



## OPEN ACCESS

## EDITED BY

Piotr Romaniuk,  
Medical University of Silesia, Poland

## REVIEWED BY

Pragyan Monalisa Sahoo,  
Utkal University, India  
Milos Stepovic,  
University of Kragujevac, Serbia  
Babita Panda,  
ITM University Raipur, India

## \*CORRESPONDENCE

Narimasa Kumagai  
✉ kumanari@seinan-gu.ac.jp

RECEIVED 12 February 2024

ACCEPTED 28 June 2024

PUBLISHED 17 July 2024

## CITATION

Kumagai N and Jakovljević M (2024) Random forest model used to predict the medical out-of-pocket costs of hypertensive patients. *Front. Public Health* 12:1382354. doi: 10.3389/fpubh.2024.1382354

## COPYRIGHT

© 2024 Kumagai and Jakovljević. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Random forest model used to predict the medical out-of-pocket costs of hypertensive patients

Narimasa Kumagai<sup>1\*</sup> and Mihajlo Jakovljević<sup>2,3,4</sup>

<sup>1</sup>Faculty of Economics, Seinan Gakuin University, Fukuoka, Japan, <sup>2</sup>UNESCO-TWAS, Section of Social and Economic Sciences, Trieste, Italy, <sup>3</sup>Shaanxi University of Technology, Hanzhong, China, <sup>4</sup>Department of Global Health Economics and Policy, University of Kragujevac, Kragujevac, Serbia

**Background:** Precise prediction of out-of-pocket (OOP) costs to improve health policy design is important for governments of countries with national health insurance. Controlling the medical expenses for hypertension, one of the leading causes of stroke and ischemic heart disease, is an important issue for the Japanese government. This study aims to explore the importance of OOP costs for outpatients with hypertension.

**Methods:** To obtain a precise prediction of the highest quartile group of OOP costs of hypertensive outpatients, we used nationwide longitudinal data, and estimated a random forest (RF) model focusing on complications with other lifestyle-related diseases and the nonlinearities of the data.

**Results:** The results of the RF models showed that the prediction accuracy of OOP costs for hypertensive patients without activities of daily living (ADL) difficulties was slightly better than that for all hypertensive patients who continued physician visits during the past two consecutive years. Important variables of the highest quartile of OOP costs were age, diabetes or lipidemia, lack of habitual exercise, and moderate or vigorous regular exercise.

**Conclusion:** As preventing complications of diabetes or lipidemia is important for reducing OOP costs in outpatients with hypertension, regular exercise of moderate or vigorous intensity is recommended for hypertensive patients that do not have ADL difficulty. For hypertensive patients with ADL difficulty, habitual exercise is not recommended.

## KEYWORDS

activities of daily living (ADL), exercise, hypertension, Japan, out-of-pocket (OOP) costs, prediction, random forest

## 1 Introduction

Given the poverty caused by rising health care costs for chronically ill patients, accurate prediction of out-of-pocket (OOP) costs is important to prevent catastrophic health care expenditures for those patients because OOP costs and prepaid private health financing are also expected to grow, although less than growth in government spending (1). On the one hand, research by Hwang et al. (2) showed a positive, nearly linear relationship between OOP costs and the number of chronic diseases. On the other hand, Zhang et al. (3) used a

data-driven ensemble learning procedure and found that the top-ranking factors that best predicted OOP costs were insurance type, age, asthma, family size, race, and number of physician visits.

Although three or more activities of daily living (ADL) dependencies are associated with the need for long-term care, the onset of ADL difficulties in chronically ill persons may differ from that of healthy older adults (4). Therefore, both ADL difficulties and major chronic diseases should be considered when analyzing physician visits and predicting OOP costs.

Hypertension is the leading cause of stroke and ischemic heart disease, leading to premature death in Japan and worldwide (5–7). Individuals with well-controlled hypertension who are undergoing treatment have a lower risk of developing cardiovascular or cerebrovascular diseases (8, 9). In Japan, where most hypertensive patients tend to adhere to their family physicians' treatment plans, regular physician visits every 30 days are effective in stabilizing blood pressure in hypertensive patients (10).

The prevalence of hypertension control among hypertensive patients taking antihypertensive medications increased from 1980 to 2016 in Japan (11). Furthermore, 72% (31 million) of the hypertensive patients were poorly controlled in 2017. Poorly controlled hypertension during treatment is expensive. A study by Japanese researchers showed that grade 3 untreated hypertension is likely to be extremely costly (12). Hence, controlling blood pressure in patients with hypertension through physician visits and medical expenses is an important issue for the Japanese government. The estimates of the average treatment effects on treated patients suggest that three consecutive years of physician visits negatively impacted the poor subjective health of hypertensive patients (10). Therefore, this study focuses on the patients who had physician visits during the past two consecutive years.

This study estimates the predicted OOP costs of outpatients in Japan with hypertension who are middle-aged and older people. We used data from a longitudinal survey conducted by the Ministry of Health (2005–2020) and estimated a random forest (RF) model focusing on complications with other lifestyle-related diseases and nonlinearities. The Stata command “rforest” in Stata version 18 (StataCorp) was used to estimate RF models.

