# Unravelling the mental health status of respondents to population health surveys using tree-based methods

A. Ba, E. Gallic, P. Michel, A. Paraponaris

10th Scientific Symposium of the AHEAD Research Network
December 12, 2025

Aix Marseille Université
amse
école d'économie d'aix-marseille
aix-marseille school of economics

# Motivations

**Toinette, Act II, Scene 2, in Molière, *The imaginary invalid*, 1673**

*He walks, sleeps, eats and drinks like anyone; but it does not exclude that he is extremely ill.*

Photo credit: Christophe Raynaud de Lage, coll. Comédie-Française

# Motivations

- Puzzle: research on **anxious and depressive disorders**
  - **under-reported** by patients and/or **under-diagnosed** by health professionals (Falagas et al., 2007; Freeling et al., 1985; Higgins, 1994; McQuaid et al., 1999; Sheehan, 2004),
  - mental health troubles may be **over-diagnosed** (Aragonès et al., 2006; Klinkman et al., 1998).
- Both situations may lead to:
  - detrimental care,
  - mismatch between people in need and those who receive antidepressant drugs and/or anxiolytics.

# Related Literature

- Unrecognized anxiety-depressive disorders may:
  - be fueled by specific **social** and **occupational** situations;
  - have strong **adverse consequences on outcomes** regarding:
    - health (Falagas et al., 2007),
    - healthcare consumption (Rost et al., 1998; Sheehan, 2004),
    - occupation (Broadhead et al., 1990; Egede, 2007; Asami et al., 2014; Lim et al., 2000; Simon et al., 2001; Hilton et al., 2010; Stewart et al., 2003).

# This paper


Source: Davodeau, E. and Hermenier, C. (2019). *Les couloirs aériens*. Futuropolis.

- Documentation of the big picture of **people with unrecognized mental health troubles**.
- Using **survey data** of French 15+ yo matched with **yearly healthcare consumption** data from the French Sickness Fund.
- **Classification** of people using tree-based machine learning methods.
- Description of **factors associated with non-recognition** of anxiety and depression symptoms (using SHAP values).

# Main Results

Several **profiles** emerge (through descriptive statistics and predictions from ML algo.):

- A strong **income effect**: lower personal disposable income is one of the most important predictors of unrecognised anxiety/depression.
- Marked **occupational arduousness**:
  - feeling of not having enough time to do one's job,
  - low decision latitude,
  - demanding or physically painful working conditions.
- A significant **gender effect**, consistent with epidemiological evidence on the prevalence of anxiety and depressive disorders.
- Reduced **social participation**: weaker engagement in group activities, fewer social interactions (friends, colleagues, organisations).
- **Specific medical consumption patterns**: higher use of pharmacy products and more consultations with general practitioners, partly reflecting chronic conditions.

# Roadmap

# 2. Data

# Source: two datasets matched at the individual level

- **Survey Data** (Enquête Santé et Protection Sociale, ESPS, 2012)
  - Representative sample of individuals aged 15+ covered by social security,
  - Socio-professional characteristics,
  - Self-reported health, chronic conditions, MHI-5 mental health score,
  - Regional characteristics,
  - Working conditions,
  - Social participation and frequency of social interactions.
- **Healthcare consumption data** (French health insurance liquidation data)
  - Expenses, reimbursement, co-payment, extra-fees, deductible, no. medical sessions,
  - Outpatient, GP, Specialist, Pharmacy, Physiotherapist, Nurse, Dentist, Equipment, Transport, Optical, Prostheses, Emergency.
- **After cleaning data and targeting people:** $N = 5,293$.

# 3. Methodology

# 3.1. Classification

# A Classification Problem

The survey population is segmented using two variables:

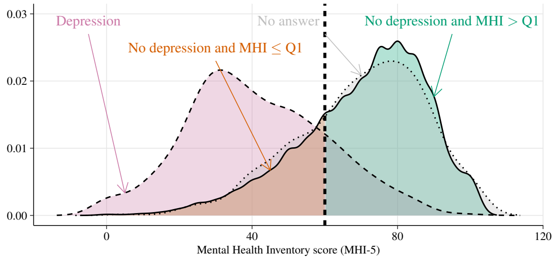
- Self-report of **depression status** (Experienced depression over the past 12 months)
- Calculation of the **Mental Health Inventory (MHI-5) score** (Ware et al., 2001)
  - Based on the answers to 5 items (mental burden, concerning nervousness, self-motivation, peacefulness, sadness and happiness),
  - The MHI-5 scores takes values between 0 (poorest mental health) to 100,
  - The first quartile of the distribution in the sample is $Q_1 = 60$.

# A Classification Problem

Among respondents who **did not report having experienced depression** during the previous 12 months:

- $\overline{D}_{Q1^-}$: people with a **low MHI-5 score** ($\leq Q_1$) (***imaginary healthy patients***)
- $\overline{D}_{Q1^+}$: people with a **high MHI-5 score** ($> Q_1$)

Figure 1: Density of MHI-5 scores according to self-reported depression status.




Notes: The MHI-5 has a score of 0 to 100, where a score of 100 represents optimal mental health. Vertical bar: first quartile of the distribution when all individuals were considered, regardless of their self-reported depression status (MHI-5 score= 60).

# A Classification Problem

Table 1: Size of groups of individuals according to their MHI-5 score and self-reported depression status.

| | Self-reported status | | | | | |
|---|---|---|---|---|---|---|
| | Depression | | **No Depression** | | Total | |
| **MHI-5** $\leq Q_1$ | 572 | (4.8%) | **3,193** | (26.9%) | 3,765 | (31.7%) |
| **MHI-5** $> Q_1$ | 101 | (0.9%) | **8,006** | (67.5%) | 8,107 | (68.2%) |
| Total | 673 | (5.6%) | **11,199** | (94.3%) | 11,872 | (100%) |

Notes: This table shows the number of individuals in each sub-population based on self-reported depression status and the evaluated MHI-5 score. $Q_1$ represents the first quartile of the MHI-5 score distribution for all respondents. The value of $Q_1$ is 60.

