie et al. 2014). Technologically, based on the ESTEM, the difference caused by different external treatments is called the treatment effect. For a simple example, the well-being status of elderly people would hypothetically differ if they received the external treatment typically associated with young populations. Of course, the population of an age group could not be treated as another age group widely in reality, so the predictions are counterfactual. Conversely, base heterogeneity effects reflect inherent or intrinsic differences and can be examined by comparing populations of varying ages under identical external treatments. Specifically, base heterogeneity captures intrinsic variations in well-being across age groups, independent of external conditions, thus elucidating fundamental differences attributable solely to age itself.

Relatively lower goodness-of-fit and limited predictive power have consistently posed challenges for previous studies employing traditional regression methods (Blanchflower and Graham, 2022; Blanchflower and Piper, 2022). Moreover, empirical studies often neglect cross-validation, treating training and validation accuracies interchangeably. If the study only needs to fit the age-well-being relationship, the regression results can still meet the requirements. However, ESTEM specifically requires models with robust predictive capabilities because accurate counterfactual predictions constitute a critical component of its analytical process. Machine learning techniques are gaining increasing attention due to their superior predictive performance compared to conventional regression methods (Bent jac et al. 2021; Chen and Guestrin, 2016). In this study, we replace the linear regression model with an advanced tree-based machine learning model, namely extreme gradient boosting (XGBoost) (Chen and Guestrin, 2016). We solve several technical problems and provide an example of conducting analyses based on the ESTEM powered by machine learning.

## Methods

### Materials

*Survey information.* Our study is based on the individual-level survey conducted by Gallup, Inc. globally, named Gallup World Poll (GWP). The current GWP dataset covers 18 years from 2005 to 2022, including 17 waves of surveys. It should be noted that the first wave was conducted in 2005 and 2006, and each wave of surveys from the second wave onwards is completed within a year. The current version dataset contains 2.594 million individual observations from 168 countries or regions. Specifically, at least 1000 individuals are sampled in each country and each wave of the GWP survey. The GWP is the largest global dataset, mainly concentrating on human well-being, and it has been widely used in previous studies (Blanchflower, 2021; Diener et al. 2013; Jebb et al. 2018). The detailed procedures, method and data collection methodology in the GWP are concretely reported on Gallup's website (<https://www.gallup.com/178667/gallup-world-poll-work.aspx>).

Certain observations are excluded during the data cleaning process, so 1,911,212 observations were finally taken into account in the actual analysis. In the first three waves, the income is not asked, so the first three waves are excluded entirely. In other waves, if the respondents do not answer the income, we use the mean income of those respondents' countries in the corresponding wave to fulfill the unanswered income values. The mean income is computed by averaging other available values in the

GWP survey in the corresponding country and wave. If the income question is not asked for a country in a particular wave, the data for that country are deleted in that wave. After this step, 2,172,297 observations are retained. However, it should be noted that missing individual income values are rare in the data provided by Gallup. In this step, a total of 624 data are imputed, accounting for 0.03% of the total retained data. Of these 624 missing values, 623 occurred in Germany during the 2011 survey wave, which originally included 10,105 respondents. One additional missing value occurs in the Canadian survey conducted in 2016. However, due to the absence of other key variables, only the single Canadian observation with imputed income data is ultimately retained.

Furthermore, respondents are required to have provided data on well-being, the dependent variable in our analysis, reducing the total to 2,141,833 observations. This step keeps 2,141,833 observations. Previous studies indicate that disability significantly affects human well-being (Fredrickson et al. 2013; Kobau et al. 2010), so we drop the observations without available answers. Here, 2,025,803 observations are retained. Additionally, age, gender, and employment are strictly needed. After dropping the observation with no-answer items, in total, 1,911,212 observations are reserved in our dataset. The observation counts in each country and each wave are summarized in Supplementary Materials Table S1.

*Subjective well-being measurement.* Subjective well-being (SWB) has long been regarded as a reasonable indicator of human well-being (Diener, 1984; Diener et al. 2018; Oswald and Wu, 2010). Overall life evaluation is a critical approach to measure subjective well-being (Diener et al. 2018; Helliwell and Aknin, 2018; Kahneman and Deaton, 2010): it extracts well-being from people's thoughts about the quality of their overall life. In the GWP, an 11-point Cantril ladder is employed to evaluate overall human well-being; that is, the respondents should imagine a ladder with 11 steps (Diener et al. 2018). The ladder is from the lowest step of the ladder, numbered as 0, which represents the worst possible life for the respondents, to the highest step, numbered as 10, which represents the best possible life for the respondents. Respondents then selected the step on which they believe they are currently located. The number of the selected step is the respondent's evaluation of life, which is a number from 0 to 10. Because the Cantril ladder is straightforward to understand and widely used in previous studies (Blanchflower, 2021; Jebb et al. 2018), it is taken as the dependent variable in our study.

*Independent variables.* In this study, our dataset includes 63 independent variables, which are "Wave", "Country", "Household Income", "Health Disability", "Female Dummy", "Age", "Marital Status", "Employment", "Children Under 15", "Feeling of Income", "Income Level", "Having Relatives to Rely on", "Living Standard Changing Direction", "Enough Food", "Enough Shelter", "Well Rested", "Respected", "Smiling", "Interesting Things", "Enjoyment", "Physical Pain", "Worry", "Sadness", "Stress", "Anger", "Satisfied with City", "Economic Changing Direction", "Good Time to Find Job", "Satisfied with Public Transportation", "Satisfied with Road", "Satisfied with Education", "Satisfied with Air Quality", "Satisfied with Water Quality", "Satisfied with Healthcare", "Satisfied with Affordable House", "Satisfied with Opportunity to Make Friends", "Good Place for Ethical Minority", "Good Place for Gay or Lesbian", "Good Place for Immigrants", "Donated Recently", "Did Volunteer Recently", "Helped Stranger", "Voiced Opinion to Official", "Confidence in Local Police", "Safety of Alone Night Walking", "Stolen", "Assaulted", "Religion Importance", "Children Respected", "Opportunity for Children Learning", "Women Respected",

“Satisfied with Poverty Alleviation”, “Satisfied with Environmental Efforts”, “Freedom of Choosing Life”, “Confidence in Military”, “Confidence in Judicial System”, “Confidence in National Government”, “Confidence in Financial System”, “Confidence in Election Honesty”, “Freedom of Media”, “Corruption within Business”, “Corruption within Government”, and “Approval of Leadership Performance”. Table 1 summarizes all variables except “Wave” and “Country”. The details of each question in the survey and value explanations are listed in Supplementary Materials Table S2.

**Relationship between well-being and age investigation.** Previous empirical studies repeatedly indicated a U-shaped relationship between age and well-being, based primarily on the consistent significance of the age-squared term in linear or generalized linear regression analyses (Blanchflower and Oswald, 2008; López Ulloa et al. 2013). However, there is no solid evidence supporting this, although it is usually consistent with people’s intuition and observation. In other words, the phenomenon of the midlife crisis is common and widely noticed (Giuntella et al. 2023; Lachman, 2015). Due to the limitations of data volume and technology, linear regression is a compromise but effective method, i.e., linear regression is not good at fitting non-linear relationships, but its explanation is straightforward. Machine learning models are designed to optimize predictive accuracy by minimizing prediction errors (Chen and Guestrin, 2016). Additionally, machine learning models make no statistical assumptions about the shape of relationships, which increases the ability to fit the linear relationship (Bentéjac et al. 2021; Chen and Guestrin, 2016).

**XGBoost and its fine-tuning.** To detect the empirical relationship between age and well-being, we first use machine learning to fit the total dataset. Second, we choose a reasonable explanation method to explain the machine learning methods, because those models are not as straightforward as the linear models (Molnar, 2020; Molnar et al. 2020). In this study, we take the XGBoost regressor as the main algorithm to replace the linear regression or other linear methods in previous studies to detect the empirical relationship. Our analysis is set as a regression task, as in previous studies, since the dependent variable is an 11-point Cantril Ladder well-being evaluation. We use the Shapley additive explanation (SHAP) method (Lundberg et al. 2020) to explain the XGBoost results based on the total dataset.

The XGBoost has several significant advantages. The XGBoost is a decision tree-based model (Chen and Guestrin, 2016). Decision trees have a strong ability to handle complex, non-linear relationships between variables with relatively “rich” tabular data (Bentéjac et al. 2021). Furthermore, decision trees could process various data types, including binary variables, continuous variables, and categorical variables. Furthermore, decision trees are completely non-parametric, that is, they do not assume any specific distribution for the data (Chen and Guestrin, 2016). However, decision trees are prone to overfitting, especially when the tree is grown extremely deep. Ensemble methods, such as gradient boosting and random forest, and taking decision trees as base learners, could significantly enhance performance. The traditional gradient boosting process is inherently sequential, making it challenging to parallelize (Chen and Guestrin, 2016). The XGBoost model is an optimized gradient boosting model, which supports parallel computation and even GPU-acceleration. Of course, other technologies, such as CatBoost (Prokhorenkova et al. 2018) and LightGBM (Ke et al. 2017), random forest based on XGBoost, are also compatible with GPU acceleration. After fine-tuning those models, XGBoost performs the best in

generalization, specifically, the  $R^2$  of fine-tuned XGBoost, CatBoost, LightBoost, and random forest are 37.81%, 37.53%, 37.38%, and 36.46%. Furthermore, artificial neural networks are also considered. However, they require extensive hyperparameter tuning, including architecture, activation functions, learning rates, and batch sizes, which can be computationally intensive and time-consuming. In a fairly large search range, artificial neural networks do not outperform tree-based models. Additionally, when combined with SHAP for interpretation, artificial neural networks are relatively slower than XGBoost with GPU acceleration. Therefore, we select XGBoost as the main algorithm in the analyses

The training process of the XGBoost regressor for the total dataset is as follows:

$$XGB_{tot} = \Theta(Xtrain_{tot}, ytrain_{tot}, Hyperparameter_{tot}) \quad (1)$$

where  $XGB_{tot}$  represents the well-train XGBoost regression model based on the total dataset,  $tot$  represents the total dataset including all independent variables and all observations,  $Xtrain_{tot}$  represents the independent variables of the training dataset split from the total dataset,  $ytrain_{tot}$  represents the dependent variables of the training dataset,  $Hyperparameter_{tot}$  represents a set of hyperparameters to train a high-accuracy XGBoost model, and  $\Theta$  represents the training process of XGBoost algorithm, which builds an ensemble of decision trees sequentially, where each new tree corrects the residual errors of the previous ones using gradient boosting to minimize a defined loss function. The split ratio between the training dataset and test dataset is 9:1. In other words, we randomly sample 90% data as the training dataset and the other 10% data is left as the test dataset.

The hyperparameter set,  $Hyperparameter_{tot}$ , includes the number of trees (“n\_estimators”), learning rate (“learning\_rate”), the maximum depth of each tree (“max\_depth”), the subsample ratio of training instance (“subsample”), the minimum loss reduction required for a new split to be added to the tree (“gamma”), the minimum sum of instance weight needed in a child leaf (“min\_child\_weight”), the maximum step size that a XGBoost model’s weight can change at each boosting iteration (“max\_delta\_step”), L1 regularization term on weights (“reg\_alpha”), and L2 regularization term on weights (“reg\_lambda”). The abbreviation in brackets remains consistent with XGBoost’s Python “xgboost” 2.0.1 API to facilitate reproduction and imitation by other researchers. To distinguish them from textual words, they will always be enclosed in quotation marks. We adopt the cross-validation method to search for the best hyperparameters. The cross-validation method is 10-fold cross-validation, but due to the limits of computing resources, we only conduct 3 folds of it to balance the time consumption and stability of the searching process. The metric of the searching process is  $R^2$  of the test dataset. The  $R^2$  of the test dataset is computed as follows:

$$\widehat{ytest}_{tot} = XGB_{tot}(Xtest_{tot}) \quad (2)$$

$$R^2_{test\ tot} = 1 - \frac{(ytest_{tot} - \widehat{ytest}_{tot})^2}{(ytest_{tot} - \overline{ytest}_{tot})^2} \quad (3)$$

where  $\widehat{ytest}_{tot}$  represents the predicted values corresponding to the input test dataset,  $Xtest_{tot}$ , of well-trained XGBoost model  $XGB_{tot}$ ,  $\overline{ytest}_{tot}$  is the mean real value of an independent variable, and  $R^2_{test\ tot}$  represents the  $R^2$  of the test dataset for the model trained by the training dataset from the total dataset. Combining Eqs. (1)–(3), it is clear that the  $R^2_{test\ tot}$  is highly related to hyperparameters.