Can the governments of countries with national health insurance precisely predict OOP costs and improve health policy design? The challenge of fiscal sustainability of healthcare financing due to increased OOP expenditures is particularly alarming in the Global South's low- and middle-income countries (LMICs) worldwide (13). This is because of huge socioeconomic inequalities

as shown by GINI indices and consecutive poor affordability of medical care (14), particularly in the vast rural and remote peripheries of these nations (15). This imbalance in health expenditure dynamics was created, to a large extent, through the decade-long evolution of the morbidity landscape, much earlier in industrialized wealthy societies of the Global North (16). Namely, traditional, mostly curable infectious diseases of the short clinical course (17) were gradually replaced by incurable, life-time chronic noncommunicable diseases (NCDs) such as hypertension, diabetes and all consequences of atherosclerosis (18). Such disorders remain expensive and difficult to treat, imposing a far higher burden to the health system. Risk sharing occurring in early modern health systems of the late XIX and early XX century have resolved this burden only to a limited extent (19). Technological innovation (20) and significant life expectancy extension worldwide associated with population aging have only worsened the situation (21). Therefore, health econometric exploration of the underlying causes in large, aged societies of Asia (22) remain capable of revealing hidden, underlying patterns of spending (23). They may also provide glimpses of possible policy strategies targeted at both prevention and more effective technical and allocative efficacy of provision and delivery of medical care (24). We believe that our study makes a significant practical contribution to the prediction of OOP costs and, therefore, the improvement of national health insurance and health policy design.

The remainder of this paper is organized as follows: Section 2 provides an overview of the machine learning method, RF, and its split selection based on an ensemble learning algorithm. Section 3 presents the estimation results of the RF model. Section 4 discusses the variable importance of OOP costs for outpatients with hypertension. Section 5 concludes the paper.

## 2 Methods

### 2.1 Random forest model

RF has the advantage of treating missing values because splits at any node can occur even if some independent variables are missing. Very few restrictions are imposed on the choice of explanatory variables because no functional form is assumed (25). Using an RF model, we can predict the dichotomous variable and obtain the predictions with no significant bias (26). The RF model does not estimate the coefficients of the explanatory variables in the same way as econometric models and easily adapts to the nonlinearities found in the data; therefore, it tends to predict better than linear regression (27). Researchers can use several explanatory variables to eliminate the effects of overtraining. The RF estimate was computed with a minimum leaf size of 1 and 800, averaged over the results of 100 trees with no maximum depth (these are the defaults for rforest). The full model has 45 features.

We split a nationally representative sample of the older adults in Japan which is explained below into two subsets: 50% of the data are used for training, and 50% of the data are used for validation. When an RF model randomly selects explanatory variables from among all the explanatory variables, the sample of the regression tree is split. RF uses entropy for split selection in classification cases. At each internal node of the decision tree, entropy (E) is given by Equation (1).

Abbreviations: hi25\_ln\_sum\_treat\_r\_1, lagged variable of logged real out-of-pocket (OOP) costs (2020 = 100); diabetes\_b, dummy variable for having diabetes; heart\_b, dummy variable for having heart diseases; lipid\_b, dummy variable for having lipidemia; stroke\_b, dummy variable for having stroke; cancer\_b, dummy variable for having cancer; m\_menh24, moderate mental health (4 < K6 < 13); s\_menh24, serious mental health (12 < K6); drinkh\_b, drinking habit; pinactiv\_b, no habitual exercise; prpa23\_b, moderate or vigorous regular exercise; age2, age squared; sexb, gender (male = 1); eduatt\_1, low education (junior high school); eduatt\_56, high education (university or graduate school); inmate\_b, dummy variable for living together with family members excluding spouse; inc\_earned, dummy variable for earned income during the past month; excretion\_b2, needing assistance for using the toilet; walk\_b, having difficulties in walking; walk\_b2, needing assistance for walking.

$$E = - \sum_{i=1}^c p_i \times \log(p_i) \quad (1)$$

where  $c$  is the number of unique classes, and  $p_i$  is the prior probability for each class (27).

Each tree is constructed using different bootstrap samples. Each bootstrap sample randomly leaves out approximately one-third of the observations. These are referred to as the out-of-bag (OOB) samples. For classification problems, the OOB error used for validation represents the classification error. Thus, we can use an OOB error tested against training data subsets that are not included in subtree construction. The RF model analysis aims to explore the extent to which explanatory variables affect the dependent variable, and the effects are assessed using variable importance. Ranking the importance of the factors is challenging in data analysis. In supervised learning there are problems with multiple factors. Ranking the importance of factors involves assessing the impact of each factor on the response variable and creating quantitative measures for comparison (3).

Generally, variable importance, which measures the magnitude of the forecast error when explanatory variables are randomly selected, is used. The larger the prediction error, the greater the importance of the variables that can be evaluated. Variable importance is calculated by adding the improvement in the objective function given in the splitting criterion over all internal nodes of a tree and across all trees in the forest. The variable importance score is normalized by dividing all scores by the maximum score (27).

## 2.2 Data

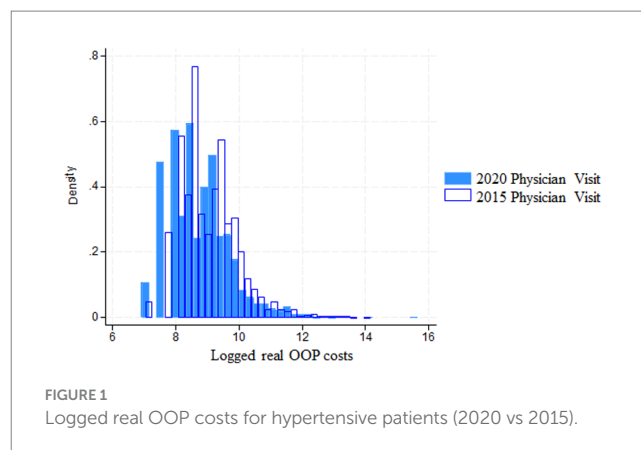
We used longitudinal data over 16 consecutive years (2005–2020), obtained from the Longitudinal Survey of Middle and Older Persons (LSMOP) of the Japanese Ministry of Health, Labor, and Welfare (MHLW). Data were collected using a combination of interviews and self-administered questionnaires. The LSMOP collects information on family situation, health status, and employment status. In the second-wave of the survey, the LSMOP asked about the educational attainment of the respondents and their spouses.