In the **sample of indiv. who did not report having experienced depression**:

- **No Depression & MHI-5** $\leq Q_1$: $3,193$ of $11,199$ (28.5%): **Imaginary healthy**
- **No Depression & MHI-5** $> Q_1$: 8,006 of $11,19$ (71.5%).

# Classification Task

- **Objective**: discriminate between $\overline{D}_{Q1^-}$ (**imaginary healthy**) and $\overline{D}_{Q1^+}$ (**healthy**) individuals.
- **Data splitting**:
  - **Training sample** (60%): used for model estimation and **repeated k-fold cross-validation** to tune hyperparameters,
  - **Validation sample** (20%): used to select the **probability threshold** that maximizes **sensitivity**,
  - **Test sample** (20%): held out for the final evaluation of each model.
- **Fit three classifier**: Random Forest, XGBoost, and penalized logistic regression.
- **Model comparison**:
  - Final evaluation on the **test** sample,
  - Primary metric: **sensitivity** (correct identification of $\overline{D}_{Q1^-}$),
  - Secondary metrics reported: specificity, PPV, NPV, and AUC.

# Ensemble Methods

- **Random Forest** (Breiman, 2001) is an **ensemble of decision trees** built on bootstrap samples of the data. Each tree is trained on a slightly different subset of observations.
  - At each split, a **random subset of variables** is considered (decorrelates the trees),
  - Final prediction: average of predicted prob. across all trees.
  - Hyperparameters: `mtry`=9, `min.node.size`=50, `ntree`=500
- **XGBoost** (Chen and Guestrin, 2016) builds an **additive sequence of decision trees**, where each new tree focuses on the **errors made by the previous ones**.
  - Trees are learned sequentially to **minimize a logistic loss**,
  - Misclassified observations receive more weight in later iterations,
  - A learning rate controls how much each tree contributes to the final model.
  - Hyperparameters: `nrounds`=500, `max_depth`=5, `colsample_bytree`=8, `eta`=0.01, `gamma`=5, `min_child_weight`=150, `subsample`=0.9

# Penalized Logistic Regression

- **Penalized Logistic Regression** (Friedman et al., 2010): a logistic regression with a **penalty** added to the loss function to avoid overfitting and to stabilize estimation in the presence of many correlated predictors:

$$\min_{\beta_0,\beta} \frac{1}{N}\sum_{i=1}^{N} l(y_i, \beta_0 + \beta^T x_i) + \underbrace{\lambda\left[(1-\alpha)\|\beta\|_2^2/2 + \alpha\|\beta\|_1\right]}_{\text{elastic net penalty}},$$

where $l(\cdot)$ is the negative log-likelihood, $\alpha$ controls the balance between L1 and L2 penalties, and $\lambda$ controls the overall regularisation strength.

  - Hyperparameters: `alpha`=0.2909091, `lambda`=0.01157483.

# 3.2. Model Explanations

# Explanation of Predictions

- Once the model is estimated, we would like to **explain the predictions**, using an **XAI technique**.
- Generally, XAI techniques are used to **decompose model predictions** into **feature-wise contributions** (Friedman, 2001; Ribeiro et al., 2016; Mothilal et al., 2020).
- **Cooperative game theory** offers a useful framework for attributing predictions to input features, with the **Shapley value** (Shapley, 1951; Hart, 1989) being a widely used allocation rule (Lundberg and Lee, 2017; Heskes et al., 2020; Covert et al., 2021; Zhang and Xu, 2023).
  - In a **collaborative game**, the Shapley value corresponds to the **average expected marginal contribution of a player** (considering all possible combinations of players).

# SHAP

**Objective**: Attribute a ML model prediction $f(\mathbf{x})$, for an individual $\mathbf{x} \in \mathbb{R}^d$, to each of the $d$ features used to train the ML model $f$:

$$f(\mathbf{x}) = \phi_0 + \sum_{j=1}^{d} \phi(j)\,, \text{ with } \phi_0 = \mathbb{E}_{\mathbf{X}}[f(\mathbf{X})] \ .$$

- **SHAP** (Lundberg and Lee, 2017) uses Shapley values from **cooperative game theory** (Shapley, 1951) to compute $\phi(j)$ (**SHAP value**), depending on a **value function** $v$.

$$\forall j \in \{1, \cdots, d\}, \ \ \phi(j) = \sum_{A \subseteq \{1,\cdots,d\} \setminus \{j\}} \frac{|A|!(d - |A| - 1)!}{d!} \left[ v(A \cup \{j\}; \mathbf{x}) - v(A; \mathbf{x}) \right].$$

- To select $v$, **one desired constraint** is that $v(D; \mathbf{x}) = f(\mathbf{x})$. We can choose:

$$v(A; \mathbf{x}) = \mathbb{E}[f(\mathbf{X}) \mid \mathbf{X}_A = \mathbf{x}_A]\,, \text{ for the individual } \mathbf{x} = (\mathbf{x}_A, \mathbf{x}_{A^c}) \ .$$

# 3.3. Clustering

# Clustering on SHAP values

- **SHAP values** are computed **at the level of each observation**.
- It may be possible to **group the respondents** depending on their SHAP values.
- We focus on people predicted "**Imaginary healthy patient**" by the model.
- To do so:
  - **Hierarchical clustering**,
  - Selecting the number of groups depending on the Silhouette score Rousseeuw (1987).