We use Bayesian hyperparameter optimization to search for the best hyperparameter set (Turner et al. 2021). The Bayesian hyperparameter optimization normally contains four steps:

**Table 1 Data summary.**

	Mean	Std	Min	25%	50%	75%	Max
SWB	5.521	2.414	0	4	5	7	10
Household Income	25,339.301	925,219.104	0.000	4190.875	10510.192	26,055.600	898,033,649.954
Health Disability	1.754	0.431	1	2	2	2	2
Female Dummy	0.532	0.499	0	0	1	1	1
Age	41.429	17.552	15	27	39	54	99
Marital Status	2.344	1.710	1	1	2	2	8
Employment	3.646	2.134	1	1	4	6	6
Children Under 15	0.522	0.500	0	0	1	1	1
Feeling of Income	2.323	0.965	1	2	2	3	4
Income Level	3.236	1.415	1	2	3	5	5
Having Relatives to Rely On	0.791	0.407	0	1	1	1	1
Living Standard Changing Direction	0.175	0.820	-1	-1	0	1	1
Enough Food	0.322	0.467	0	0	0	1	1
Enough Shelter	0.245	0.430	0	0	0	0	1
Well Rested	0.674	0.469	0	0	1	1	1
Respected	0.865	0.342	0	1	1	1	1
Smiling	0.703	0.457	0	0	1	1	1
Interesting Things	0.520	0.500	0	0	1	1	1
Enjoyment	0.689	0.463	0	0	1	1	1
Physical Pain	0.307	0.461	0	0	0	1	1
Worry	0.380	0.485	0	0	0	1	1
Sadness	0.234	0.423	0	0	0	0	1
Stress	0.329	0.470	0	0	0	1	1
Anger	0.197	0.398	0	0	0	0	1
Satisfied with City	0.758	0.429	0	1	1	1	1
Economic Changing Direction	0.054	0.851	-1	-1	0	1	1
Good Rime to Find Job	0.337	0.473	0	0	0	1	1
Satisfied with Public Transportation	0.560	0.496	0	0	1	1	1
Satisfied with Road	0.523	0.499	0	0	1	1	1
Satisfied with Education	0.604	0.489	0	0	1	1	1
Satisfied with Air Quality	0.713	0.453	0	0	1	1	1
Satisfied with Water quality	0.668	0.471	0	0	1	1	1
Satisfied with Healthcare	0.544	0.498	0	0	1	1	1
Satisfied with Affordable house	0.463	0.499	0	0	0	1	1
Satisfied with Opportunity to make friends	0.684	0.465	0	0	1	1	1
Good Place for Ethitical Minority	0.575	0.494	0	0	1	1	1
Good Place For Gay Or Lesbian	0.287	0.452	0	0	0	1	1
Good Place for Immigrants	0.569	0.495	0	0	1	1	1
Donated Recently	0.306	0.461	0	0	0	1	1
Did Volunteer Recently	0.201	0.401	0	0	0	0	1
Helped Stranger	0.495	0.500	0	0	0	1	1
Voiced Opinion to Official	0.159	0.366	0	0	0	0	1
Confidence in Local Police	0.558	0.497	0	0	1	1	1
Safety of Alone Night Walking	0.590	0.492	0	0	1	1	1
Stolen	0.139	0.346	0	0	0	0	1
Assaulted	0.047	0.211	0	0	0	0	1
Religion Importance	0.637	0.481	0	0	1	1	1
Children Respected	0.640	0.480	0	0	1	1	1
Opportunity for Children Learning	0.671	0.470	0	0	1	1	1
Women Respected	0.566	0.496	0	0	1	1	1
Satisfied with Poverty Alleviation	0.357	0.479	0	0	0	1	1
Satisfied with Environmental Efforts	0.494	0.500	0	0	0	1	1
Freedom of Choosing Life	0.723	0.448	0	0	1	1	1
Confidence in Military	0.586	0.493	0	0	1	1	1
Confidence in Judicial System	0.431	0.495	0	0	0	1	1
Confidence in National Government	0.420	0.494	0	0	0	1	1
Confidence in Financial System	0.531	0.499	0	0	1	1	1
Confidence in Election Honesty	0.402	0.490	0	0	0	1	1
Freedom of Media	0.543	0.498	0	0	1	1	1
Corruption Within Business	0.610	0.488	0	0	1	1	1
Corruption Within Government	0.581	0.493	0	0	1	1	1
Approval of Leadership Performance	0.406	0.491	0	0	0	1	1

initialization based on several sets of hyperparameters, surrogate function construction, next set of hyperparameters selection and metric estimation, and updating the surrogate function. The surrogate probabilistic function approximates the objective function based on a Gaussian process. The third and fourth steps should be iterated several times to obtain the set of hyperparameters that could achieve high performance. The Bayesian hyperparameter optimization is conducted by the Python library “scikit-optimize” 0.8.1. The specific pseudo-code of the optimization and acquisition functions is well-recorded in the webpage of “scikit-optimize” (Gilles and Manoj, 2016). In this study, we set 20 iterations. Simply, the surrogate function’s input is a set of hyperparameters, and the output is the estimated  $R^2$  of the test dataset. The surrogate functions are continuously optimized by interactions. The ranges of these hyperparameters are as follows: 100–5000 for “n\_estimators”; 0.001–0.1 for “learning\_rate”; 3–16 for “max\_depth”; 0.5–1.0 for “subsample”; 0.001–10 for “min\_child\_weight”; 0.001–10 for “max\_delta\_step”; 0.001–10 for “gamma”; 0.001–10 for “reg\_alpha”, and 0.001–10 for “reg\_lambda”. The loss function is set as a square error function. Additionally, other hyperparameters are retained as the default settings of the Python library “xgboost” 2.0.1. We compare the difference between Bayes hyperparameter optimization results with 20 iterations and grid searching results with more than 3000 possible hyperparameter sets. The Bayes hyperparameter optimization results are relatively better. Although further fine-tuning the hyperparameters by grid searching has the possibility to increase the performance, the time costs are too high, and the improvement might be marginal. Hence, we use Bayesian hyperparameter optimization to fine-tune all XGBoost models in this study.

*Contributions of independent variables to well-being.* Tree-based ensemble models, like XGBoost, are entirely non-parametric, making it challenging to interpret their results (Molnar, 2020). SHAP offers a novel and effective approach to estimating the contribution of each specific independent variable to the dependent variable individually in machine learning models (Lundberg et al. 2020). The method leverages cooperative game theory and Shapley values to ensure that the contributions of the independent variables to the complex model’s predictions are fairly and evenly distributed (Lundberg et al. 2020; Molnar, 2020). Naturally, Shapley values are computed by evaluating the change in the predictions based on the well-trained machine learning model before and after adding a certain independent variable to all possible subsets of other independent variables, then averaging these marginal contributions. The contributions of each independent variable at the individual level could be written as follows:

$$SHAP_{test\_tot} = SHAP(XGB_{tot}, X_{test\_tot}) \quad (4)$$

where  $XGB_{tot}$  represents the XGBoost regression model trained by the training dataset split from the total dataset, SHAP represents the standard SHAP algorithm, and  $SHAP_{test\_tot}$  represents the SHAP values of each independent variable and each observation in the test dataset. Theoretically, we could directly use the well-trained XGBoost model and SHAP algorithm to explain all observations, even though the model is overfitting. The reason is that the SHAP method enumerates all the subsets, including the independent variable of interest, and among hundreds of subsets, only one is the same as the input dataset with all independent variables. If the model learned the observations in the training process, the prediction performance would be significantly better than general observations. To avoid this situation, we only investigate the observations in the test dataset. Of course, we could use 10-fold explanations for the whole dataset to solve this

issue, which is similar to 10-fold cross-validation. Specifically, we divide the total dataset into 10 folds, use nine folds to train the XGBoost model, employ SHAP and the trained XGBoost to explain the left one fold’s observations, and iterate this process until enumerating all possible combinations. However, the SHAP method is computationally expensive. Each test dataset has approximately 200,000 observations. With a rational setting, we need at least 100 GPU hours to complete the computation. Additionally, we just want to observe an intuitive relationship, and nearly 200,000 observations are representative enough.

**Exogenous switching treatment effect model (ESTEM).** To investigate the treatment effects and base heterogeneity effects for the well-being difference among three age groups, we conduct an ESTEM. The ESTEM is to estimate causal effects based on counterfactual analyses. Similar to other causal inference methods, the ESTEM also relies on several key assumptions. First, we assume no unobserved confounding. In other words, all variables influencing both the treatment assignment variable, namely age, and the outcome are observed and included in the model. Blocking confounding is a core basis for structural causal inference (Pearl, 2003, 2009). This assumption is often considered plausible because age is not subject to manipulation and is generally treated as an exogenous variable. However, we acknowledge that variables interacting with age, such as income, employment status, or health, may themselves be influenced by complex causal relationships. While our models attempt to control for such confounding factors, residual endogeneity cannot be entirely ruled out. Therefore, we interpret the causal implications of ESTEM with appropriate caution and frame them as suggestive rather than definitive. Second, we assume the machine learning models are flexible and powerful enough to accurately capture the true and real-world status. Although our models’ performance has not reached amazing levels, such as 95% accuracy, they have achieved significant improvements in accuracy compared to traditional technologies. The third assumption is that the covariate distributions across age groups have sufficient overlap to enable credible counterfactual predictions. Nearly 2 million valid data ensure that the effect investigation is based on a randomized study. These assumptions align with core principles from foundational causal inference literature, including the Rubin causal effect estimation (Rubin, 1974) and Pearl’s structural causal inference (Pearl, 2003, 2009).

The ESTEM divides all observations into several groups as sub-datasets by an exogenous variable, then uses each sub-dataset to train corresponding models, takes each model to predict each sub-dataset, and compares the differences between the predicted values. Making it easy, the pattern of external treatments to a specific group of people could be learned by a complex model. If the dependent variables of a certain group of observations are predicted by another model, that could be regarded as if this group is treated as another group externally. Of course, this is impossible in reality, so this method is counterfactual. In this study, we divide the entire dataset into three age groups: the young population aged 40 or less, the middle-aged population aged 40 and above but not exceeding 65, and the elderly population aged more than 65. Here, we present three separate training processes for the young population, middle-aged population, and elderly population as follows:

$$\begin{cases} XGB_{yo} = \Theta(X_{train_{yo}}, y_{train_{yo}}, Hyperparameter_{yo}) \\ XGB_{ma} = \Theta(X_{train_{ma}}, y_{train_{ma}}, Hyperparameter_{ma}) \\ XGB_{el} = \Theta(X_{train_{el}}, y_{train_{el}}, Hyperparameter_{el}) \end{cases} \quad (5)$$

where  $XGB_{yo}$ ,  $XGB_{ma}$ , and  $XGB_{el}$  represent the well-trained XGBoost regression models based on the young population,

middle-aged population, and elderly population datasets, respectively,  $Xtrain_{yo}$ ,  $Xtrain_{ma}$ , and  $Xtrain_{el}$  represent the independent variables of the training dataset split from the young population, middle-aged population, and elderly population datasets, respectively,  $ytrain_{yo}$ ,  $ytrain_{ma}$ , and  $ytrain_{el}$  represent the dependent variables of those three datasets, and **Hyperparameter**<sub>yo</sub>, **Hyperparameter**<sub>ma</sub>, and **Hyperparameter**<sub>el</sub> represent three sets of hyperparameters to train high-accuracy XGBoost models for each sub-dataset. We also adopt the cross-validation method to search for the best hyperparameter sets, as in the previous process, as Eqs. (1)–(3). Additionally, it must be noted that the independent variable “age” is not included in the sub-dataset when training the models and predicting.