The samples for the LSMOP were randomly selected using a two-stage sampling procedure. First, 2,515 districts in 2005 were randomly selected from the 5,280 districts covered by the “Comprehensive Survey of Living Conditions” conducted by the MHLW in 2004 (another nationwide survey). Second, 40,877 respondents aged 50–59 years, as of October 30, 2005, were randomly selected from each district.

The proportion of the total sum of 16 consecutive respondents to the subjects who responded to the second wave of the survey in 2006 was approximately 53% (10). A total of 34,240 individuals responded to the first wave of the survey (response rate: 83.8%), whereas 32,285 participants returned the questionnaires for the second wave (response rate: 92.2%). Almost half of the respondents in the 2006 survey dropped out of the study.

The sample in this study consisted of hypertensive patients who had continued physician visits during the past two consecutive years. Male patients and those with diabetes, lower educational attainment, no habitual exercise, or smoking habits tended not to have physician visits during the past two consecutive years (10).

Based on a doctor’s diagnosis, hypertension was defined as a systolic blood pressure (BP)  $\geq 140$  mmHg, a diastolic BP  $\geq 90$  mmHg,



or use of antihypertensive medication. The hypertension control rate was defined as the proportion of patients with systolic BP  $< 140$  mmHg and diastolic BP  $< 90$  mmHg among the hypertensive patients taking antihypertensive medication (11). The proportion of patients with cancer, diabetes, heart disease, lipidemia, and strokes among the older adults with hypertension is higher than among the older adults without hypertension (10).

## 3 Results

Figure 1 shows the logged real OOP costs in 2020 for hypertensive patients who had physician visits during 2019–20 (blue). It further shows hypertensive patients who had physician visits during 2014–15 (white). Two changes can be seen: first, the peak of the blue histogram is located more toward the left than the white histogram; and second, the tail of the right hem of the blue graph is getting longer as the proportion of patients aged 70–74 with a 20% co-payment rate is higher in 2020 than in 2015. This is associated with a shift to the left side of the histogram. Hypertensive patients in their 60s, whose co-payment rate was 30% more, contributed the right hem having a longer tail than the 70–74 age group in 2020. We pay attention to these points when interpreting the statistical data in Table 1.

Table 1 shows the sample characteristics of the hypertensive patients who continued their physician visits during the past two consecutive years. The proportion of aged 70–74 in the highest quartile group of OOP costs was 0.089, which was significantly different from those of comparison group at the 1% level. The mean OOP cost in the highest quartile was approximately 38,900 Japanese yen per month (data not shown). By contrast, the proportion of patients with chronic diseases in the highest quartile group of OOP costs, such as diabetes, heart disease, stroke, and cancer, was much higher than that of the overall sample. Therefore, we conjectured that the number of chronic diseases contributes to higher OOP costs.

Approximately 18% of hypertensive patients who continued physician visits during the past two consecutive years did not respond to questions regarding their OOP costs at the time of the survey. Compared with the overall sample, a higher percentage of non-respondents were never married, less educated, had a serious mental health status, had no habitual exercise, and smoked (Supplementary Table S1). The number of variables to randomly investigate was the square root of the number of independent variables, and the number of iterations was 500.

TABLE 1 Sample characteristics of hypertensive patients who continued physician visits during the past two consecutive years.

Variables	Highest quartile of OOP costs			All		
	N	Mean or proportion	SD	N	Mean or proportion	SD
Dependent variable						
Logged real out-of-pocket (OOP) costs (2020 = 100)				57,710	9.093	0.999
Demographic variables						
Age	14,247	<b>63.73</b>	4.16	57,710	64.48	4.48
Dummy variable for aged 70–74	14,247	<b>0.089</b>	0.284	57,710	0.149	0.356
Gender (male = 1)	14,247	<b>0.542</b>	0.498	57,710	0.523	0.499
Married (reference)	14,247	<b>0.917</b>	0.275	57,710	0.945	0.227
Never married	14,247	<b>0.082</b>	0.275	57,710	0.055	0.227
Divorced or widowed	14,247	<b>0.000</b>	0.017	57,710	0.000	0.009
Dummy variable for living together with family members excluding spouse	14,246	<b>0.555</b>	0.497	57,693	0.539	0.498
Dummy variable for earned income during the past month	11,968	<b>0.578</b>	0.494	50,241	0.584	0.493
Educational attainment						
Junior high school	14,247	<b>0.172</b>	0.378	57,710	0.162	0.368
High school (reference)	14,247	<b>0.497</b>	0.500	57,710	0.514	0.500
Vocational school or junior college	14,247	0.145	0.352	57,710	0.144	0.351
University or graduate school	14,247	0.179	0.383	57,710	0.175	0.380
Sum of K6						
Serious mental health (12 < K6)	10,766	<b>0.085</b>	0.279	45,867	0.054	0.225
Moderate mental health (4 < K6 < 13)	10,766	<b>0.313</b>	0.464	45,867	0.244	0.429
Objective health status						
Dummy variable for having diabetes	14,059	<b>0.306</b>	0.461	56,997	0.196	0.397
Dummy variable for having heart diseases	14,004	<b>0.152</b>	0.359	56,790	0.092	0.289
Dummy variable for having lipidemia	13,946	<b>0.326</b>	0.469	56,712	0.292	0.454
Dummy variable for having stroke	13,884	<b>0.075</b>	0.263	56,274	0.043	0.202
Dummy variable for having cancer	13,903	<b>0.090</b>	0.286	56,533	0.042	0.200
Perceived health						
Dummy variable for felt worse symptoms of high blood pressure than its onset during the past year	14,247	<b>0.017</b>	0.129	57,710	0.009	0.093
Lifestyle						
Drinking habit	14,247	<b>0.249</b>	0.432	57,710	0.237	0.425
No habitual exercise	14,221	<b>0.346</b>	0.476	57,604	0.325	0.468
Moderate or vigorous regular exercise	14,247	<b>0.280</b>	0.449	57,710	0.307	0.461
Smoking habit	14,247	0.156	0.363	57,710	0.154	0.361
Having difficulties in activities of daily living (ADL)						
Dummy variable for walking	14,247	0.230	0.421	57,710	0.151	0.358
Dummy variable for getting in and out of bed	14,247	0.234	0.423	57,710	0.153	0.360
Dummy variable for standing up from a chair	14,247	0.234	0.424	57,710	0.153	0.360
Dummy variable for dressing	14,247	0.232	0.422	57,710	0.152	0.359
Dummy variable for washing face	14,247	0.234	0.424	57,710	0.153	0.360
Dummy variable for eating meals	14,247	0.235	0.424	57,710	0.153	0.360
Dummy variable for using the toilet	14,247	0.234	0.423	57,710	0.153	0.360
Dummy variable for bathing and showering	14,247	0.226	0.418	57,710	0.149	0.356
Dummy variable for stair climbing	14,247	0.225	0.417	57,710	0.148	0.356