# 4. Results

# 4.1. Descriptive Statistics

# Descriptive Statistics

| Variable | Self-reported No depression ($n = 5,305$) | MHI-5 $\leq Q1$ ($n = 1,598$) | MHI-5 $> Q1$ ($n = 3,707$) | p-value |
|---|---|---|---|---|
| MHI-5 Score | 70 ($\pm$18) | 48 ($\pm$12) | 80 ($\pm$10) | $< 10^{-3}$ |
| **Social and demographic characteristics** | | | | |
| Age | 49 ($\pm$19) | 50 ($\pm$18) | 48 ($\pm$18) | $< 10^{-3}$ |
| Gender: *Female* | 52% | 59% | 49% | $< 10^{-3}$ |
| Couple | 65% | 60% | 68% | $< 10^{-3}$ |
| **Health status and healthcare consumption** | | | | |
| *Good/very good SAH* | 93% | 84% | 97% | $< 10^{-3}$ |
| Self-reported long-term condition | 19% | 27% | 16% | $< 10^{-3}$ |
| Long-term cond. recog. by health insur. | 19% | 26% | 17% | $< 10^{-3}$ |
| No. GP visits | 4.7 ($\pm$5.0) | 6.1 ($\pm$6.2) | 4.1 ($\pm$4.3) | $< 10^{-3}$ |
| No. Specialists visits | 3.4 ($\pm$4.4) | 4.1 ($\pm$5.0) | 3.1 ($\pm$4.1) | $< 10^{-3}$ |
| Outpatient expenses (€) | 1470 ($\pm$2528) | 2024 ($\pm$3306) | 1230 ($\pm$2060) | $< 10^{-3}$ |
| Waiver GP | 3% | 7% | 2% | $< 10^{-3}$ |
| Waiver Dental care | 13% | 19% | 10% | $< 10^{-3}$ |

# Descriptive Statistics: Health Condition

| Variable | Self-reported No depression ($n = 5,305$) | MHI-5 $\leq Q1$ ($n = 1,598$) | MHI-5 $> Q1$ ($n = 3,707$) | p-value |
|---|---|---|---|---|
| **Self-reported health-related conditions** | | | | |
| Asthma | 6.8% | 10.3% | 5.2% | $< 10^{-3}$ |
| Bronchitis | 5.8% | 10.0% | 4.0% | $< 10^{-3}$ |
| Heart Attack | 0.6% | 1.2% | 0.4% | $0.002$ |
| Artery Disease | 1.9% | 3.1% | 1.5% | $< 10^{-3}$ |
| Hypertension | 11.9% | 16.0% | 10.2% | $< 10^{-3}$ |
| Stroke | 0.5% | 0.9% | 0.4% | $0.037$ |
| Osteoarthritis | 13.36% | 18.7% | 11.0% | $< 10^{-3}$ |
| Low Back Pain | 20.4% | 28.7% | 16.8% | $< 10^{-3}$ |
| Neck Pain | 14.6% | 22.1% | 11.4% | $< 10^{-3}$ |
| Diabetes | 8.4% | 14.0% | 6.0% | $< 10^{-3}$ |
| Allergy | 14.6% | 19.6% | 12.4% | $< 10^{-3}$ |
| Cirrhosis | 0.1% | 0.2% | 0.1% | $0.253$ |
| Urinary Incontinence | 4.2% | 7.3% | 2.8% | $< 10^{-3}$ |

# Descriptive Statistics: Other variables

| Variable | Self-reported No depression ($n = 5,305$) | MHI-5 $\leq Q1$ ($n = 1,598$) | MHI-5 $> Q1$ ($n = 3,707$) | p-value |
|---|---|---|---|---|
| **SES and working conditions** | | | | |
| Personal disposable income (€) | 1609 ($\pm$1008) | 1424 ($\pm$806) | 1690 ($\pm$1074) | $< 10^{-3}$ |
| Has to hurry to do job (always/often) | 24% | 29% | 22% | $< 10^{-3}$ |
| Very little freedom to do job, always/often | 8% | 11% | 7% | $< 10^{-3}$ |
| Job allows to learn new things, always/often | 25% | 21% | 28% | $< 10^{-3}$ |
| Colleagues help to carry out tasks, always/often | 20% | 16% | 22% | $< 10^{-3}$ |
| Repetitive work under time constraints, always/often | 9% | 9% | 7% | $< 10^{-3}$ |
| Must carry heavy loads, always/often | 11% | 12% | 10% | $< 10^{-3}$ |
| Painful postures, always/often | 17% | 19% | 17% | $< 10^{-3}$ |
| Harmful/toxic substances/products, always/often | 10% | 10% | 10% | 0.002 |
| **Social participation** | | | | |
| Participation in group activities | 35% | 29% | 38% | $< 10^{-3}$ |
| Meeting w/ family outside hh, everyday/$\geq$1 a week | 56% | 54% | 58% | $< 10^{-3}$ |
| Meeting w/ people in organizations, everyday/$\geq$1 a week | 20% | 16% | 22% | $< 10^{-3}$ |
| Meeting w/ colleagues outside work, everyday/$\geq$1 a week | 14% | 12% | 14% | $< 10^{-3}$ |