To assess the role of age in well-being disparity among age groups, we evaluate the counterfactual well-being status of each age group. Based on this method, the effects of age on well-being could be disentangled by comparing the predicted well-being status under the actual and counterfactual situations. The predictions of the actual and counterfactual well-being of each age group are computed as follows:

$$\widehat{ytest}_{yo}^{yo} = XGB_{yo}(Xtest_{yo}) \tag{6}$$

$$\widehat{y}_{yo}^{ma} = XGB_{yo}(X_{ma}) \tag{7}$$

$$\widehat{y}_{yo}^{el} = XGB_{yo}(X_{el}) \tag{8}$$

$$\widehat{y}_{ma}^{yo} = XGB_{ma}(X_{yo}) \tag{9}$$

$$\widehat{ytest}_{ma}^{ma} = XGB_{ma}(Xtest_{ma}) \tag{10}$$

$$\widehat{y}_{ma}^{el} = XGB_{ma}(X_{el}) \tag{11}$$

$$\widehat{y}_{el}^{yo} = XGB_{el}(X_{yo}) \tag{12}$$

$$\widehat{y}_{el}^{ma} = XGB_{el}(X_{ma}) \tag{13}$$

$$\widehat{ytest}_{el}^{el} = XGB_{el}(Xtest_{el}) \tag{14}$$

where  $\widehat{ytest}_{yo}^{yo}$ ,  $\widehat{ytest}_{ma}^{ma}$ , and  $\widehat{ytest}_{el}^{el}$  are predicted well-being status of observations in test datasets of young population, middle-aged population, and elderly population sub-datasets estimated by well-trained corresponding models, respectively, which are the actual cases,  $\widehat{y}_{yo}^{ma}$ ,  $\widehat{y}_{yo}^{el}$ ,  $\widehat{y}_{ma}^{yo}$ ,  $\widehat{y}_{ma}^{el}$ ,  $\widehat{y}_{el}^{yo}$ , and  $\widehat{y}_{el}^{ma}$  are counterfactual predicted well-being. Since we use 90% of the data in a sub-dataset to train the model for an age group, this model cannot be directly applied for the prediction of all data in this age group. Overfitting is the main reason. Similarly, we employ a 10-fold process to predict all actual cases. Moreover, since these 10-fold predictions are from 10 not-exactly-the-same models, this might lead to lower reliability and stability of predictions. We conduct a 10-fold prediction 10 times based on different fold division strategies. In this way, each age group would obtain 100 different models based on the same hyperparameters. The 10-fold 10-epoch computations could be written as follows:

$$\begin{cases} XGB_{yo@fe} = \Theta(Xtrain_{yo@fe}, ytrain_{yo@fe}, Hyperparameter_{yo}) \\ XGB_{ma@fe} = \Theta(Xtrain_{ma@fe}, ytrain_{ma@fe}, Hyperparameter_{ma}) \\ XGB_{el@fe} = \Theta(Xtrain_{el@fe}, ytrain_{el@fe}, Hyperparameter_{el}) \end{cases} \tag{15}$$

$$\widehat{ytest}_{yo@fe}^{yo} = XGB_{yo@fe}(Xtest_{yo@fe}) \tag{16}$$

$$\widehat{y}_{yo@fe}^{ma} = XGB_{yo@fe}(X_{ma}) \tag{17}$$

$$\widehat{y}_{yo@fe}^{el} = XGB_{yo@fe}(X_{el}) \tag{18}$$

$$\widehat{y}_{ma@fe}^{yo} = XGB_{ma@fe}(X_{yo}) \tag{19}$$

$$\widehat{ytest}_{ma@fe}^{ma} = XGB_{ma@fe}(Xtest_{ma@fe}) \tag{20}$$

$$\widehat{y}_{ma@fe}^{el} = XGB_{ma@fe}(X_{el}) \tag{21}$$

$$\widehat{y}_{el@fe}^{yo} = XGB_{el@fe}(X_{yo}) \tag{22}$$

$$\widehat{y}_{el@fe}^{ma} = XGB_{el@fe}(X_{ma}) \tag{23}$$

$$\widehat{ytest}_{el@fe}^{el} = XGB_{el@fe}(Xtest_{el@fe}) \tag{24}$$

where  $f$  represents the fold indicator,  $e$  represents the epoch indicator,  $XGB_{yo@fe}$  represents the XGBoost model trained for the young population in the  $f$  fold of the  $e$  epoch,  $\widehat{ytest}_{yo@fe}^{yo}$  represents the well-being predictions of the test data of the young population sub-dataset based on the corresponding model in the  $f$  fold of the  $e$  epoch,  $\widehat{y}_{yo@fe}^{ma}$  represents the prediction of the middle-aged population based on the model trained by the young population sub-dataset in the  $f$  fold of the  $e$  epoch, and the explanations of other symbols are similar. In the 10-fold 10-epoch computations, each actual case should be estimated 10 times, and each counterfactual case should be calculated 100 times. We use the mean values of each individual for further computation:

$$apy_{yo}^{yo} = iwm(\widehat{ytest}_{yo@fe}^{yo}) \tag{25}$$

$$apy_{yo}^{ma} = iwm(\widehat{y}_{yo@fe}^{ma}) \tag{26}$$

$$apy_{yo}^{el} = iwm(\widehat{y}_{yo@fe}^{el}) \tag{27}$$

$$apy_{ma}^{yo} = iwm(\widehat{y}_{ma@fe}^{yo}) \tag{28}$$

$$apy_{ma}^{ma} = iwm(\widehat{ytest}_{ma@fe}^{ma}) \tag{29}$$

$$apy_{ma}^{el} = iwm(\widehat{y}_{ma@fe}^{el}) \tag{30}$$

$$apy_{el}^{yo} = iwm(\widehat{y}_{el@fe}^{yo}) \tag{31}$$

$$apy_{el}^{ma} = iwm(\widehat{y}_{el@fe}^{ma}) \tag{32}$$

$$apy_{el}^{el} = iwm(\widehat{ytest}_{el@fe}^{el}) \tag{33}$$

where  $apy_{yo}^{yo}$ ,  $apy_{yo}^{ma}$ ,  $apy_{yo}^{el}$ ,  $apy_{ma}^{yo}$ ,  $apy_{ma}^{ma}$ ,  $apy_{ma}^{el}$ ,  $apy_{el}^{yo}$ ,  $apy_{el}^{ma}$ , and  $apy_{el}^{el}$  represents the individual-wise average predicted well-being status of the young population, middle-aged population, and elderly population based on models trained by the young population, middle-age population, and elderly population, and  $iwm$  represents individual-wise mean method.

The treatment effects could be estimated as follows:

$$TE_{mag1-mag2}^{pag} = \overline{apy_{mag1}^{pag}} - \overline{apy_{mag2}^{pag}} \tag{34}$$

where  $TE_{mag1-mag2}^{pag}$  represents the treatment effect of changing from the external treatment for age group  $mag1$  to the external

treatment for age group mag2 on the population of age group pag,  $\overline{apy}_{mag1}^{pag}$  represents the average value of the predictions of the population in age group pag based on the model trained by mag1 population,  $\overline{apy}_{mag2}^{pag}$  represents the average value of the predictions of the population in age group pag based on the model trained by mag2 population, and pag, mag1, and mag2 represent one of the age group from young, middle-aged, and elderly age group. The base heterogeneity effects could be computed as follows:

$$BHE_{mag}^{pag1-pag2} = \overline{apy}_{mag1}^{pag1} - \overline{apy}_{mag2}^{pag2} \quad (35)$$

where  $BHE_{mag}^{pag1-pag2}$  represents the base heterogeneity effects between age group pag1 and pag2 based on the model trained by the population in age group mag,  $\overline{apy}_{mag1}^{pag1}$  represents the average value of the predictions of the population in age group pag1 based on the model trained by mag population,  $\overline{apy}_{mag2}^{pag2}$  represents the average value of the predictions of the population in age group pag2, and pag1, pag2, and mag represent one of the age group. In order to confirm the significance of treatment effects and base heterogeneity effects, we conduct the *t*-test between each group of average predicted values computed by Eqs. (25)–(33).

The temporal variations of base heterogeneity effects and treatment effects could be estimated because the wave order of the survey is included as an independent variable in the analysis. In Eqs. (16)–(24), we need to refine the input data by the wave order. Specifically, to compute base heterogeneity effects and treatment effects in a year, say 2016, we only predict the well-being status of observations surveyed in 2016. The country-level variation of treatment effects should be calculated directly based on all observations in a single country. This is to reduce the complexity of the results and make them easier to understand. The country-level treatment effects are computed as follows:

$$\overline{apy}_{mag}^{country} = \overline{apy}_{mag}^{yo@country} \cup \overline{apy}_{mag}^{ma@country} \cup \overline{apy}_{mag}^{el@country} \quad (36)$$

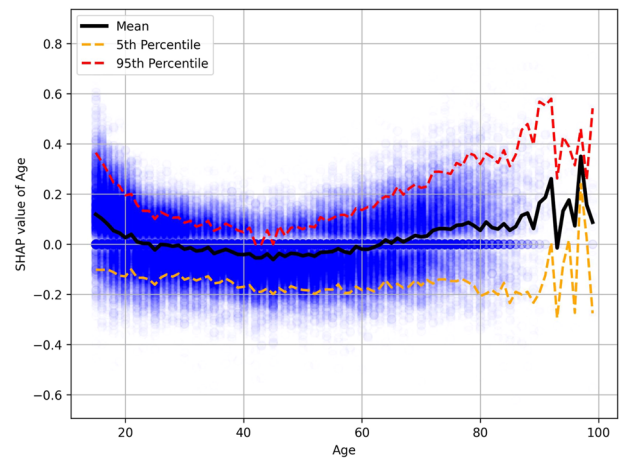
$$CTE_{mag1-mag2}^{country} = \overline{apy}_{mag1}^{country} - \overline{apy}_{mag2}^{country} \quad (37)$$

where  $\overline{apy}_{mag}^{country}$  represents the union of individual-wise average predicted well-being status of the young population, middle-aged population, and elderly population in a certain country, country, based on the model trained by the population in age group mag,  $CTE_{mag1-mag2}^{country}$  represents country-level treatment effects between age group mag1 and mag2 in the country, country. It should be noted that country-level base heterogeneity effects are challenging to detect since the difference in dataset sizes of the three age groups is more significant within a single country. Therefore, in this study, we do not go further to investigate the base heterogeneity within countries.

## Results

**Results of intuitive relationship between well-being and age.** To estimate the intuitive relationship between well-being and age, we conduct two steps: first, calibrating the best model to fit the relationship with human well-being, and second, using SHAP to estimate the age’s contribution to human well-being. It should be noted that the independent variables of this model include age, which is different from other models in the ESTEM.

To fine-tune the hyperparameters of the XGBoost for the datasets, including the independent variable, age, we conduct Bayesian hyperparameter optimization with 20 iterations based on our cross-validation process. The best hyperparameter set includes “n\_estimators” of 2441, “learning\_rate” of 0.0309, “max\_depth” of 11, “subsample” of 0.653, “min\_child\_weight”



**Fig. 1** The relationship between SHAP value of age and age.

of 0.167, “max\_delta\_step” of 0.382, “gamma” of 0.001, “reg\_alpha” of 1.124, and “reg\_lambda” of 0.007. The average test  $R^2$  of the fine-tuned XGBoost is 38.96% from three single-fold  $R^2$  38.93%, 38.94%, and 39.02%. Its average training  $R^2$  is 60.09% from three single-fold  $R^2$  60.10%, 60.08%, and 60.09%. Obviously, there is somewhat overfitting. Reducing model complexity can help mitigate overfitting; however, it may also lead to a decrease in the model’s accuracy on the test data. Therefore, we keep the best hyperparameter set. Relatively, the traditional model, ordinary least squares, performs worse in the 10-fold cross-validation. Specifically, its mean validation  $R^2$  is 33.81%, which is 5.15% lower than XGBoost’s performance on this dataset. In other words, XGBoost has a 15.23% performance improvement over the traditional model.

Figure 1 illustrates the relationship between the SHAP values of age and age. The SHAP values are explained as the contributions of age to human well-being. The middle ages contribute to human well-being the least, and the U-shape could be easily detected. In this way, the intuitive relationship, or empirical relationship, between human well-being and age is consistent with most previous studies (Blanchflower, 2021; Blanchflower and Oswald, 2008; Diener et al. 2018; Stone et al. 2010). It should be emphasized that the XGBoost model here is different from the models in ESTEM. The aim of this XGBoost model and SHAP values is to probe whether the fairly and objectively distributed contribution of age to well-being varies among different ages. This preliminary analysis is a crucial prerequisite for the subsequent ESTEM analysis, and the observed variations across age groups have indeed been confirmed.

## Well-being differences among age groups and reasons for the differences

*Well-being differences among age groups.* The dataset is divided into three age groups: the young population, the middle-aged population, and the elderly population. The young population sub-dataset has 1,031,174 observations; the middle-aged population sub-dataset has 663,573 observations; and the elderly population sub-dataset has 216,465 observations. The mean SWB values of the young population, the middle-aged population, and the elderly population are 5.462, 5.546, and 5.728, respectively. To verify the statistical significance of these differences, pairwise *t*-tests were performed among the three groups. All pairwise comparisons yielded highly significant differences (*p*-values < 0.001), indicating that average SWB significantly increases with age.