(Continued)

TABLE 1 (Continued)

Variables	Highest quartile of OOP costs			All		
	N	Mean or proportion	SD	N	Mean or proportion	SD
Dummy variable for carrying of shopped items	14,247	0.213	0.410	57,710	0.144	0.351
Needing assistance for ADL						
Dummy variable for walking	14,247	0.009	0.096	57,710	0.004	0.064
Dummy variable for getting in and out of bed	14,247	0.005	0.073	57,710	0.002	0.046
Dummy variable for standing up from a chair	14,247	0.005	0.070	57,710	0.002	0.046
Dummy variable for dressing	14,247	0.008	0.087	57,710	0.003	0.056
Dummy variable for washing face	14,247	0.005	0.069	57,710	0.002	0.048
Dummy variable for eating meals	14,247	0.004	0.064	57,710	0.002	0.042
Dummy variable for using the toilet	14,247	0.005	0.073	57,710	0.002	0.047
Dummy variable for bathing and showering	14,247	0.013	0.113	57,710	0.006	0.074
Dummy variable for stair climbing	14,247	0.015	0.120	57,710	0.006	0.080
Dummy variable for carrying of shopped items	14,247	0.026	0.159	57,710	0.011	0.105

Non-respondents of OOP costs are excluded. Boldface indicates that the mean or the proportion of variables in the highest quartile group of copayments is statistically different from those of comparison group at the 1% level. Source: Longitudinal Survey of Middle-aged and Older Persons 2005–2020.

To predict logged real OOP costs of hypertensive outpatients, the following 45 candidate covariates were included: age, gender, dummy variable for aged 70–74 as the proxy variable of lower copayment rate (20%), marital status, earned income during the past month, educational attainment, living together with family members excluding spouse, non-communicable diseases (coded as 5 dummy variables) such as having diabetes, mental health variables, perceived health, having difficulties in ADL or needing assistance for ADL (coded as 20 dummy variables), lifestyle variables such as no habitual exercise (coded as 4 dummy variables), logged real OOP costs during the past year.

Table 2 compares the prediction accuracies of the three RF models: full model (including all covariates), sub-model (1) (no difficulties with ADL), and sub-model (2) (difficulties with ADL or needing assistance). The lowest OOB error of the three models was 0.2229 for Model (1). This implies that the prediction accuracy of OOP costs for hypertensive patients without ADL difficulties was slightly better than that for all hypertensive patients who had continued their physician visits during the past two consecutive years.

In the full model, the top 10 variables of importance were the highest quartile of OOP costs in the previous year, age, age squared, presence of diabetes, living together with family members excluding the spouse, having lipidemia, no habitual exercise, moderate or vigorous regular exercise, gender, moderate mental health status (Figure 2). Interestingly, most of these variables are consistent with the significant variables in the doubly robust estimation of Kumagai et al. (10), who analyzed the determinants of subjective poor health among hypertensive patients. This means that most determinants of poor subjective health and high medical costs are the same.

Hypertensive patients with diabetes or hyperlipidemia are often unaware of their symptoms and often allow their condition to progress/worsen untreated. As the above variables indicate, moderate or vigorous regular exercise is important to prevent the progression of arteriosclerosis, which is associated with poor subjective health. No habitual exercise is thought to contribute to the development of ischemic heart diseases such as angina pectoris and myocardial infarction.

The sample was divided based on the presence or absence of difficulties in ADL. Sub-model (1) is the group with no ADL difficulty. The variables of importance for sub-models (1) and (2) are almost the same as those in the full model (Figure 3). However, in sub-model (1), the degree of importance of moderate or vigorous regular exercise is greater than that in the full model. This coincides with the results (28) showing that individuals with moderate or vigorous regular exercise tend to have a higher health stock. In contrast, the degree of importance of moderate mental health in sub-model (2) was greater than that in the full model. The lack of habitual exercise might have been associated with moderate mental health in the ADL difficulty group.

## 4 Discussion

Increasing the prevalence of complex diseases leads to more OOP payments (29). Furthermore, increases in OOP payments can have regressive effects, potentially leading to a substantial number of people being pushed into poverty due to health care costs (30, 31). Thus, accurate predictions of health spending patterns can help predict how countries will fare in the health care sector (31).