# 4.2. Classification with Ensemble Machine Learning

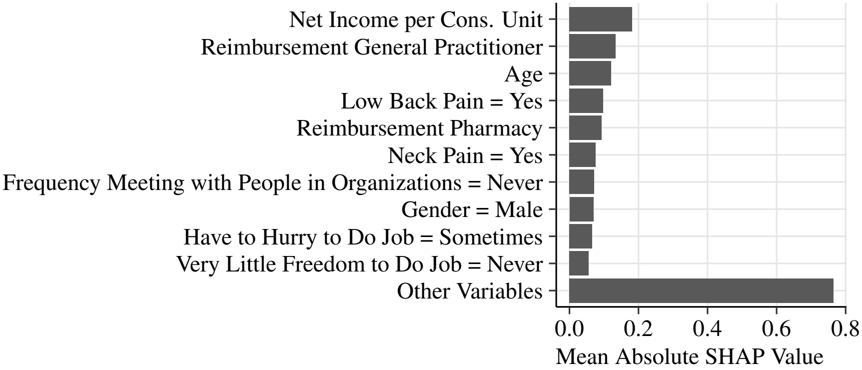
# Performance of the Estimation

| Method | Accuracy | | Sensitivity | | Specificity | | ROC-AUC | |
|---|---|---|---|---|---|---|---|---|
| | Train | Test | Train | Test | Train | Test | Train | Test |
| Random Forest | 0.812 | 0.726 | 0.383 | 0.144 | 0.997 | 0.977 | 0.96 | 0.72 |
| **XGBoost** | 0.676 | 0.667 | 0.666 | 0.690 | 0.680 | 0.658 | 0.74 | 0.71 |
| Pen. Log. Reg. | 0.740 | 0.749 | 0.276 | 0.301 | 0.941 | 0.942 | 0.75 | 0.73 |

- **Accuracy**: Overall proportion of correct classifications (**Imaginary Healthy** + **Healthy**) among all predictions.
- **Sensitivity**: measures how well the model detects the **Imaginary Healthy** (proportion of **imaginary healthy** who are correctly classified as such).
- **Specificity**: the proportion of **healthy individuals** correctly classified as healthy.
- **ROC–AUC**: summarizes the trade-off between sensitivity and specificity across all thresholds (higher AUC: better overall discrimination).

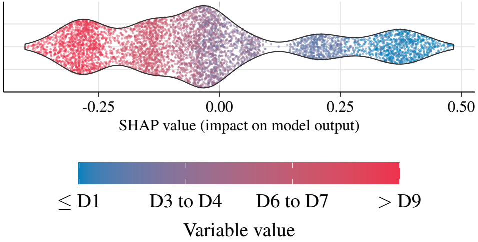
# 4.3. Interpretation with SHAP Values

# Variable Importance (SHAP Values)

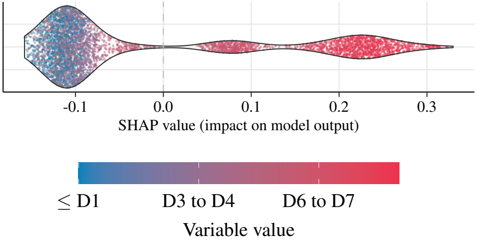


# Variable Importance (2/4)

## Impact of Variables on the Prob. of Being Classified as an **Imaginary Healthy Patient**.




(a) Net Income per Cons. Unit (€)

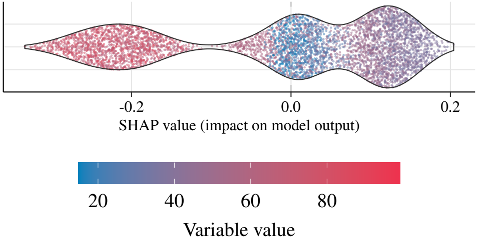



(b) Reimbursement General Practitioner (€)

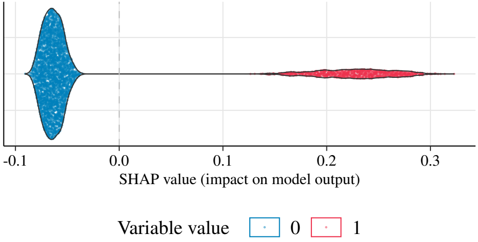
Note: The plots are ordered by variable importance with respect to the average absolute SHAP values. Each dot represents an individual. For points with a negative abscissa, the variable of interest has a downward effect on the probability of being classified as an imaginary healthy patient ($\overline{D}_{Q1-}$). For quantitative variables, the color of the points depends on the level of the variable of interest, ranging from blue for low values to red for high values. For the net income per consumption unit variable, the color scale ranges according to the empirical quantiles of the variable.

# Variable Importance (3/4)

Impact of Variables on the Prob. of Being Classified as an **Imaginary Healthy Patient**.




(a) Age

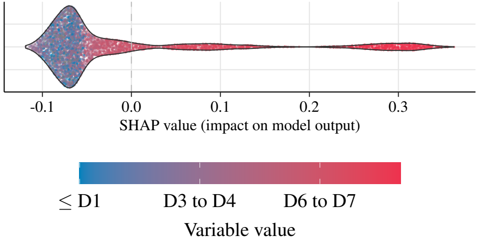



(b) Low Back Pain

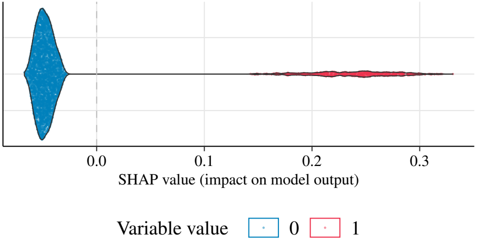
Note: The plots are ordered by variable importance with respect to the average absolute SHAP values. Each dot represents an individual. For points with a negative abscissa, the variable of interest has a downward effect on the probability of being classified as an imaginary healthy patient ($\overline{D}_{Q1-}$). For quantitative variables, the color of the points depends on the level of the variable of interest, ranging from blue for low values to red for high values.

# Variable Importance (4/4)

Impact of Variables on the Probability of Being Classified as an **Imaginary Healthy Patient**.