**Table 2 Summary of test  $R^2$  in 10-epoch 10-fold Prediction.**

Sub-dataset	Model	Average test $R^2$ (%)	Maximum test $R^2$ (%)	Minimum test $R^2$ (%)	Standard deviation (%)	Number of predictions for each obs.
Young population	Young population	35.23	36.04	34.50	0.28	10
Middle-age population	Young population	39.83	39.86	39.80	0.02	100
Elderly population	Young population	38.74	38.83	38.60	0.04	100
Young population	Middle-age population	33.14	33.18	33.08	0.02	100
Middle-age population	Middle-age population	40.72	41.53	40.00	0.31	10
Elderly population	Middle-age population	40.77	40.81	40.71	0.02	100
Young population	Elderly population	29.79	29.91	29.66	0.06	100
Middle-age population	Elderly population	37.64	37.72	37.56	0.03	100
Elderly population	Elderly population	40.43	41.69	39.17	0.51	10

*Models for three age groups fine-tuning and their performance.* We employ Bayesian hyperparameter optimization with 20 iterations based on the cross-validation process to calibrate the best hyperparameter sets for three models for three age groups. For the young population, the average test  $R^2$  of the best model is 35.23% based on three single-fold test  $R^2$  34.80%, 35.51%, and 35.39%, while its average training  $R^2$  is 42.76% based on three single-fold training  $R^2$  42.78%, 42.75%, and 42.74%. In this way, the overfitting status exists in the model for the young population. Therefore, 10-epoch 10-fold predictions are necessary to solve the overfitting. The best hyperparameter set of the model for young population includes “n\_estimators” of 1136, “learning\_rate” of 0.0252, “max\_depth” of 8, “subsample” of 0.797, “min\_child\_weight” of 0.033, “max\_delta\_step” of 3.066, “gamma” of 6.636, “reg\_alpha” of 0.001, and “reg\_lambda” of 0.030. The accuracy of the traditional model can only reach 30.56% on the young population dataset. Its test  $R^2$  difference between the best and the worst performing folds is 3.54%.

For the middle-aged population, the average test  $R^2$  of the best model is 40.95% based on three single-fold test  $R^2$  41.30%, 40.59%, and 40.95%, while its average training  $R^2$  is 60.31% based on three single-fold training  $R^2$  60.31%, 60.35%, and 60.26%. The traditional model could only achieve 36.22% on the middle-aged population dataset. The best hyperparameter set of the model for middle-aged population contains “n\_estimators” of 3304, “learning\_rate” of 0.0162, “max\_depth” of 13, “subsample” of 0.896, “min\_child\_weight” of 1.762, “max\_delta\_step” of 0.100, “gamma” of 0.060, “reg\_alpha” of 10, and “reg\_lambda” of 0.003.

For the elderly population, the average test  $R^2$  of the best model is 40.46% based on three single-fold test  $R^2$  40.21%, 40.46%, and 40.72%, while its average training  $R^2$  is 67.24% based on three single-fold training  $R^2$  67.14%, 67.27%, and 67.30%. The best hyperparameter set of the model for elderly population contains “n\_estimators” of 540, “learning\_rate” of 0.0306, “max\_depth” of 13, “subsample” of 0.557, “min\_child\_weight” of 0.002, “max\_delta\_step” of 0.381, “gamma” of 1.205, “reg\_alpha” of 6.590, and “reg\_lambda” of 0.016. The ordinary least square model could reach 38.99% test accuracy in the 10-fold cross-validation on the elderly population dataset, which is still slightly lower than XGBoost’s performance.

Table 2 summarizes the test  $R^2$  in a 10-epoch 10-fold prediction. Each row in Table 2 is a summary of 100 test  $R^2$  computed based on 9:1 cross-validation. The model for a certain age group predicts the corresponding observations ten times and predicts the observations in other age groups 100 times. From an overall view, the model’s performance for the young population

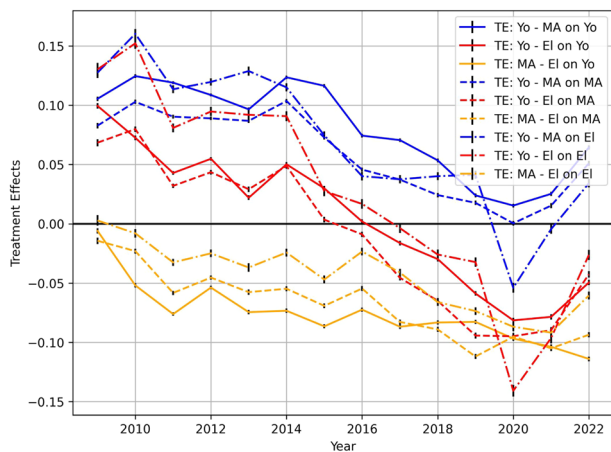
prediction is relatively worse than other age groups. The main reason is that the data size of the young population sub-dataset is large. The variation within this sub-dataset is more complex, which is more challenging to be completely grasped by models. However, the performance of each model is still acceptable, since the XGBoost models have significantly improved the prediction ability compared with linear models. The XGBoost algorithm could estimate variable importance by calculating the average gain, which is the improvement in the model’s loss function when a variable is used to split the data across all decision trees. The feature importance based on each age-group model is computed (listed in Supplementary Materials Table S3). The importance of the feeling of income is high in all three models. Specifically, for the young population, its importance is 14.91%; for the middle-aged population, its importance is 17.56%; and for the elderly population, the value is 8.20%. The results reflect that the middle-aged population is the most sensitive to the feeling of income.

*Overall base heterogeneity and treatment effects.* Table 3 illustrates the base heterogeneity and treatment effects of age on well-being. All three age groups All three age groups exhibit the lowest predicted well-being when subjected to external treatments corresponding to the middle-aged population. The treatment effects between young-population and middle-age-population treatment, the difference between average predictions of an age group treated as young population and middle-age population, are significantly positive. Specifically, individuals from any age group show higher well-being when externally treated as younger compared to being treated as middle-aged. The treatment effects between middle-aged and elderly populations are significantly negative. If a person is treated as a middle-aged population, her/his well-being status would be lower than the situation if she/he is treated as an elderly population. To summarize, the treatment for the middle-aged population significantly reduces human well-being. There is no significant difference between the young population treated as the young and elderly population. The well-being status that of middle-aged population is treated as young population is higher than that they are treated as the elderly population. The well-being status of the elderly population is treated as young population is lower than that they are treated as the elderly population. Regarding base heterogeneity effects, older age groups consistently demonstrated significantly higher intrinsic well-being compared to younger groups, with all pairwise differences reaching statistical significance.

**Table 3 Treatment and base heterogeneity effects.**

		Treatment			Treatment effect		
		Treated as young population	Treated as middle-age population	Treated as elderly population	$TE_{Yo-MA}^{mag}$	$TE_{MA-EI}^{mag}$	$TE_{Yo-EI}^{mag}$
Age group	Young Population	5.462	5.382	5.459	0.080***	-0.077***	0.003
	Middle-age Population	5.605	5.545	5.615	0.060***	-0.070***	-0.010***
	Elderly Population	5.749	5.681	5.726	0.068***	-0.045***	0.023***
Base heterogeneity effect	$BHE_{pag}^{Yo-MA}$	-0.143***	-0.163***	-0.156***			
	$BHE_{pag}^{MA-EI}$	-0.144***	-0.136***	-0.111***			
	$BHE_{pag}^{Yo-EI}$	-0.287***	-0.299***	-0.267***			

Yo represents the young population, MA represents the middle-aged population, and EI represents the elderly population. TE and BHE estimations follow Eqs. (34) and (35) mag and pag could be the young population, middle-aged population, and elderly population.  
 \*\*\*p-value < 1%; \*\*p-value < 5%; \*p-value < 10%.



**Fig. 2 Temporal variation of treatment effect.** TE: Yo-MA on Yo represents the treatment effect between the young population treated as the young and middle-aged population. Other explanations are similar.

**Variations of base heterogeneity and treatment effects**

*Temporal variation of base heterogeneity and treatment effects.* Figure 2 illustrates the temporal variation of the treatment effects. The treatment effects between treated as young population and treated as the middle-aged population in each age group are gradually decreasing temporally in terms of numbers. Since the treatment effects of switching from treatments for the young population to treatments for the middle-aged population are mainly positive, people in any age group treated as the young population are prone to have higher human well-being compared with the situations treated as the middle-aged population. The treatment for the young population and the treatment for the middle-aged population become closer. Table 4 reports the treatment effects of each year.

Additionally, the difference in any age group between the treatment for the middle-aged population and the treatment for the elderly population gradually becomes larger. People in any age group treated as an elderly population tend to have higher human well-being compared with those treated as the middle-aged population. The treatment effects between populations in any age group treated as a young population and treated as an elderly population are gradually reversing. Before 2015, the population treated as the young population tended to achieve a higher level of human well-being compared with the population treated as the elderly population. After 2017, the treatment effects

were reversed, i.e., the population treated as the elderly population could obtain higher human well-being. Indeed, in terms of mean values, the human well-being of the elderly group is higher than that of other age groups, and this difference is also increasing, as shown in Fig. 3.

To ensure that the temporal trends in treatment effects are not random perturbations from a baseline, we compute the standard errors of the temporal mean for each case based on the individual treatment effects. The small standard errors and concentrated distribution demonstrate the reliability and robustness of the temporal variation of the treatment effects. In the process of data processing, model training, and result prediction, we control the potential impacts of time. At the data level, Gallup’s long-term, consistent and random sampling method avoids sampling errors as much as possible. Furthermore, their questionnaire remains the same from the first wave in 2005. During the data preprocessing process, we also retain most of the data to ensure that the distribution does not change significantly. In the model training phase, we take the wave as an independent variable to reduce the impact of the year. In the result prediction process, we employ a 10-epoch 10-fold prediction, that is, all individuals’ well-being is the mean of results predicted by 10 different models with the same hyperparameter setting. The aggregation of multiple models increases the robustness of the results.

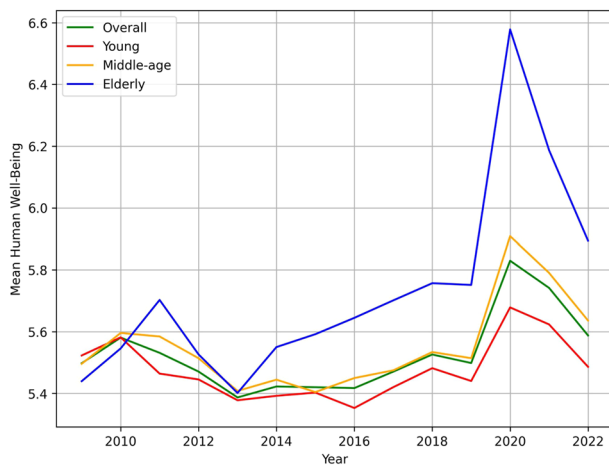
Figure 4 demonstrates the temporal variation of the base heterogeneity effects. The base heterogeneity effects between the young population and the middle-aged are temporally stable, without significant and clear change trends. The base heterogeneity effects between the young population and elderly population, and between the middle-aged population and elderly population, become larger temporally, accompanied by a certain level of vibration. Table 5 lists the base heterogeneity effects of each year. Those temporal variations in base heterogeneity effects are mainly caused by the change in the elderly population, supported by Fig. 3.