The determinants of OOP costs for lifestyle-related outpatients are not uniform, as some outpatients with hypertension also have other lifestyle-related diseases, such as diabetes (32). Indeed, according to the 2021 Survey on Trends in Lifestyle-Related Diseases (33), outpatients with diabetes or hypertension accounted for the first or second largest share of healthcare expenditures among 10 lifestyle-related diseases (34). The survey (33) shows that hypertension and hyperlipidemia had the first and second highest rates of physician visits, respectively, and the year-to-year increase in outpatient costs for diabetes and hyperlipidemia was above 5%.

Most contemporary health systems have achieved a certain degree of universal health coverage, particularly with regards to prevention, screening, diagnostics, and treatment of major NCDs such as hypertension (35). Given the long and clinically unpredictable course

TABLE 2 Comparison of prediction accuracy.

	All	No difficulties in ADL	Having difficulties in ADL or needing assistance
N	57,710	46,905	10,805
Features	45	25	35
OOB error	0.2355	0.2229	0.2956

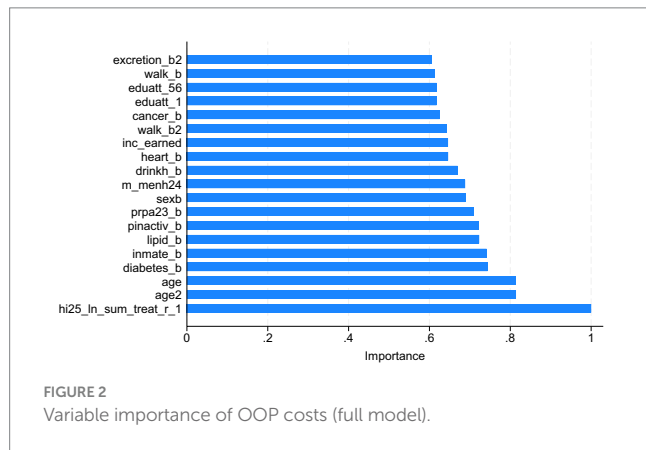


FIGURE 2 Variable importance of OOP costs (full model).

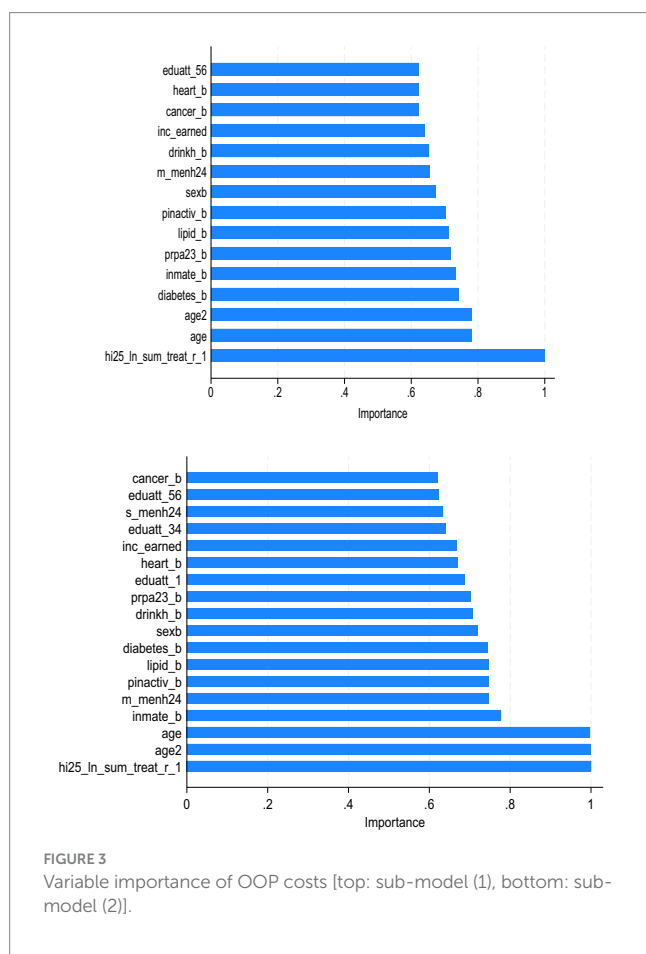


FIGURE 3 Variable importance of OOP costs [top: sub-model (1), bottom: sub-model (2)].

sudden and predictable complications of hypertensive disease such as cerebral stroke, myocardial infarction or renal insufficiency, may result in life threatening conditions (37) with necessary hospital admissions to the intensive care unit (38). These treatments may last up to a few weeks and frequently lead to long-term impairment, disability, months-long absenteeism, decreased working ability and even premature mortality (39). Such consequences are the real-life toll and societal burdens of hypertension leading to huge OOP spending by patients and their families (40). Furthermore, this leads to the catastrophic household expenditure phenomenon which is widespread in the Global South LMICs countries without strong Japanese welfare and risk sharing agreements (41), which leads to families falling into debt and poverty traps (42). Therefore, we should carefully observe our Japanese OOP cost projections, particularly compared to Emerging BRICs Markets (43) as the leading global drivers of real GDP growth and consecutive health spending worldwide (44).

We estimated the predicted logged real OOP costs of hypertensive outpatients, with attention to complications of other lifestyle-related diseases and nonlinearities in the data, using RF models. The variables of importance shown in Figures 2, 3 indicate that preventing complications of diabetes or lipidemia is important for reducing OOP costs in outpatients with hypertension. Therefore, regular exercise of moderate or vigorous intensity is recommended in the no ADL difficulty group. In contrast, for hypertensive patients with ADL difficulties, habitual exercise must not be remedied to prevent the development of ischemic heart disease (45).

Providing accurate prediction of OOP payments for chronically ill patients and reducing the upper limit of OOP payments is important to prevent their catastrophic health care expenditures. However, due to data limitations, this study cannot link data on public pension benefits and OOP payments for older people to examine the desired upper limit of OOP payments.