(a) Deduct. Pharmacy (€)

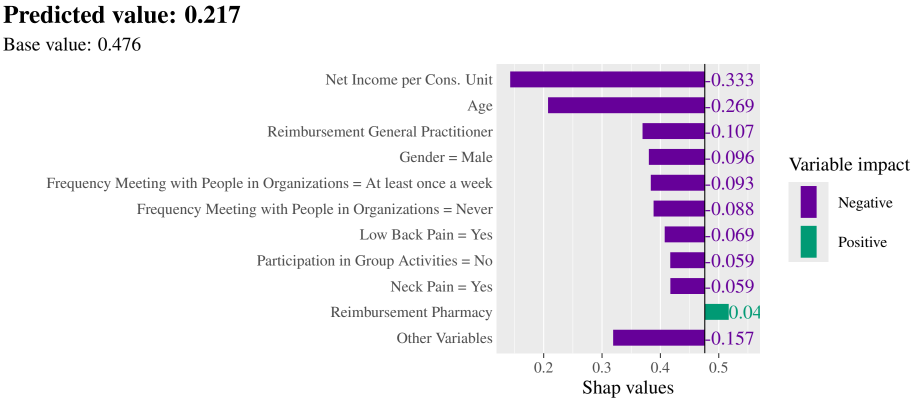



(b) Neck Pain

Note: The plots are ordered by variable importance with respect to the average absolute SHAP values. Each dot represents an individual. For points with a negative abscissa, the variable of interest has a downward effect on the probability of being classified as an imaginary healthy patient ($\overline{D}_{Q1-}$). The color of the points depends on the level of the variable of interest, ranging from blue for low values to red for high values.

# Individual Effects (1/2)

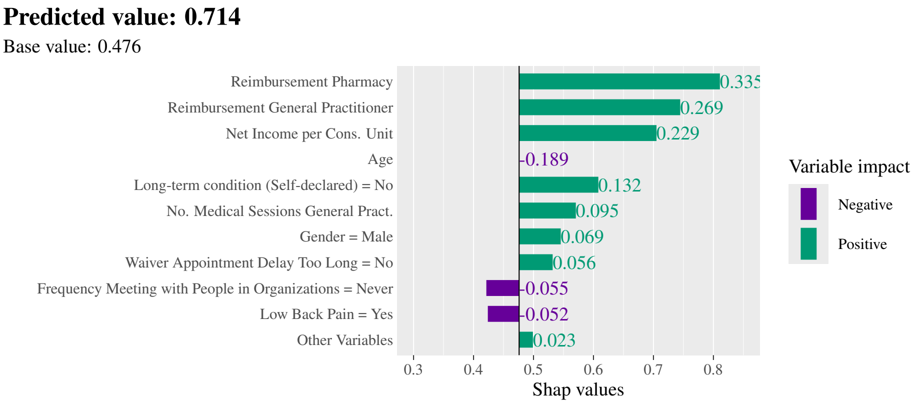
Decomposition of the Contribution of the Most Influential Variables to the Prediction Deviation of Being Classified as an **Imaginary Healthy** Patient from the Baseline Value, for an **Individual with a Low Predicted Probability**




Note: Baseline value: average probability of being classified by the preferred model as an imaginary healthy patient ($\overline{D}_{Q1-}$) in the dataset.

# Individual Effects (2/2)

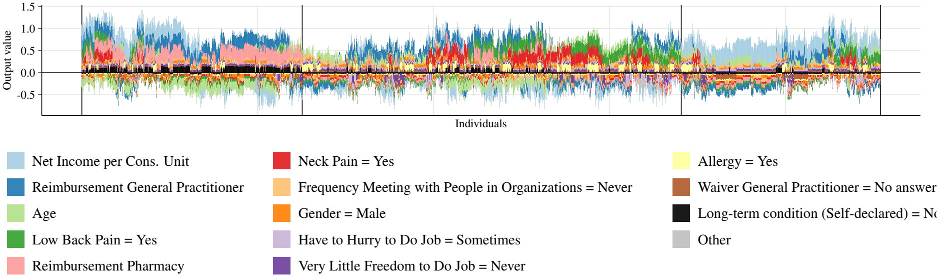
Decomposition of the Contribution of the Most Influential Variables to the Prediction Deviation of Being Classified as an **Imaginary Healthy** Patient from the Baseline Value, for an **Individual with a High Predicted Probability**




Note: Baseline value: average probability of being classified by the preferred model as an imaginary healthy patient ($\overline{D}_{Q1-}$) in the dataset.

# Clustering (1/2)

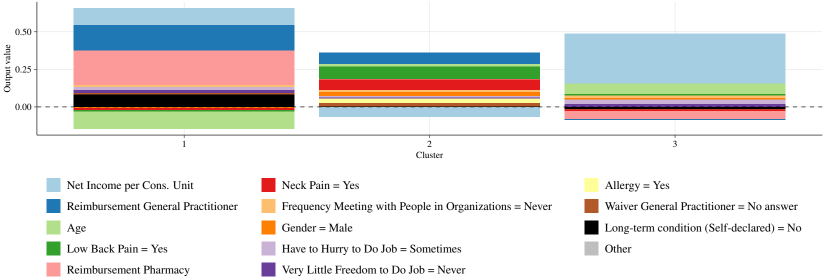
## Decomposition of the Effect of the Most Important Variables on the Probability of Being Predicted as an Imaginary Healthy Patient for Individuals Predicted as Such.




Note: The reported contribution is relative to the baseline value (the average probability of being classified by the preferred model as an imaginary healthy patient ($\overline{D}_{Q1-}$) in the dataset) centered in zero. The observations are ordered according to their relative distance from each other (Ward's distance).

# Clustering (2/2)

## Decomposition of the Effect of the Most Important Variables on the Probability of Being Predicted as an Imaginary Healthy Patient for Individuals Predicted as Such.




la

Note: The reported contribution is relative to the baseline value (the average probability of being classified by the preferred model as an imaginary healthy patient ($\overline{D}_{Q1-}$) in the dataset) centered in zero.

(1) elderly & high-consumption patients; (2) average profiles with pain disorders; (3) younger economically deprived individuals with low healthcare use.