*Country-level variation of treatment effects.* The country-level treatment effects are summarized in Table 6. We have roughly divided these countries into five styles based on how harshly they treat each age group: Younger Enjoy, Older Enjoy, Middle-age Suffer, Middle-age Enjoy, and No Trend. These typologies illustrate generalized patterns in the relationship between predicted well-being and age, as depicted conceptually in Fig. 5. According to our computational definition of treatment effects, they can be represented as three vectors in the coordinate system in Fig. 5. The red, blue, and green vectors represent the shifts of external treatments from the treatments for the young population to the

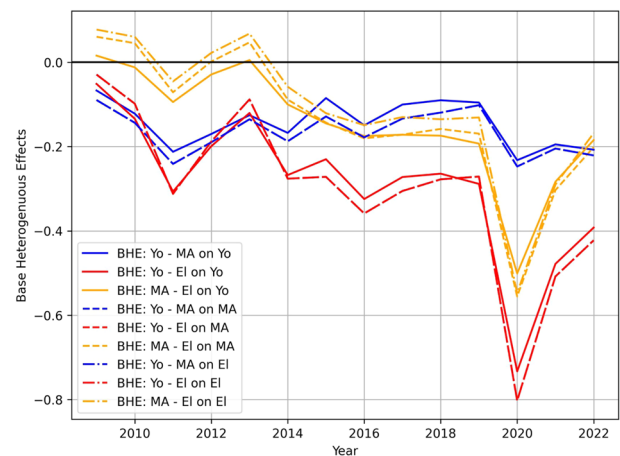
**Table 4 Temporal variation of treatment effects.**

Year	$TE_{Yo-MA}^{Yo}$	$TE_{Yo-EI}^{Yo}$	$TE_{MA-EI}^{Yo}$	$TE_{Yo-MA}^{MA}$	$TE_{Yo-EI}^{MA}$	$TE_{MA-EI}^{MA}$	$TE_{Yo-MA}^{EI}$	$TE_{Yo-EI}^{EI}$	$TE_{MA-EI}^{EI}$
2009	0.105***	0.099***	-0.006	0.083***	0.068***	-0.014	0.128***	0.130***	0.003
2010	0.125***	0.073***	-0.052***	0.103***	0.080***	-0.023**	0.160***	0.152***	-0.008
2011	0.119***	0.043***	-0.076***	0.090***	0.032***	-0.058***	0.113***	0.081***	-0.033*
2012	0.109***	0.055***	-0.054***	0.089***	0.044***	-0.045***	0.120***	0.095***	-0.025
2013	0.097***	0.022***	-0.075***	0.087***	0.029***	-0.058***	0.129***	0.092***	-0.037*
2014	0.124***	0.050***	-0.073***	0.104***	0.049***	-0.055***	0.115***	0.091***	-0.024
2015	0.116***	0.030***	-0.087***	0.073***	0.003	-0.069***	0.075***	0.028	-0.047**
2016	0.074***	0.002	-0.072***	0.046***	-0.009	-0.055***	0.040**	0.017	-0.023
2017	0.071***	-0.016**	-0.087***	0.037***	-0.046***	-0.083***	0.038**	-0.004	-0.041**
2018	0.054***	-0.030***	-0.083***	0.024**	-0.065***	-0.089***	0.040**	-0.026	-0.066***
2019	0.024***	-0.059***	-0.083***	0.018*	-0.094***	-0.112***	0.041***	-0.032**	-0.074***
2020	0.015**	-0.081***	-0.097***	0	-0.095***	-0.095***	-0.054***	-0.141***	-0.087***
2021	0.025***	-0.079***	-0.104***	0.015	-0.090***	-0.105***	-0.005	-0.096***	-0.092***
2022	0.064***	-0.050***	-0.114***	0.051***	-0.043***	-0.094***	0.034**	-0.026	-0.060***

Yo represents the young population, MA represents the middle-aged population, and EI represents the elderly population. TE estimations follow Eq. (34).  
 \*\*\*p-value < 1%; \*\*p-value < 5%; \*p-value < 10%.



**Fig. 3** The average human well-being variation of each age group.



**Fig. 4 Temporal variation of base heterogeneity effect.** BHE: Yo-MA on Yo represents the base heterogeneity effect between the young population treated as the young and middle-aged population. Other explanations are similar).

treatments for the middle-aged population, from the treatments for the middle-aged population to the treatments for the elderly population, and from the treatments for the young population to the treatments for the elderly population, respectively. The projections of these vectors on the y-axis are the treatment effects. If the treatment effect is not statistically significant, the corresponding vector is level. If two or more values are significantly positive, the country would be classified as Younger Enjoy, that is, people who are treated as younger people are prone to have a better well-being status. In the real calculation, Younger Enjoy includes three cases, illustrated as Fig. 5a–c. Figure 5a shows that all three treatment effects are significantly positive, while Fig. 5b, c have two significantly positive values and one insignificant value. Conversely, if two or more values are significantly negative, the country would be classified as Older Enjoy, shown in Fig. 5d–f, meaning that people who are considered relatively older tend to have better well-being. If the treatment effects between the young population and middle-aged population are negative and the treatment effects between the middle-aged population and elderly population are positive, the country should be labeled as Middle-aged Enjoy, as demonstrated in Fig. 5g. At present, we do not consider the treatment effect between the young population and the elderly population. If people are considered middle-aged, they tend to have the highest levels of well-being. If, in a country, the treatment effects between the young population and middle-

aged population are positive and the treatment effects between the middle-aged population and elderly population are negative, the country should be labeled as Middle-age Suffer, shown in Fig. 5h. In other words, if the people are treated as the middle-aged population, they tend to have the lowest well-being. Among three values of the country-level treatment effects between three age groups in a certain country, if two or more of them are not significant, then that country would be classified as No Trend, i.e., no significant trend exists, illustrated by Fig. 5i. Among 165 countries, the counts of each style are listed as follows: 51 of Older Enjoy, 49 of Younger Enjoy, 48 of Middle-age Suffer, 10 of Middle-age Enjoy, and 6 of No Trend. Although Middle-age Suffer is not absolutely mainstream in those countries, treatments for the middle-aged population are basically in a relatively unfavorable state.

**Discussion**

Based on a large global dataset and cutting-edge technologies, the reasons for the low human well-being in middle-aged people are investigated. In the GWP survey, we employ ESTEM based on machine-learning technology to analyze more than 1.9 million observations from 168 countries or regions during 2009–2022.

**Table 5 Temporal variation of base heterogenous effects.**

Year	$BHE_{Yo-MA}^{Yo-MA}$	$BHE_{Yo-EI}^{Yo-EI}$	$BHE_{MA-EI}^{MA-EI}$	$BHE_{Yo-MA}^{Yo-MA}$	$BHE_{Yo-EI}^{Yo-EI}$	$BHE_{MA-EI}^{MA-EI}$	$BHE_{Yo-MA}^{Yo-MA}$	$BHE_{Yo-EI}^{Yo-EI}$	$BHE_{MA-EI}^{MA-EI}$
2009	-0.067***	-0.051***	0.016	-0.090***	-0.029*	0.060***	-0.090***	-0.029	0.078***
2010	-0.122***	-0.134***	-0.012	-0.144***	-0.098***	0.045***	-0.144***	-0.098***	0.060***
2011	-0.212***	-0.307***	-0.094***	-0.241***	-0.312***	-0.071***	-0.241***	-0.312***	-0.045***
2012	-0.170***	-0.199***	-0.029**	-0.190***	-0.188***	0.002	-0.190***	-0.188***	0.022*
2013	-0.126***	-0.120***	0.005	-0.135***	-0.088***	0.047***	-0.135***	-0.088***	0.068***
2014	-0.168***	-0.268***	-0.100***	-0.188***	-0.276***	-0.089***	-0.188***	-0.276***	-0.058***
2015	-0.085***	-0.230***	-0.145***	-0.129***	-0.272***	-0.143***	-0.129***	-0.272***	-0.121***
2016	-0.150***	-0.324***	-0.175***	-0.178***	-0.359***	-0.180***	-0.178***	-0.359***	-0.149***
2017	-0.100***	-0.272***	-0.172***	-0.134***	-0.305***	-0.172***	-0.134***	-0.305***	-0.130***
2018	-0.090***	-0.264***	-0.174***	-0.119***	-0.278***	-0.158***	-0.119***	-0.278***	-0.135***
2019	-0.096***	-0.289***	-0.193***	-0.102***	-0.271***	-0.169***	-0.102***	-0.271***	-0.131***
2020	-0.232***	-0.733***	-0.501***	-0.247***	-0.803***	-0.556***	-0.247***	-0.803***	-0.547***
2021	-0.195***	-0.478***	-0.284***	-0.205***	-0.508***	-0.303***	-0.205***	-0.508***	-0.290***
2022	-0.208***	-0.392***	-0.185***	-0.221***	-0.422***	-0.201***	-0.221***	-0.422***	-0.168***

Yo represents the young population, MA represents the middle-aged population, and EI represents the elderly population. BHE estimations follow Eq. (35).  
 \*\*\*p-value < 1%; \*\*p-value < 5%; \*p-value < 1%.

We contribute to the literature in the following several aspects. First, with high-accuracy machine learning models on extensive global data, our findings empirically confirm the U-shaped relationship between age and subjective well-being, wherein middle-aged individuals consistently experience the lowest levels of well-being. Second, middle-aged people receive the worst external treatment, and the external treatments for the young and elderly populations are relatively favorable and similar in nature. Third, the base heterogeneity difference shows that older people generally experience higher human well-being naturally and inherently. Fourth, the external treatments for young and middle-aged people are gradually becoming more stringent compared with those for the elderly. Fifth, according to the temporal variation of the base heterogeneity difference between each age group, older people are inherently more likely to achieve higher levels of human well-being temporally. This study demonstrates that external treatments may be an important factor in causing a midlife crisis. Additionally, we also explore the trends of treatment effects and base heterogeneity effects among several age groups. Our study provides insights into human well-being variation among age groups. These findings can contribute to society by informing policies and programs aimed at improving quality of life, tailoring social services to meet the specific needs of various age groups, and fostering a better understanding of how well-being evolves with age.

The empirical relationship between age and human well-being in this study is U-shaped, corroborating findings from multiple previous studies (Blanchflower, 2021; Blanchflower and Oswald, 2008; Diener et al. 2018; Stone et al. 2010). Although the U-shaped relationship is widely accepted, it is still controversial, as other studies have found inverted U-shaped and linear relationships (Easterlin, 2006; Frijters and Beatton, 2012). Our results support the validity of U-shaped associations at the global scale using big data and relatively high-precision models. SHAP objectively isolates the contribution of age to well-being, because SHAP values represent a fairly distributed contribution of each input variable to the model’s output (Lundberg et al. 2020). Consequently, our findings are statistically stable and reliable, having minimized interaction effects among independent variables. Confirming these age-related differences is also a prerequisite for applying the ESTEM approach.

Two prevailing explanations are typically cited for the observed U-shaped pattern. First, midlife crises and the various symptoms accompanying them are the critical reasons for the lower human well-being of middle-aged people (Giuntella et al. 2023; Lachman,

2015). Specifically, middle-aged people experience more sleeping problems (Krueger and Friedman, 2009), concentration difficulties, financial stress (Plagnol, 2011), and extreme depression (Giuntella et al. 2023), which could significantly reduce human well-being. This midlife decline aligns with midlife crisis theories, i.e., midlife is a time of role strain and identity transition marked by conflicting demands and life reassessment (Lachman, 2015). From a psychological perspective, individuals may experience a mismatch between their earlier aspirations and current achievements, triggering emotional distress (Hardie, 2014). Sociologically, midlife often coincides with overlapping roles, which intensifies stress and reduces autonomy, the key components of subjective well-being (Giuntella et al. 2023; Lachman, 2015). This is also consistent with our feature importance analysis, that is, the middle-aged population is impacted by the feeling of income the most. Second, people born in some decades are more likely to achieve higher well-being (Shu et al. 2023; Sutin et al. 2013; Van Landeghem, 2012). For instance, Shu et al. (2023) indicate that individuals born between 1956 and 1961 faced challenges at several critical stages in their lives, including education, employment, economic stability, and social connections, and they report lower human well-being compared to other cohorts. This cohort happens to be middle-aged people. Sutin et al. (2013) display that while well-being generally increased with age for everyone, those cohorts that experienced the economic hardships of the early 20th century reported lower well-being compared to those born during more prosperous times. However, these explanations do not distinguish whether the causes of this U-shaped relationship are externally or inherently idiosyncratic. For example, the previous studies do not illustrate whether mental disorders are triggered by external pressure and treatments, or caused by the vulnerable status of a certain life stage. Our study proves that external treatments for three age groups are U-shaped. Specifically, the middle-aged population is treated relatively more harshly.