Furthermore, it was well documented in the seminal literature that the importance of the education of patients about the risk factors may substantially influence their long-term behavior (46). Responsible attitude toward patient’s own disease may lead to pre-cautious diet, exercise, healthy lifestyle and utilization of artificial intelligence assisted small, wearable medical devices (47). Such an equipment may provide the attending physician and the patient alike with reliable 24/7 supervision of the most, sensitive fluctuating clinical indicators such as arterial tension, cardiac rhythm, glycemia control or partial pressures of blood oxygen (48).

All these medical surveillance data are possible to be obtained to the scale of in-depth observation of inner bodily dynamics that was unthinkable only a few years ago. Constant education of health providers about the ongoing AI-related, technological revolution in clinical medicine is rapidly expanding the horizon of our understanding of arterial hypertension’s unpredictable dynamics in the real-world setting (49). Alongside this change, adoption of random forest and similar modeling approaches to the health econometrics field allow us to better cope with “a boomerang phenomenon.” Namely, rather insufficient density of physicians and nursing staff, have significant impact to the delayed discovery of hypertension persistence in the vast number of patients due to poor coverage of general population with screening measures (50).

In return this fact leads to a neglected, self-evolving disorder leading to microangiopathic changes of small blood vessels and accelerated atherosclerosis of large blood vessels (51). Both histopathological changes drive occurrence of unpredictable, severe

of the disease, most systems cover provision and reimbursement of essential medicines for hypertension and regular physician visits in primary care through the medical insurance premiums (36). However,

and hard to treat clinical complications ranging from malignant cardiac arrhythmias, myocardial infarction, cerebral stroke etc. Most of these conditions require lengthy and exceptionally expensive intensive care unit admissions which might have been partially prevented with mass public health screening procedures of far more humble than overall budget impact (52). This is neglect of NCDs is widely known as the boomerang effect. Among other OECD countries the causal bottleneck insufficiency is prominent in Japan with 269 practicing physicians per 100,000 inhabitants as of 2020 (53). Room for successful intervention yet remains as witnessed by the findings of current study.

## 5 Conclusion

This study predicted the highest quartile group of OOP costs for hypertensive outpatients, focusing on complications with other lifestyle-related diseases and nonlinearities of the data. The estimation results of the RF models showed that the prediction accuracy of OOP costs for hypertensive patients without ADL difficulties was slightly better than that for all hypertensive patients who had continued their physician visits during the past two consecutive years. The important variables in the highest quartile of OOP costs were age, diabetes or lipidemia, lack of habitual exercise, and moderate or vigorous regular exercise. Habitual exercise must not be remedied to prevent the development of ischemic heart disease in patients with hypertension and ADL difficulties. In the present study, sub-model (2) recorded the highest OOB error of the three models; the mechanism by which the exclusion of older adults with no ADL difficulties contributed to the lower prediction accuracy is unclear. Future studies should be conducted to clarify this point.

## Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding author.

## References

- Dieleman JL, Campbell M, Chapin A, Eldrenkamp E, Fan VY, Haakenstad A, et al. Future and potential spending on health 2015–40: development assistance for health, and government, prepaid private, and out-of-pocket health spending in 184 countries. *Lancet*. (2017) 389:2005–30. doi: 10.1016/S0140-6736(17)30873-5
- Hwang W, Weller W, Ireys H, Anderson G. Out-of-pocket medical spending for care of chronic conditions. *Health Aff*. (2001) 20:267–78. doi: 10.1377/hlthaff.20.6.267
- Zhang C, Ding Y, Peng Q. Who determines United States healthcare out-of-pocket costs? Factor ranking and selection using ensemble learning. *Health Inf Sci Syst*. (2021) 9:22. doi: 10.1007/s13755-021-00153-9
- Fong JH. Disability incidence and functional decline among older adults with major chronic diseases. *BMC Geriatr*. (2019) 19:323. doi: 10.1186/s12877-019-1348-z
- Ikeda N, Inoue M, Iso H, Ikeda S, Satoh T, Noda M, et al. Adult mortality attributable to preventable risk factors for non-communicable diseases and injuries in Japan: a comparative risk assessment. *PLoS Med*. (2012) 9:e1001160. doi: 10.1371/journal.pmed.1001160
- Lawes CM, Vander Hoorn S, Rodgers A. International Society of Hypertension. Global burden of blood-pressure-related disease, 2001. *Lancet*. (2008) 371:1513–8. doi: 10.1016/S0140-6736(08)60655-8
- Fujiyoshi A, Ohkubo T, Miura K, Murakami Y, Nagasawa SY, Okamura T, et al. Blood pressure categories and long-term risk of cardiovascular disease according to age group in Japanese men and women. *Hypertens Res*. (2012) 35:947–53. doi: 10.1038/hr.2012.87

## Author contributions

NK: Writing – original draft, Formal analysis, Conceptualization.  
MJ: Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the JSPS KAKENHI (Grant No. JP20K01739). The funder had no role in the design of the study; the collection, analysis, and interpretation of the data; and the writing of the manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2024.1382354/full#supplementary-material>