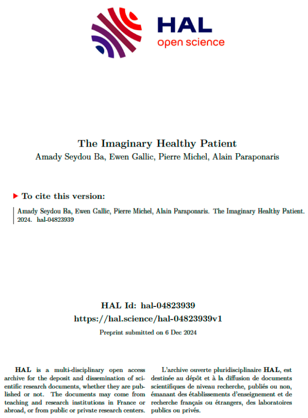
# 5. Conclusion

# Conclusion

- Use of Machine Learning techniques to characterize individuals unaware of the presence of anxiety and depressive disorders
- Main results:
  - A significant **gender effect** consistent with epidemiological knowledge of the prevalence of anxiety and depressive disorders in the general population.
  - An **income effect**.
  - An important influence of **working and employment conditions** (low decision latitude, work intensity, demanding schedules).
  - Contributions from specific **patterns of medical consumption** (GP visits, pharmacy expenditures) → unrecognised underlying health problems in several clusters.
  - A **social participation** effect.
  - **Three distinct subgroups** of **imaginary-healthy** individuals: (1) elderly & high-consumption patients; (2) average profiles with pain disorders; (3) younger economically deprived individuals with low healthcare use.
- **Robustn. checks**: MHI-3 and Self-Assessed Health used as substitutes to MHI-5.

# Research Agenda

- **Potential explanations for misperception of personal mental health troubles**
  - Psychometric properties of MHI-5: sensitivity and specificity,
  - Disease denial and illnesses masked by drugs,
  - Differential item functioning.
- Use of alternative measures from cooperative game theory as substitute to **Shapley's values** to measure the marginal contributions of different characteristics of each individual on the probability of ignoring their bad health condition (*e.g.*, the core).


HAL open science

**The Imaginary Healthy Patient**

Amady Seydou Ba, Ewen Gallic, Pierre Michel, Alain Paraponaris

▶ **To cite this version:**

Amady Seydou Ba, Ewen Gallic, Pierre Michel, Alain Paraponaris. The Imaginary Healthy Patient. 2024. hal-04823939

**HAL Id: hal-04823939**

**https://hal.science/hal-04823939v1**

Preprint submitted on 6 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HAL

Working Paper


N° 6 novembre-décembre 2024 BIMESTRIELLE PP. 801-986

Revue d'économie politique

Machine Learning et Économie

**Arthur Charpentier, Emmanuel Flachaire**

Machine Learning et économies

**Amady Seydou Ba, Ewen Gallic, Pierre Michel, Alain Paraponaris**

The Imaginary Healthy Patient — *Le bien portant imaginaire*

**Isabelle Cadoret, Aureline Fargeas, Véronique Thelen**

Prediction of Energy Poverty in France: A Machine Learning Approach — *Méthodes de Machine Learning pour la prédiction de la précarité énergétique en France*

**Pierre Durand, Gaëtan Le Quang, Arnold Vialfont**

Non-linéarités et interactions entre variables : quelques apports des méthodes d'apprentissage automatique interprétables à la régulation bancaire — *Nonlinerities and Interations between Variables: some Contributions of Machine Learning Methods Interpretable to Banking Regulation*

**Thomas Chuffart, Cyril Dell'Eva**

Media Coverage of the ECB: a Textual Analysis — *La couverture médiatique relative à la BCE : une analyse textuelle*

**Charles Condevaux, Stéphane Mussard**

Le traitement automatisé du language naturel : l'apport des *transformers* et les enjeux pour les sciences économiques — *Automated Natual Language Processing: The Contribution of Transformers and Challenges for Economics*

Lefebvre Dalloz

DALLOZ

Special issue:

Machine learning & economics

# A. References

# References I

Aragonès, E., Piñol, J. L. and Labad, A. (2006). The overdiagnosis of depression in non-depressed patients in primary care. *Family Practice* 23: 363–368, doi: 10.1093/fampra/cmi120.

Asami, Y., Goren, A. and Okumura, Y. (2014). Work Productivity Loss with Depression, Diagnosed and Undiagnosed, among Employed Respondents in an Internet-Based Survey Conducted in Japan. *Value in Health* 17: A463, doi: 10.1016/j.jval.2014.08.1289.

Breiman, L. (2001). Random Forests. *Machine Learning* 45: 5–32, doi: 10.1023/A:1010933404324.

Broadhead, W. E., Blazer, D. G., George, L. K. and Tse, C. K. (1990). Depression, disability days, and days lost from work in a prospective epidemiologic survey. *JAMA* 264: 2524–2528, doi: 10.1001/jama.1990.03450190056028.

Chen, T. and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16. New York, NY, USA: Association for Computing Machinery, 785–794, doi: 10.1145/2939672.2939785.

Covert, I., Lundberg, S. and Lee, S.-I. (2021). Explaining by removing: A unified framework for model explanation. *Journal of Machine Learning Research* 22: 1–90.

Egede, L. E. (2007). Failure to Recognize Depression in Primary Care: Issues and Challenges. *Journal of General Internal Medicine* 22: 701–703, doi: 10.1007/s11606-007-0170-z.

# References II

Falagas, M. E., Vardakas, K. Z. and Vergidis, P. I. (2007). Under-diagnosis of common chronic diseases: prevalence and impact on human health. *International Journal of Clinical Practice* 61: 1569–1579, doi: 10.1111/j.1742-1241.2007.01423.x.

Freeling, P., Rao, B. M., Paykel, E. S., Sireling, L. I. and Burton, R. H. (1985). Unrecognised depression in general practice. *Br Med J (Clin Res Ed)* 290: 1880–1883, doi: 10.1136/bmj.290.6485.1880.

Friedman, J., Hastie, T. and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33, doi: 10.18637/jss.v033.i01.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics* 29: 1189–1232, doi: 10.1214/aos/1013203451.

Hart, S. (1989). Shapley Value. In Eatwell, J., Milgate, M. and Newman, P. (eds), *Game Theory*, The New Palgrave. London: Palgrave Macmillan UK, 210–216, doi: 10.1007/978-1-349-20181-5_25.