The trend of base heterogeneity difference demonstrates that older people are prone to experience higher human well-being naturally and inherently. Older people have a higher age but also live longer and spend more time achieving the goals of their possible lives. Additionally, a well-built theory, socioemotional selectivity theory, declaims that as individuals age, they prioritize emotionally meaningful goals and relationships (Carstensen et al. 2000; Löckenhoff and Carstensen, 2004). As shown in our results, the elderly population is the least influenced by the feeling of income. The goal shift from the material aspect to the spiritual

**Table 6 Country-level treatment effects of age.**

Country	Country name	CTE <sup>Country</sup> <sub>Yo-MA</sub>	CTE <sup>Country</sup> <sub>MA-EI</sub>	CTE <sup>Country</sup> <sub>Yo-EI</sub>	Style
AFG	Afghanistan	0.026*	0.185***	0.211***	Younger Enjoy
AGO	Angola	0.211***	0.576***	0.787***	Younger Enjoy
ALB	Albania	0.231***	-0.105***	0.126***	Middle-age Suffer
ARE	United Arab Emirates	0.009	-0.385***	-0.376***	Older Enjoy
ARG	Argentina	0.271***	0.002	0.273***	Younger Enjoy
ARM	Armenia	0.201***	0.044***	0.245***	Younger Enjoy
AUS	Australia	-0.311***	-0.224***	-0.535***	Older Enjoy
AUT	Austria	0.038***	-0.037***	0.002	Middle-age Suffer
AZE	Azerbaijan	0.127***	0.117***	0.245***	Younger Enjoy
BDI	Burundi	-0.022	0.077***	0.055***	Younger Enjoy
BEL	Belgium	0.028***	-0.016	0.012	No Trend
BEN	Benin	0.017	-0.115***	-0.097***	Older Enjoy
BFA	Burkina Faso	-0.049***	-0.189***	-0.237***	Older Enjoy
BGD	Bangladesh	-0.086***	-0.165***	-0.250***	Older Enjoy
BGR	Bulgaria	0.090***	0.070***	0.161***	Younger Enjoy
BHR	Bahrain	0.035**	-0.196***	-0.161***	Middle-age Suffer
BIH	Bosnia and Herzegovina	0.298***	0.129***	0.427***	Younger Enjoy
BLR	Belarus	0.135***	0.213***	0.349***	Younger Enjoy
BLZ	Belize	-0.075	-0.171***	-0.246***	Older Enjoy
BOL	Bolivia	0.197***	-0.301***	-0.104***	Middle-age Suffer
BRA	Brazil	0.094***	-0.258***	-0.164***	Middle-age Suffer
BTN	Bhutan	0.02	-0.030**	-0.01	No Trend
BWA	Botswana	0.257***	0.123***	0.380***	Younger Enjoy
CAF	Central African Republic	0.058***	-0.100***	-0.042**	Middle-age Suffer
CAN	Canada	-0.259***	-0.163***	-0.423***	Older Enjoy
CHE	Switzerland	-0.040***	-0.076***	-0.116***	Older Enjoy
CHL	Chile	0.283***	0.095***	0.378***	Younger Enjoy
CHN	China	0.009	-0.332***	-0.323***	Older Enjoy
CIV	Ivory Coast	0.019	-0.157***	-0.138***	Older Enjoy
CMR	Cameroon	0.174***	-0.061***	0.113***	Middle-age Suffer
COD	Democratic Republic of the Congo	0.070***	-0.222***	-0.152***	Middle-age Suffer
COG	Republic of the Congo	0.011	0.108***	0.119***	Younger Enjoy
COL	Colombia	0.072***	-0.025*	0.046***	Middle-age Suffer
COM	Comoros	0.007	0.01	0.017	No Trend
CRI	Costa Rica	0.031**	-0.054***	-0.023*	Middle-age Suffer
CYP	Cyprus	0.178***	-0.119***	0.059***	Middle-age Suffer
CZE	Czechia	0.212***	0.170***	0.382***	Younger Enjoy
DEU	Germany	0.122***	0.025***	0.146***	Younger Enjoy
DNK	Denmark	-0.072***	-0.253***	-0.325***	Older Enjoy
DOM	Dominican Republic	0.412***	-0.357***	0.055***	Middle-age Suffer
DZA	Algeria	0.002	0.057***	0.059***	Younger Enjoy
ECU	Ecuador	0.341***	0.098***	0.439***	Younger Enjoy
EGY	Egypt	0.112***	-0.151***	-0.039***	Middle-age Suffer
ESP	Spain	0.098***	0.059***	0.157***	Younger Enjoy
EST	Estonia	0.147***	0.006	0.152***	Younger Enjoy
ETH	Ethiopia	0.001	-0.130***	-0.129***	Older Enjoy
FIN	Finland	-0.019*	0.086***	0.067***	Middle-age Enjoy
FRA	France	0.092***	-0.006	0.087***	Younger Enjoy
GAB	Gabon	0.184***	0.241***	0.425***	Younger Enjoy
GBR	United Kingdom	-0.027***	-0.136***	-0.163***	Older Enjoy
GEO	Georgia	0.171***	-0.029*	0.141***	Middle-age Suffer
GHA	Ghana	-0.014	-0.011	-0.025*	No Trend
GIN	Guinea	-0.062***	-0.097***	-0.159***	Older Enjoy
GMB	Gambia	-0.084***	0.397***	0.312***	Middle-age Enjoy
GRC	Greece	0.375***	0.119***	0.494***	Younger Enjoy
GTM	Guatemala	0.127***	-0.102***	0.025*	Middle-age Suffer
HKG	Hong Kong	-0.113***	0.024	-0.089***	Older Enjoy
HND	Honduras	0.263***	-0.139***	0.124***	Middle-age Suffer
HRV	Croatia	0.260***	0.257***	0.517***	Younger Enjoy
HTI	Haiti	0.101***	0.001	0.102***	Younger Enjoy
HUN	Hungary	0.233***	0.052***	0.286***	Younger Enjoy
IDN	Indonesia	0.045***	-0.346***	-0.301***	Middle-age Suffer
IND	India	-0.012*	-0.087***	-0.099***	Older Enjoy
IRL	Ireland	0.049***	-0.214***	-0.165***	Middle-age Suffer
IRN	Iran	0.294***	-0.053***	0.241***	Middle-age Suffer
IRQ	Iraq	0.019	-0.042***	-0.024*	Older Enjoy

**Table 6 (continued)**

Country	Country name	CTE <sub>Country</sub> Yo-MA	CTE <sub>Country</sub> MA-EI	CTE <sub>Country</sub> Yo-EI	Style
ISL	Iceland	-0.044***	-0.058***	-0.102***	Older Enjoy
ISR	Israel	0.177***	0.048***	0.226***	Younger Enjoy
ITA	Italy	0.189***	0.078***	0.268***	Younger Enjoy
JAM	Jamaica	-0.081***	-0.244***	-0.325***	Older Enjoy
JOR	Jordan	0.094***	-0.106***	-0.011	Middle-age Suffer
JPN	Japan	-0.149***	0.019	-0.131***	Older Enjoy
KAZ	Kazakhstan	-0.073***	-0.188***	-0.261***	Older Enjoy
KEN	Kenya	0.122***	0.143***	0.265***	Younger Enjoy
KGZ	Kyrgyzstan	0.017	-0.185***	-0.169***	Older Enjoy
KHM	Cambodia	-0.160***	-0.294***	-0.454***	Older Enjoy
KOR	South Korea	0.354***	-0.036**	0.319***	Middle-age Suffer
KWT	Kuwait	-0.089***	-0.395***	-0.484***	Older Enjoy
LAO	Laos	-0.160***	-0.075***	-0.236***	Older Enjoy
LBN	Lebanon	0.266***	-0.068**	0.198***	Middle-age Suffer
LBR	Liberia	0.01	-0.041*	-0.031	No Trend
LBY	Libya	0.007	-0.426***	-0.418***	Older Enjoy
LKA	Sri Lanka	-0.136***	-0.148***	-0.284***	Older Enjoy
LSO	Lesotho	0.392***	-0.026	0.366***	Younger Enjoy
LTU	Lithuania	0.153***	0.063***	0.215***	Younger Enjoy
LUX	Luxembourg	0.064***	-0.054***	0.01	Middle-age Suffer
LVA	Latvia	0.139***	-0.02	0.120***	Younger Enjoy
MAR	Morocco	-0.081***	-0.163***	-0.245***	Older Enjoy
MDA	Moldova	0.230***	0.219***	0.449***	Younger Enjoy
MDG	Madagascar	-0.01	-0.104***	-0.114***	Older Enjoy
MDV	Maldives	0.082*	-0.299***	-0.217***	Middle-age Suffer
MEX	Mexico	0.075***	0.236***	0.311***	Younger Enjoy
MKD	North Macedonia	0.251***	-0.029*	0.222***	Middle-age Suffer
MLI	Mali	-0.041***	0.095***	0.054***	Middle-age Enjoy
MLT	Malta	0.115***	-0.086***	0.029**	Middle-age Suffer
MMR	Myanmar	-0.302***	-0.339***	-0.641***	Older Enjoy
MNE	Montenegro	0.276***	-0.062***	0.214***	Middle-age Suffer
MNG	Mongolia	-0.240***	-0.123***	-0.363***	Older Enjoy
MOZ	Mozambique	0.056**	0.170***	0.226***	Younger Enjoy
MRT	Mauritania	-0.125***	-0.054*	-0.179***	Older Enjoy
MUS	Mauritius	-0.055***	-0.376***	-0.431***	Older Enjoy
MWI	Malawi	0.259***	-0.166***	0.093***	Middle-age Suffer
MYS	Malaysia	0.049***	-0.032***	0.017	Middle-age Suffer
NAM	Namibia	0.062***	0.104***	0.166***	Younger Enjoy
NER	Niger	-0.043***	-0.156***	-0.199***	Older Enjoy
NGA	Nigeria	-0.044***	0.224***	0.181***	Middle-age Enjoy
NIC	Nicaragua	0.353***	-0.286***	0.067***	Middle-age Suffer
NLD	Netherlands	-0.001	-0.086***	-0.087***	Older Enjoy
NOR	Norway	-0.050***	-0.139***	-0.190***	Older Enjoy
NPL	Nepal	0.035***	-0.258***	-0.223***	Middle-age Suffer
NZL	New Zealand	-0.193***	-0.301***	-0.494***	Older Enjoy
PAK	Pakistan	-0.042***	0.150***	0.108***	Middle-age Enjoy
PAN	Panama	0.149***	0.001	0.150***	Younger Enjoy
PER	Peru	0.026**	0.092***	0.118***	Younger Enjoy
PHL	Philippines	0.063***	-0.295***	-0.232***	Middle-age Suffer
POL	Poland	0.163***	-0.102***	0.061***	Middle-age Suffer
PRI	Puerto Rico	-0.044	0.238***	0.195***	Younger Enjoy
PRT	Portugal	0.238***	0.038**	0.276***	Younger Enjoy
PRY	Paraguay	0.269***	-0.039***	0.230***	Middle-age Suffer
PSE	West Bank and Gaza	0.235***	-0.063***	0.172***	Middle-age Suffer
QAT	Qatar	-0.159***	0.029	-0.130***	Older Enjoy
ROU	Romania	0.316***	-0.033*	0.283***	Middle-age Suffer
RUS	Russia	0.194***	0.109***	0.303***	Younger Enjoy
RWA	Rwanda	0.028**	0.104***	0.131***	Younger Enjoy
SAU	Saudi Arabia	-0.241***	-0.236***	-0.477***	Older Enjoy
SDN	Sudan	-0.001	0.030*	0.029*	Younger Enjoy
SEN	Senegal	0.078***	-0.176***	-0.097***	Middle-age Suffer
SGP	Singapore	-0.056***	-0.092***	-0.148***	Older Enjoy
SLE	Sierra Leone	-0.196***	0.425***	0.229***	Middle-age Enjoy
SLV	El Salvador	0.334***	-0.166***	0.167***	Middle-age Suffer
SOM	Somalia	0.386***	0.071***	0.457***	Younger Enjoy
SRB	Serbia	0.295***	0.021	0.317***	Younger Enjoy

**Table 6 (continued)**

Country	Country name	CTE <sub>Yo-MA</sub> <sup>Country</sup>	CTE <sub>MA-EI</sub> <sup>Country</sup>	CTE <sub>Yo-EI</sub> <sup>Country</sup>	Style
SSD	South Sudan	0.194***	-0.342***	-0.148***	Middle-age Suffer
SUR	Suriname	-0.112**	0.103***	-0.009	Middle-age Enjoy
SVK	Slovakia	0.166***	0.129***	0.295***	Younger Enjoy
SVN	Slovenia	0.366***	-0.083***	0.284***	Middle-age Suffer
SWE	Sweden	-0.200***	-0.109***	-0.309***	Older Enjoy
SWZ	Eswatini	0.330***	-0.140***	0.190***	Middle-age Suffer
SYR	Syria	-0.103***	0.192***	0.089***	Middle-age Enjoy
TCD	Chad	-0.013	-0.622***	-0.635***	Older Enjoy
TGO	Togo	-0.106***	0.089***	-0.017	Middle-age Enjoy
THA	Thailand	0.158***	-0.242***	-0.084***	Middle-age Suffer
TJK	Tajikistan	0.082***	0.098***	0.180***	Younger Enjoy
TKM	Turkmenistan	-0.066***	0	-0.066***	Older Enjoy
TTO	Trinidad and Tobago	-0.448***	0.073	-0.375***	Older Enjoy
TUN	Tunisia	0.012	-0.171***	-0.159***	Older Enjoy
TUR	Turkey	0.063***	-0.109***	-0.047***	Middle-age Suffer
TWN	Taiwan	-0.019	-0.106***	-0.124***	Older Enjoy
TZA	Tanzania	-0.138***	-0.060***	-0.197***	Older Enjoy
UGA	Uganda	0.012	0.105***	0.117***	Younger Enjoy
UKR	Ukraine	0.242***	0.173***	0.415***	Younger Enjoy
URY	Uruguay	0.194***	-0.091***	0.102***	Middle-age Suffer
USA	United States of America	-0.125***	-0.262***	-0.387***	Older Enjoy
UZB	Uzbekistan	-0.272***	-0.669***	-0.941***	Older Enjoy
VEN	Venezuela	0.137***	0.047**	0.184***	Younger Enjoy
VNM	Vietnam	0.043***	-0.154***	-0.111***	Middle-age Suffer
XKX	Kosovo	0.188***	-0.136***	0.051***	Middle-age Suffer
XNC		0.045**	-0.453***	-0.408***	Middle-age Suffer
XNK		-0.02	-0.01	-0.03	No Trend
XSR		0.213***	0.058***	0.271***	Younger Enjoy
YEM	Yemen	0.178***	-0.081***	0.098***	Middle-age Suffer
ZAF	South Africa	-0.031**	-0.178***	-0.209***	Older Enjoy
ZMB	Zambia	-0.091***	0.238***	0.147***	Middle-age Enjoy
ZWE	Zimbabwe	0.004	-0.163***	-0.159***	Older Enjoy

Yo represents the young population, MA represents the middle-aged population, and EI represents the elderly population. Country-level treatment effects estimations follow Eq. (37).  
 \*\*\*p-value < 1%; \*\*p-value < 5%; \*p-value < 10%.