- SHEP Cooperative Research Group. Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension. Final results of the systolic hypertension in the elderly program (SHEP). *JAMA J Am Med Assoc*. (1991) 265:3255–64.
- Staessen JA, Fagard R, Thijs L, Celis H, Arabidze GG, Birkenhager WH, et al. Randomised double-blind comparison of placebo and active treatment for older patients with isolated systolic hypertension. The systolic hypertension in Europe (Syst-Eur) trial investigators. *Lancet*. (1997) 350:757–64. doi: 10.1016/S0140-6736(97)05381-6
- Kumagai N, Nishimura S, Jakovljević M. Could high continuity of care (COC) have a negative impact on subjective health of hypertensive patients? A Japanese perspective. *Cost Eff Resour Alloc*. (2023) 21:39. doi: 10.1186/s12962-023-00448-6
- Hisamatsu T, Segawa H, Kadota A, Ohkubo T, Arima H, Miura K. Epidemiology of hypertension in Japan: beyond the new 2019 Japanese guidelines. *Hypertens Res*. (2020) 43:1344–51. doi: 10.1038/s41440-020-0508-z
- Nakamura K, Miura K, Nakagawa H, Okamura T, Okuda N, Nishimura K, et al. Treated and untreated hypertension, hospitalization, and medical expenditure: an epidemiological study in 314622 beneficiaries of the medical insurance system in Japan. *J Hypertens*. (2013) 31:1032–42. doi: 10.1097/HJH.0b013e32835f5747
- Jakovljević M, Liu Y, Cerda A, Simonyan M, Correia T, Mariita RM, et al. The global south political economy of health financing and spending landscape—history and presence. *J Med Econ*. (2021) 24:25–33. doi: 10.1080/1366998.2021.2007691