Heskes, T., Sijben, E., Bucur, I. G. and Claassen, T. (2020). Causal Shapley Values: Exploiting Causal Knowledge to Explain Individual Predictions of Complex Models. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. and Lin, H. (eds), *Advances in Neural Information Processing Systems*, *33*. Curran Associates, Inc., 4778–4789.

Higgins, E. S. (1994). A review of unrecognized mental illness in primary care. Prevalence, natural history, and efforts to change the course. *Archives of Family Medicine* 3: 908–917, doi: 10.1001/archfami.3.10.908.

# References III

Hilton, M. F., Scuffham, P. A., Vecchio, N. and Whiteford, H. A. (2010). Using the interaction of mental health symptoms and treatment status to estimate lost employee productivity. *Australian and New Zealand Journal of Psychiatry* 44: 151–161, doi: 10.3109/00048670903393605.

Klinkman, M. S., Coyne, J. C., Gallo, S. and Schwenk, T. L. (1998). False Positives, False Negatives, and the Validity of the Diagnosis of Major Depression in Primary Care. *Archives of Family Medicine* 7: 451, doi: 10.1001/archfami.7.5.451.

Lim, D., Sanderson, K. and Andrews, G. (2000). Lost productivity among full-time workers with mental disorders. *The Journal of Mental Health Policy and Economics* 3: 139–146, doi: 10.1002/mhp.93.

Lundberg, S. M. and Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S. and Garnett, R. (eds), *Advances in Neural Information Processing Systems*, *30*. Curran Associates, Inc.

McQuaid, J. R., Stein, M. B., Laffaye, C. and McCahill, M. E. (1999). Depression in a Primary Care Clinic: the Prevalence and Impact of an Unrecognized Disorder. *Journal of Affective Disorders* 55: 1–10, doi: 10.1016/S0165-0327(98)00191-8.

Mothilal, R. K., Sharma, A. and Tan, C. (2020). Explaining machine learning classifiers through diverse counterfactual explanations. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20. New York, NY, USA: Association for Computing Machinery, 607–617, doi: 10.1145/3351095.3372850.

# References IV

Ribeiro, M. T., Singh, S. and Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16. New York, NY, USA: Association for Computing Machinery, 1135–1144, doi: 10.1145/2939672.2939778.

Rost, K., Zhang, M., Fortney, J., Smith, J., Coyne, J. and Smith, G. R. (1998). Persistently poor outcomes of undetected major depression in primary care. *General Hospital Psychiatry* 20: 12–20, doi: 10.1016/s0163-8343(97)00095-9.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20: 53–65, doi: 10.1016/0377-0427(87)90125-7.

Shapley, L. S. (1951). Notes on the n-Person Game – II: The Value of an n-Person Game. Researach Memorandum ATI 210720, RAND Corporation, Santa Monica, California.

Sheehan, D. V. (2004). Depression: underdiagnosed, undertreated, underappreciated. *Managed Care (Langhorne, Pa.)* 13: 6–8.

Simon, G. E., Barber, C., Birnbaum, H. G., Frank, R. G., Greenberg, P. E., Rose, R. M., Wang, P. S. and Kessler, R. C. (2001). Depression and Work Productivity: The Comparative Costs of Treatment Versus Nontreatment. *Journal of Occupational and Environmental Medicine* 43: 2, doi: 10.1097/00043764-200101000-00002.

Stewart, W. F., Ricci, J. A., Chee, E., Hahn, S. R. and Morganstein, D. (2003). Cost of Lost Productive Work Time Among US Workers With Depression. *JAMA* 289: 3135–3144, doi: 10.1001/jama.289.23.3135.

# References V

Ware, J. E., Kosinski, M. and Dewey, J. E. (2001). *How to score version 2 of the SF-36 health survey: (standard & acute forms) ; [SF-36v2]*. Lincoln, RI: QualityMetric, 3rd ed.

Zhang, N. and Xu, H. (2023). Fairness of ratemaking for catastrophe insurance: Lessons from machine learning. *Information Systems Research* 35: 469–488, doi: 10.1287/isre.2022.1195.

# B. MHI-5 Score

# MHI-5 score

The MHI-5 score is calculated using the answers to 5 questions:

**In the last 4 weeks, how often:**

1. Did you feel very nervous?
2. Did you feel calm and relaxed?
3. Did you feel sad and downcast?
4. Have you felt happy?
5. You felt so discouraged that nothing could cheer you up?

In the **French version**, respondents can choose between the following answers, for each question:

- Always (1)
- Most of the time (2)
- Sometimes (3)
- Rarely (4)
- Never (5)

# MHI-5 score

- Each answer by respondent $j = \{1, 2, \ldots, N\}$ to question $i = \{1, 2, \ldots, 5\}$ is transcribed using its numerical value.
- For the answers to the questions about feeling calm and happy, the scale is reversed.
- The MHI-5 score of individual $j$ is then calculated as follows:

$$\text{MHI-5}_j = 100 \times \frac{\sum_{i=1}^{5} A_{i,j} - \min\left(\sum_{j=1}^{N}\sum_{i=1}^{5} A_{i,j}\right)}{\max\left(\sum_{j=1}^{N}\sum_{i=1}^{5} A_{i,j}\right) - \min\left(\sum_{j=1}^{N}\sum_{i=1}^{5} A_{i,j}\right)}$$

- The values, by construction, range between 0 and 100.
- The higher the score, the better the mental health.