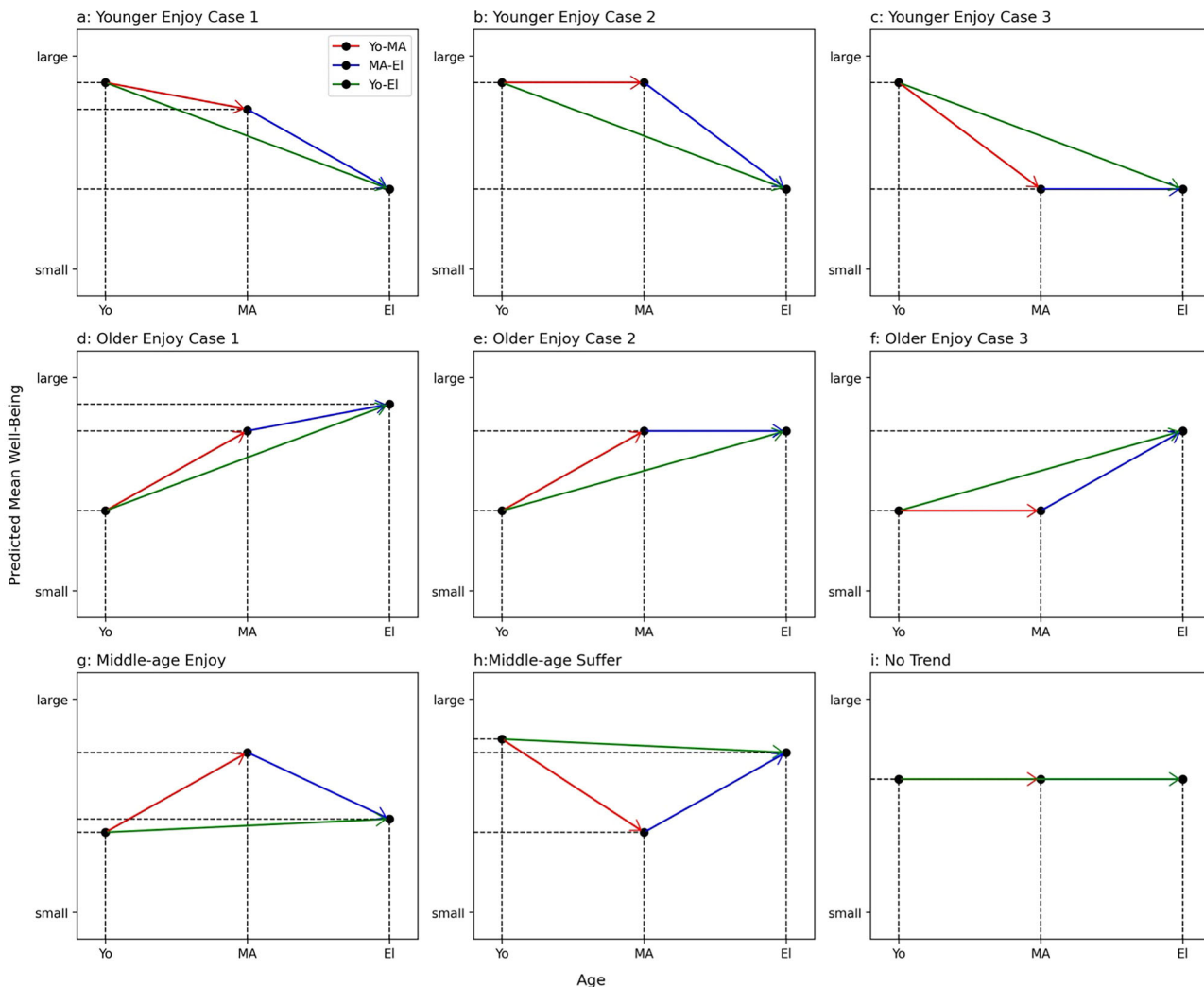
aspect that occurs with aging objectively reduces personal stress and inherently improves well-being, which is a potential reason (Löckenhoff and Carstensen, 2004). On the one hand, aging consistently reduces human well-being (Li and Managi, 2023), since various health problems, attitudes toward life change, and social relationships come with aging (Bamidis et al. 2014; Luu and Palczewski, 2018; Steptoe et al. 2015; Stone et al. 2010). On the other hand, the effect of time could not be ignored. Because the SWB method in this research is the Cantril ladder, the elderly population has more time to climb the higher steps. In this way, inherently, elderly people are prone to have a higher global human well-being evaluation. It should be noted that in this study, we only investigate global well-being based on life evaluation; the findings might be inconsistent with the research using other well-being indicators, such as hedonic well-being (Stone et al. 2010) and eudemonic well-being (Bussière et al. 2021). The base heterogeneity effects could be regarded as the effects of time. Combining previous studies and our findings, the effects of time and aging are completely in opposite directions. Because they cancel each other out, the empirical result is U-shaped.

Temporally, the treatments for the young and middle-aged population are becoming similar, and the differences in treatments between the elderly population and other age groups are gradually increasing. The base heterogeneity effects between the elderly population and the other two age groups become larger gradually. Because a cohort of people becomes an elderly population, who tend to experience higher human well-being, the heterogeneity effects vary temporally. This supports previous studies showing that there are indeed gaps between generations

(Shu et al. 2023). Additionally, although the situation varies from country to country, middle-aged population treatment in most countries is in a relatively unfavorable situation. Therefore, adopting strategies that can reduce stress in middle-aged people is an important means to improve their level of well-being.

Our innovation in the method is also noteworthy. First, we use tree-based machine learning methods to replace the traditional regression method, which is more suitable for grasping non-linear relationships. Second, the SHAP method could fairly distribute the contribution to each independent variable. It is an essential way to illustrate the relationship between age and well-being accurately at the statistical and social science levels. In fact, our model accuracy significantly exceeds previous studies based on linear regressions or similar technologies, as normally their  $R^2$  is <25%, e.g., Blanchflower and Graham (2022); Blanchflower and Piper (2022). Seemingly, we just slightly improved the accuracy compared with previous studies based on regression technologies. However, the  $R^2$  in the previous studies is the training accuracy, i.e., the regression method in that way is completely unable to monitor and avoid overfitting. The test  $R^2$  is the effective and necessary metric to check whether the model really grasps the relationship. If the generalization of models in the ESTEM is poor, the counterfactual prediction would be totally unreliable. When data is abundant, using more complex machine learning models can effectively reduce the impact of this problem.

Society and governments should pay more attention to the middle-aged population. They are the backbone of social and family development; therefore, they also bear relatively more pressure. Inherently, the middle-aged population should feel



**Fig. 5** Schematic diagrams of Younger Enjoy, Older Enjoy, Middle-age Suffer, Middle-age Enjoy, and No Trend.

achieving a better life than young people, but the responsibilities or expectations have put them into a midlife crisis. Policies should take into account the needs and dilemmas of middle-aged people in order to achieve a sustainable society. First, governments and enterprises should promote vocational retraining programs specifically targeting middle-aged workers to enhance their employability and adaptability in rapidly evolving labor markets (Shu et al. 2023). In contemporary society, accelerating the adoption of AI-driven productivity tools and facilitating the integration of AI technologies among middle-aged workers constitute important policy priorities. Second, flexible working hour policies could help reduce work-life conflict, particularly for those balancing career demands with caregiving responsibilities (Giuntella et al. 2023; Lachman, 2015). Third, mental health support services tailored to midlife challenges should be expanded as a welfare (Krueger and Friedman, 2009). Based on the Gallup survey, among three age groups, in 83 countries, the middle-aged population is the least satisfied with the healthcare system, which is the absolute majority, as shown in Supplementary Materials Table S4. Additionally, unemployment insurance programs should be optimized to ensure that middle-aged individuals facing job loss receive adequate financial and reemployment support (Wanberg et al. 2016). In summary, interventions should be considered to enhance well-being by aligning social support with the unique roles and pressures of midlife. Additionally, we are aware that the harsh treatment is

gradually spreading to younger people. A life situation of the young population that becomes increasingly difficult due to external factors can lead to a variety of problems, such as a declining birth rate, a lower marriage rate, and a relatively weak economic environment. Moreover, beyond the general policy implications for the global situation, context-specific policy interventions are required because the treatment effects vary across countries. In “Older Enjoy” countries such as Japan, where the elderly benefit from favorable external conditions, the focus should be on strengthening the fairness of intergenerational resource distribution (Murayama et al. 2019). Redirecting some welfare resources or employment incentives toward the middle-aged population could improve overall well-being, as the middle-aged individuals are often in their peak working and caregiving years but may receive comparatively fewer benefits. In contrast, in “Middle-age Suffer” countries, like India, where the middle-aged receive the harshest external treatment, priority should be given to labor protection policies (Mansoor and O’Neill, 2021). As those countries are mainly developing countries, the workplace conditions, mid-career upskilling programs, and mental health services are under formulation. Improving relevant systems in such a country can effectively reduce the oppression of middle-aged people. The differentiated strategies are supported by the country-level treatment effects identified in our study to boost human well-being more equitably across age groups.

Several limitations of this study should be noted, although we adopt advanced technologies and the largest global dataset to probe the reasons for the lower human well-being among the middle-aged population. First, there are relatively large differences in sample sizes among the three age groups. This results in relatively poor generalization ability of some models. The smaller sample size of the elderly group may result in insufficient stability of model predictions. Second, some important variables, such as educational background, are not obtained in the dataset. Furthermore, most variables in the analysis are binary, which provides insufficient information in a way. Third, limited by the computing ability of the hardware and data size, we only divide the total dataset into three age groups. If we set more age groups, more interesting findings might be obtained. To enhance future studies on human well-being across different age groups, future studies should consider several improvements. First, it is recommended to balance the sample sizes across age groups to enhance the model performance in ESTEM. Stratified sampling techniques or data augmentation methods could be employed to ensure a more balanced age group representation, particularly to improve the robustness of predictions for elderly subgroups. Second, future studies could include a broader range of variables, such as years of education or vocational training, to provide deeper insights into well-being determinants. Third, using more detailed variable types beyond binary options is a way to capture detailed data effectively. Additionally, adopting more sophisticated statistical or machine learning methods could address complex datasets and reveal more intricate patterns. Lastly, expanding the age categorizations and incorporating longitudinal and cross-cultural data could uncover dynamic trends and cultural influences on well-being.

## Conclusions

This study provides robust empirical evidence of a global U-shaped relationship between age and well-being statistically and empirically, revealing that middle-aged individuals experience the lowest levels of well-being levels across diverse contexts. By leveraging extensive global data and advanced machine-learning methods, we have demonstrated that external social treatments significantly disadvantage the middle-aged population, whereas elderly individuals inherently experience higher levels of well-being. These findings underscore the critical role that external conditions play in exacerbating midlife challenges and highlight the importance of tailored policy responses. Given the apparent disparities in external treatment among age groups and their evolution over time, context-specific interventions should be prioritized. Policies targeting the middle-aged should emphasize vocational retraining, flexible work arrangements, improved mental health support, and strengthened social security systems, thereby addressing the unique stressors associated with midlife. Additionally, as younger individuals begin to face increasingly harsh external conditions, proactive measures are needed to mitigate emerging societal issues such as declining birth rates and economic instability. To optimize well-being across the lifespan, policies should be dynamic and adapt to the shifting base differences among age groups, ensuring that interventions are timely and tailored to the evolving needs of each demographic. This strategy not only promotes a more equitable distribution of resources but also supports the overall goal of enhancing human well-being in an aging society.