14. Tekle MG, Gebre EG, Medhin G, Teklu AM, Tadesse D, Gebrehiwot G, et al. Assessment of clinician's knowledge and attitude toward health extension program in Ethiopia: an exploratory factor analysis. *Global Health Econ Sustain.* (2023) 1:0887. doi: 10.36922/ghes.0887
15. Jakovljević M, Groot W, Souliotis K. Health care financing and affordability in the emerging global markets. *Front Public Health.* (2016) 4:1054409. doi: 10.3389/fpubh.2016.00002
16. Tyrovolas S, Kassebaum NJ, Stergachis A, Abraha HN, Alla F, Androudi S, et al. The burden of disease in Greece, health loss, risk factors, and health financing, 2000–16: an analysis of the global burden of disease study 2016. *Lancet Public Health.* (2018) 3:e395–406. doi: 10.1016/S2468-2667(18)30130-0
17. Local Burden of Disease Vaccine Coverage Collaborators. Mapping routine measles vaccination in low- and middle-income countries. *Nature.* (2021) 589:415–9. doi: 10.1038/s41586-020-03043-4
18. GBD 2016 Occupational Risk Factors Collaborators. Global and regional burden of disease and injury in 2016 arising from occupational exposures: a systematic analysis for the global burden of disease study 2016. *Occup Environ Med.* (2020) 77:133–41. doi: 10.1136/oemed-2019-106008
19. Jakovljević M, Vuković M, Chen CC, Antunović M, Dragojević Simić V, Velicković Radovanović R, et al. Do health reforms impact cost consciousness of health care professionals? Results from a nation-wide survey in the Balkans. *Balkan Med J.* (2016) 33:8–17. doi: 10.5152/balkanmedj.2015.15869
20. Kosaka M, Ozaki A, Kaneda Y, Saito H, Yamashita E, Murayama A, et al. Generic drug crisis in Japan and changes leading to the collapse of universal health insurance established in 1961: the case of Kobayashi Kako Co. Ltd. *Cost Effect Resour Alloc.* (2023) 21:1–3. doi: 10.1186/s12962-023-00441-z
21. Jakovljević M, Netz Y, Buttigieg SC, Adany R, Laaser U, Varjacić M. Population aging and migration—history and UN forecasts in the EU-28 and its east and south near neighborhood—one century perspective 1950–2050. *Glob Health.* (2018) 14:1–6. doi: 10.1186/s12992-018-0348-7
22. Yoon K, Kim HK, Choi M, Lee M, Jakovljević M. Analyzing the effectiveness of data-linked projects for health promotion in public health centers of South Korea. *Risk Manag Healthc Policy.* (2023) 16:1101–17. doi: 10.2147/RMHP.S413630
23. Jakovljević M, Chang H, Pan J, Guo C, Hui J, Hu H, et al. Successes and challenges of China's health care reform: a four-decade perspective spanning 1985–2023. *Cost Effect Resour Alloc.* (2023) 21:59. doi: 10.1186/s12962-023-00461-9
24. Jakovljević MB. The key role of the leading emerging BRIC markets in the future of global health care. *Serbian J Exp Clin Res.* (2014) 15:139–43. doi: 10.2478/sjccr-2014-0018
25. Breiman L. Random forests. *Mach Learn.* (2001) 45:5–32. doi: 10.1023/A:1010933404324
26. Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, Marshall S. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiol Meas.* (2014) 35:2191–203. doi: 10.1088/0967-3334/35/11/2191
27. Schonlau M, Zou RY. The random forest algorithm for statistical learning. *Stata J.* (2020) 20:3–29. doi: 10.1177/1536867X20909688
28. Kumagai N, Ogura S. Persistence of physical activity in middle age: a nonlinear dynamic panel approach. *Eur J Health Econ.* (2014) 15:717–35. doi: 10.1007/s10198-013-0518-8
29. Zhang D, Rahman KMA. Government health expenditure, out-of-pocket payment and social inequality: a cross-national analysis of China and OECD countries. *Int J Health Plann Manag.* (2020) 35:1111–26. doi: 10.1002/hpm.3017
30. Baird K. High out-of-pocket medical spending among the poor and elderly in nine developed countries. *Health Serv Res.* (2016) 51:1467–88. doi: 10.1111/1475-6773.12444
31. Sahoo PM, Rout HS, Jakovljević M. Future health expenditure in the BRICS countries: a forecasting analysis for 2035. *Glob Health.* (2023) 19:49. doi: 10.1186/s12992-023-00947-4
32. Biorac N, Jakovljević M, Stefanović D, Perović S, Janković S. Assessment of diabetes mellitus type 2 treatment costs in the republic of Serbia. *Vojnosanit Pregl.* (2009) 66:271–6. doi: 10.2298/VSP0904271B
33. National Federation of Health Insurance Societies. The 2021 survey on trends in lifestyle-related diseases (2023). Available at: [https://www.kenporen.com/toukei\\_data/pdf/chosa\\_r05\\_06\\_02.pdf](https://www.kenporen.com/toukei_data/pdf/chosa_r05_06_02.pdf) (Accessed December 22, 2023).
34. Vuković M, Gvozdenović BS, Gajić T, McCormick BP. Validation of a patient satisfaction questionnaire in primary health care. *Public Health.* (2012) 126:710–8. doi: 10.1016/j.puhe.2012.03.008
35. Sahoo PM, Rout HS, Jakovljević M. Contemporary universal health coverage in India - the case of Federal State of Odisha (Orissa). *Risk Manag Healthc Policy.* (2023) 16:1131–43. doi: 10.2147/RMHP.S406491
36. Cousin E, Duncan BB, Stein C, Ong KL, Vos T, Abbafati C, et al. Diabetes mortality and trends before 25 years of age: an analysis of the global burden of disease study 2019. *Lancet Diab Endocrinol.* (2022) 10:177–92. doi: 10.1016/S2213-8587(21)00349-1
37. Kovacević A, Dragojević-Simić V, Rancić N, Jurisević M, Gutzwiller FS, Matter-Walstra K, et al. End-of-life costs of medical care for advanced stage cancer patients. *Vojnosanit Pregl.* (2015) 72:334–41. doi: 10.2298/VSP1504334K
38. Ranabhat CL, Jakovljević M, Dhimal M, Kim CB. Structural factors responsible for universal health coverage in low-and middle-income countries: results from 118 countries. *Front Public Health.* (2020) 7:414. doi: 10.3389/fpubh.2019.00414
39. Jakovljević M, Malmose-Stapelfeldt C, Milovanovic O, Rancić N, Bokonić D. Disability, work absenteeism, sickness benefits, and cancer in selected European OECD countries—forecasts to 2020. *Front Public Health.* (2017) 5:23. doi: 10.3389/fpubh.2017.00023
40. Jakovljević MB, Nakazono S, Ogura S. Contemporary generic market in Japan—key conditions to successful evolution. *Expert Rev Pharmacoecon Outcomes Res.* (2014) 14:181–94. doi: 10.1586/14737167.2014.881254
41. Ke X, Evans DB, Kawabata K, Zeramini R, Klavus J, Murray C. Household catastrophic health expenditure: a multicountry analysis. *Lancet.* (2003) 362:111–7. doi: 10.1016/S0140-6736(03)13861-5
42. Iskander D, Picchioni F, Vouch LL, Parsons L, Guermond V, Michiels S, et al. Trapped in the service of debt. How the burdens of repayment are fuelling the health poverty trap in rural Cambodia. Project Report. London: Royal Holloway, University of London (2022).
43. Jakovljević M, Lamnisos D, Westerman R, Chattu VK. Cerda a future health spending forecast in leading emerging BRICS markets in 2030: health policy implications. *Health Res Policy Syst.* (2022) 20:23. doi: 10.1186/s12961-022-00822-5
44. Jakovljević MB. BRIC's growing share of global health spending and their diverging pathways. *Front Public Health.* (2015) 3:135. doi: 10.3389/fpubh.2015.00135
45. Borusz K, Juhász A, Nagy C, Szabó Z, Ádány R. High inequalities associated with socioeconomic deprivation in cardiovascular disease burden and antihypertensive medication in Hungary. *Front Pharmacol.* (2018) 9:839. doi: 10.3389/fphar.2018.00839
46. Wu Y, Shyong Tai E, Heng D, Tan CE, Low LP, Lee J. Risk factors associated with hypertension awareness, treatment, and control in a multi-ethnic Asian population. *J Hypertens.* (2009) 27:190–7. doi: 10.1097/HJH.0b013e328317c8c3
47. Nahavandi D, Alizadehsani R, Khosravi A, Acharya UR. Application of artificial intelligence in wearable devices: opportunities and challenges. *Comput Methods Prog Biomed.* (2022) 213:106541. doi: 10.1016/j.cmpb.2021.106541
48. Dias D, Cunha JPS. Wearable health devices—vital sign monitoring, systems and technologies. *Sensors.* (2018) 18:2414. doi: 10.3390/s18082414
49. Zamani-Alavijeh F, Araban M, Harandy TF, Bastami F, Almasian M. Sources of health care providers' self-efficacy to deliver health education: a qualitative study. *BMC Med Educ.* (2019) 19:1–9. doi: 10.1186/s12909-018-1448-z
50. Charoendee K, Sriratanaban J, Aekplakorn W, Hanvoravongchai P. Assessment of population coverage of hypertension screening in Thailand based on the effective coverage framework. *BMC Health Serv Res.* (2018) 18:1–9. doi: 10.1186/s12913-018-2996-y
51. Mathew RO, Nayer A, Asif A. The endothelium as the common denominator in malignant hypertension and thrombotic microangiopathy. *J Am Soc Hypertens.* (2016) 10:352–9. doi: 10.1016/j.jash.2015.12.007
52. Kondo M, Yamagata K, Hoshi SL, Saito C, Asahi K, Moriyama T, et al. Budget impact analysis of chronic kidney disease mass screening test in Japan. *Clin Exp Nephrol.* (2014) 18:885–91. doi: 10.1007/s10157-014-0943-8
53. Sekimoto M, Ii M. Supplier-induced demand for chronic disease care in Japan: multilevel analysis of the association between physician density and physician-patient encounter frequency. *Val Health Reg Issues.* (2015) 6:103–10. doi: 10.1016/j.vhri.2015.03.010