Go back

# C. Estimation

# Hyperparameters

Table 2: Hyperparameters.

| Hyperparameter | Possible Values | Description |
| --- | --- | --- |
| | *Random Forest* | |
| `mtry` | {3, 4, 5, 6, 7, 8, **9**} | No. variables sampled as candidates at each split |
| `splitrule` | Gini index | Splitting rule |
| `min.node.size` | {**50**, 75, 100, 150} | Minimum size of terminal nodes |
| | *Extreme Gradient Boosting* | |
| `nrounds` | 500 | No. boosting iterations |
| `max_depth` | {3, 4, **5**, 6} | Maximum depth of a tree |
| `colsample_bytree` | {.1, .2, ..., **.8**, .9} | Subsample ratio of col. when building each tree |
| `eta` | 0.01 | Learning rate |
| `gamma` | {0, **5**, 10} | Min loss reduction for further partition on a leaf node |
| `min_child_weight` | {50, 100, **150**} | Min sum of instance weight needed in a child |
| `subsample` | {0.7, 0.8, **0.9**, 1} | Subsample ratio of the training instances |
| | *Penalised Logistic Regression Model.* | |
| `alpha` | Sequence of equally distant values from 0.1 to 1 with a length of 100 | Elasticnet mixing parameter |

# Clustering (1/3)

- **Input for clustering**:
  - Use **SHAP values** as individual-level explanations,
  - For each variable, compute the **mean SHAP value** across individuals,
  - Compute the **overall mean** of these variable-wise averages,
  - Retain only variables whose **mean SHAP value** is **above the overall mean**.
- **Hierarchical clustering**:
  - Perform **hierarchical clustering** on the **selected SHAP values**,
  - Use **Euclidean distance** on **SHAP vectors** as the dissimilarity measure.
- **Choosing the number of clusters $K$**:
  - For each $K = 2, \dots, 15$: run clustering and compute the **silhouette score** (Rousseeuw, 1987),
  - Select the $K$ **that maximises** the average silhouette score.

# Clustering (2/3)

Table 3: Average Person in Each Cluster and in the Samples.

| | Cluster 1 | Cluster 2 | Cluster 3 | Pred. Imaginary | Pred. Healthy | Entire Sample |
|---|---|---|---|---|---|---|
| | $n = 626$ | $n = 1,077$ | $n = 566$ | $n = 2,269$ | $n = 3,024$ | $n = 5,293$ |
| *Imaginary healthy* | 306(48.9%) | 517(48.0%) | 248(43.8%) | 1071(47.2%) | 526(17.4%) | 1597(30.2%) |
| Accuracy | 48.9% | 48.0% | 43.8% | 47.2% | 82.6% | 67.4% |
| Net Income per Cons. Unit | 1193.67 (687.48) | 1640.95 (691.14) | 643.45 (192.42) | 1268.72 (728.7) | 1865.21 (1108.24) | 1609.51 (1008.13) |
| Reimbursement GP | 228.85 (169.81) | 131.15 (119.72) | 85.79 (108.71) | 146.79 (143.51) | 46.41 (48.95) | 89.44 (112.53) |
| Age | 66.15 (15.23) | 49.29 (16.48) | 38.39 (12.82) | 51.22 (18.39) | 46.8 (18.49) | 48.7 (18.58) |
| Low Back Pain | No (81.6%) | No (50.3%) | No (75.6%) | No (65.3%) | No (90.4%) | No (79.6%) |
| Reimbursement Pharmacy | 1794.9 (3321.96) | 306.26 (669.55) | 99.92 (178.35) | 665.5 (1937.75) | 133.41 (805.81) | 361.5 (1431.61) |
| Neck Pain | No (88%) | No (57.9%) | No (86.9%) | No (73.5%) | No (94.3%) | No (85.4%) |
| Freq. Meet. w/ People in Org. | Never (64.2%) | Never (59.4%) | Never (66.8%) | Never (62.6%) | Never (42.9%) | Never (51.3%) |
| Gender | Female (51%) | Female (71.7%) | Female (62.9%) | Female (63.8%) | Male (56.4%) | Female (52.2%) |
| Have to Hurry to Do Job | No answer (85.8%) | No answer (37.5%) | No answer (71.6%) | No answer (59.3%) | No answer (46.2%) | No answer (51.8%) |
| Very Little Freedom to Do Job | No answer (85.9%) | No answer (37.8%) | No answer (72.1%) | No answer (59.6%) | No answer (46.4%) | No answer (52.1%) |
| Allergy | No (86.7%) | No (69.5%) | No (83.9%) | No (77.8%) | No (91.1%) | No (85.4%) |
| Waiver GP | No (81%) | No (85.1%) | No (64%) | No (78.7%) | No (74%) | No (76%) |
| Long-term condition (Self-declared) | Yes (74.1%) | No (79.4%) | No (89.9%) | No (67.2%) | No (90.5%) | No (80.5%) |

Notes: This table shows the representative person in each of the three clusters in the set of individuals predicted as *imaginary healthy*, in the set of individuals predicted as healthy, and in the entire sample. For numerical variables, the within-cluster means are reported (with standard errors between brackets), whereas for categorical variables, the within-cluster mode is provided (the proportions are given in brackets). The first rows of the table give the number of *imaginary healthy* and the corresponding proportion in each sample, and the accuracy of the XBG model in that sample. GP stands for general practitioners.

Go back

# Clustering (3/3)

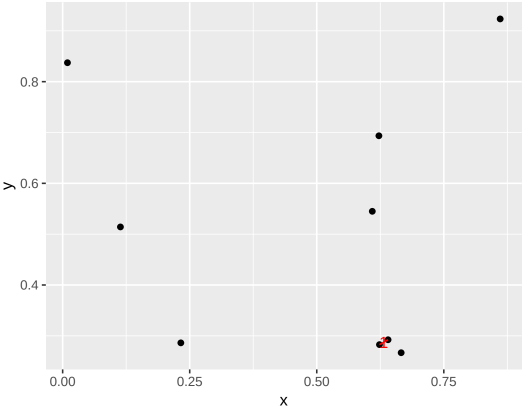



Figure 5: Step-by-step example of the hierarchical clustering algorithm.

Go back