## Data availability

Data are available from the corresponding author on reasonable request.

Received: 1 February 2025; Accepted: 6 August 2025;

Published online: 22 August 2025

## References

- Balcilar M, Gupta R, Miller SM (2015) Regime switching model of US crude oil and stock market prices: 1859 to 2013. *Energy Econ* 49:317–327
- Bamidis PD, Vivas AB, Styliadis C, Frantidis C, Klados M, Schlee W, Siountas A, Papageorgiou SG (2014) A review of physical and cognitive interventions in aging. *Neurosci Biobehav Rev* 44:206–220
- Bentéjac C, Csörgő A, Martínez-Muñoz G (2021) A comparative analysis of gradient boosting algorithms. *Artif Intell Rev* 54:1937–1967
- Blanchflower DG (2021) Is happiness U-shaped everywhere? Age and subjective well-being in 145 countries. *J Popul Econ* 34:575–624
- Blanchflower DG, Graham CL (2022) The mid-life dip in well-being: a critique. *Soc Indic Res* 161:287–344
- Blanchflower DG, Oswald AJ (2008) Is well-being U-shaped over the life cycle? *Soc Sci Med* 66:1733–1749
- Blanchflower DG, Piper A (2022) There is a mid-life low in well-being in Germany. *Econ Lett* 214:110430
- Bussi ere C, Sirven N, Tessier P (2021) Does ageing alter the contribution of health to subjective well-being? *Soc Sci Med* 268:113456
- Carstensen LL, Pasupathi M, Mayr U, Nesselrode JR (2000) Emotional experience in everyday life across the adult life span. *J Pers Soc Psychol* 79:644
- Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM Association for Computing Machinery, New York, NY, USA 785–794 <https://doi.org/10.1145/2939672.2939785>
- Diener E (1984) Subjective well-being. *Psychol Bull* 95:542–575
- Diener E, Chan MY (2011) Happy people live longer: subjective well-being contributes to health and longevity. *Appl Psychol: Health Well-Being* 3:1–43
- Diener E, Oishi S, Tay L (2018) Advances in subjective well-being research. *Nat Hum Behav* 2:253–260
- Diener E, Tay L, Oishi S (2013) Rising income and the subjective well-being of nations. *J Pers Soc Psychol* 104:267–276
- Easterlin RA (2006) Life cycle happiness and its sources: Intersections of psychology, economics, and demography. *J Econ Psychol* 27:463–482
- Fredrickson BL, Grewen KM, Coffey KA, Algoe SB, Firestone AM, Arevalo JMG, Ma J, Cole SW (2013) A functional genomic perspective on human well-being. *Proc Natl Acad Sci USA* 110:13684–13689
- Frijters P, Beaton T (2012) The mystery of the U-shaped relationship between happiness and age. *J Econ Behav Organ* 82:525–542
- García Mata O (2023) At what age do Mexicans suffer the most financial stress? *J Econ Finance Adm Sci*
- Gerstorff D, Drewelies J, Duezel S, Smith J, Wahl H-W, Schilling OK, Kunzmann U, Siebert JS, Katzorreck M, Eibich P (2019) Cohort differences in adult-life trajectories of internal and external control beliefs: a tale of more and better maintained internal control and fewer external constraints. *Psychol Aging* 34:1090
- Gilles L, Manoj K (2016) Bayesian optimization with skopt. [https://scikit-optimize.github.io/0.8/auto\\_examples/bayesian-optimization.html](https://scikit-optimize.github.io/0.8/auto_examples/bayesian-optimization.html)
- Giuntella O, McManus S, Mujcic R, Oswald AJ, Powdthavee N, Tohamy A (2023) The midlife crisis. *Economica* 90:65–110
- Hansen T, Blekesaune M (2022) The age and well-being “paradox”: a longitudinal and multidimensional reconsideration. *Eur J Ageing* 19:1277–1286
- Hardie JH (2014) The consequences of unrealized occupational goals in the transition to adulthood. *Soc Sci Res* 48:196–211
- Helliwell JF, Aknin LB (2018) Expanding the social science of happiness. *Nat Hum Behav* 2:248–252
- Jebb AT, Tay L, Diener E, Oishi S (2018) Happiness, income satiation and turning points around the world. *Nat Hum Behav* 2:33–38
- Kahneman D, Deaton A (2010) High income improves evaluation of life but not emotional well-being. *Proc Natl Acad Sci USA* 107:16489–16493
- Kassie M, Ndiritu SW, Stage J (2014) What determines gender inequality in household food security in Kenya? application of exogenous switching treatment regression. *World Dev* 56:153–171
- Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y (2017) Lightgbm: a highly efficient gradient boosting decision tree. In: *Advances in Neural Information Processing Systems*. (Guyon I et al. eds). Curran Associates, Inc. 30
- Kobau R, Sniezek J, Zack MM, Lucas RE, Burns A (2010) Well-being assessment: an evaluation of well-being scales for public health and population estimates of well-being among US adults. *Appl Psychol: Health Well-Being* 2:272–297
- Krueger PM, Friedman EM (2009) Sleep duration in the United States: a cross-sectional population-based study. *Am J Epidemiol* 169:1052–1063

- Lachman ME (2015) Mind the gap in the middle: a call to study midlife. *Res Hum Dev* 12:327–334
- Li C, Managi S (2023) Income raises human well-being indefinitely, but age consistently slashes it. *Sci Rep* 13:5905
- Löckenhoff CE, Carstensen LL (2004) Socioemotional selectivity theory, aging, and health: the increasingly delicate balance between regulating emotions and making tough choices. *J Personal* 72:1395–1424
- López Ulloa BF, Møller V, Sousa-Poza A (2013) How does subjective well-being evolve with age? A literature review. *J Popul Ageing* 6:227–246
- Lundberg SM, Erion G, Chen H, Degraev A, Prutkin JM, Nair B, Katz R, Himmelstam J, Bansal N, Lee S-I (2020) From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* 2:56–67
- Luu J, Palczewski K (2018) Human aging and disease: lessons from age-related macular degeneration. *Proc Natl Acad Sci USA* 115:2866–2872
- Malik MA, Singh SP, Jyoti J, Pattanaik F (2022) Work stress, health and wellbeing: evidence from the older adults labor market in India *Humanit Soc Sci Commun* 9:204
- Mansoor K, O'Neill D (2021) Minimum wage compliance and household welfare: an analysis of over 1500 minimum wages in India. *World Dev* 147:105653
- Margolis R, Myrskylä M (2011) A global perspective on happiness and fertility. *Popul Dev Rev* 37:29–56
- Molnar C (2020) Interpretable machine learning. Lulu. com
- Molnar C, Casalicchio G, Bischl B (2020) Interpretable machine learning—a brief history, state-of-the-art and challenges. Springer International Publishing, pp 417–431
- Murayama Y, Murayama H, Hasebe M, Yamaguchi J, Fujiwara Y (2019) The impact of intergenerational programs on social capital in Japan: a randomized population-based cross-sectional study. *BMC Public Health* 19(1):156
- Nomaguchi K, Milkie MA (2020) Parenthood and well-being: a decade in review. *J Marriage Fam* 82:198–223
- Oswald AJ, Wu S (2010) Objective confirmation of subjective measures of human well-being: evidence from the USA. *Science* 327:576–579
- Pearl J (2003) Statistics and causal inference: a review. *Test* 12:281–345
- Pearl J (2009) Causal inference in statistics: an overview. *Stat Surv* 3:96–146. 151
- Plagnol AC (2011) Financial satisfaction over the life course: The influence of assets and liabilities. *J Econ Psychol* 32:45–64
- Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulina A (2018) CatBoost: unbiased boosting with categorical features. *Adv Neural Inf Process Syst*
- Rubin DB (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 66:688
- Shu X, Chen J, Zhu Y (2023) Changing times and subjective well-being in urban China 2003–2013: an age-period-cohort approach. *Chin J Sociol* 9:321–354
- Stephens A, Deaton A, Stone AA (2015) Subjective wellbeing, health, and ageing. *Lancet* 385:640–648
- Stone AA, Schwartz JE, Broderick JE, Deaton A (2010) A snapshot of the age distribution of psychological well-being in the United States. *Proc Natl Acad Sci* 107:9985–9990
- Sutin AR, Terracciano A, Milaneschi Y, An Y, Ferrucci L, Zonderman AB (2013) The Effect of Birth Cohort on Well-Being. *Psychol Sci* 24:379–385
- Turner R, Eriksson D, McCourt M, Kiliç, Laaksonen, E, Xu Z, Guyon I (2021) Bayesian optimization is superior to random search for machine learning hyperparameter tuning: analysis of the black-box optimization challenge 2020. *NeurIPS 2020 competition and demonstration track*. PMLR, pp. 3–26
- Van Landeghem B (2012) A test for the convexity of human well-being over the life cycle: longitudinal evidence from a 20-year panel. *J Econ Behav Organ* 81:571–582
- Wanberg CR, Kanfer R, Hamann DJ, Zhang Z (2016) Age and reemployment success after job loss: an integrative model and meta-analysis. *Psychol Bull* 142:400
- Wu W, Wu H, Wu X, Gu J, Qi X (2024) A meta-analysis of life satisfaction's association with cognitive reappraisal and expressive suppression: the influences of age, gender, and cultural values. *J Happiness Stud* 25:41

## Acknowledgements

The authors express their gratitude to the seminar participants from the Research Institute of Economy, Trade and Industry (RIETI) for their comments on this study. This research was supported by the following funding agencies: JSPS KAKENHI (Grant Nos. JP20H00648 and 25K03322), and JST MiraiProgram (Grant No. JPMJMI2214).

## Author contributions

CL: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, original draft, and visualization. JM: Conceptualization, methodology, validation, formal analysis, investigation, review and editing. JZ: Data processing. BS: Data processing. AK: Conceptualization, validation, review and editing. SM: Resources, review and editing, supervision, project administration, and funding acquisition. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Ethical approval

This study used secondary data from the Gallup World Poll, conducted annually by Gallup, Inc. from 2005 to 2022 across all sampled countries. Gallup obtains ethical approval for its global survey procedures through both internal and external review processes that adhere to internationally recognized research ethics standards, including the Declaration of Helsinki and the ethical guidelines of the Council for International Organizations of Medical Sciences (CIOMS). All data collection is conducted in accordance with applicable national regulations for research involving human participants and the ethical principles set forth in the Belmont Report, disseminated by the U.S. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Gallup World Poll is conducted under the oversight of Gallup's Institutional Review Board (Gallup IRB). The IRB's framework is based on the principles of respect for persons, beneficence, and justice. Gallup upholds strict protocols to protect respondent confidentiality and anonymity throughout data collection, transfer, storage, and reporting. Gallup's procedures also comply with the professional ethical codes of the Worldwide Association for Public Opinion Research (WAPOR), the European Society for Opinion and Marketing Research (ESOMAR), the American Association for Public Opinion Research (AAPOR), the Council of American Survey Research Organizations (CASRO), and the American Statistical Association (ASA). Gallup IRB approval number: 33f44049-4067. Date of approval: June 15, 2022. Scope of approval: Global survey data collection across all countries sampled in the Gallup World Poll, 2005–2022. This study involved only secondary analysis of fully anonymized data provided directly by Gallup. The authors had no access to identifiable personal information and no interaction with survey respondents. In accordance with institutional policy and prevailing ethical standards for secondary data analysis of de-identified datasets, the ethics review committee of Kyushu University, Japan confirms that no additional ethical approval was required.

## Informed consent

Oral informed consent was obtained by Gallup, Inc. from all participants prior to data collection, during the years 2005 to 2022, in accordance with Gallup's standardized global survey protocol. Consent was administered by trained local interviewers, using a uniform script, and conducted in the participant's native language to accommodate diverse literacy levels and cultural contexts. Participants were clearly informed about the purpose of the survey, their right to decline or withdraw at any time, and the fact that their responses would remain anonymous and confidential. The oral consent process was documented at the time of the interview in line with Gallup's internal procedures and ethical standards. Gallup's informed consent procedures align with the ethical principles outlined in the Declaration of Helsinki, the CIOMS guidelines, and the Belmont Report. Participants were assured that their data would be anonymized and used solely for research and policy analysis purposes. The scope of consent included voluntary participation, data use, and dissemination of findings in a de-identified and aggregated form. No personally identifiable information was shared with the authors, and no direct interaction occurred between the researchers and participants.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1057/s41599-025-05708-9>.

**Correspondence** and requests for materials should be addressed to Shunsuke Managi.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025