Journal of Visualized Experiments

A machine learning approach to design an efficient selective screening of mild cognitive impairment

--Manuscript Draft--

Article Type:	Invited Methods Article - Author Produced Video
Manuscript Number:	JoVE59649R3
Full Title:	A machine learning approach to design an efficient selective screening of mild cognitive impairment
Section/Category:	JoVE Medicine
Keywords:	memory complaint; early detection; Mild Cognitive Impairment; sleep duration; community pharmacists; risk factors; decision trees; statistical learning
Corresponding Author:	Francisco Javier Munoz-Almaraz, Ph.D Universidad CEU Cardenal Herrera Alfara del Patriarca, Valencia SPAIN
Corresponding Author's Institution:	Universidad CEU Cardenal Herrera
Corresponding Author E-Mail:	malmaraz@uchceu.es
Order of Authors:	Francisco Javier Munoz-Almaraz
	Maria Teresa Climent
	Maria Dolores Guerrero Masia
	Lucrecia Moreno Royo
	Juan Pardo Albiach
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US$1200)

TITLE:

A Machine Learning Approach to Design an Efficient Selective Screening of Mild Cognitive Impairment

AUTHORS AND AFFILIATIONS:

Francisco J Muñoz-Almaraz[1], María Teresa Climent[2], María Dolores Guerrero[3], Lucrecia Moreno[3], Juan Pardo[1]

[1]Embedded Systems and Artificial Intelligence Group, Universidad CEU Cardenal Herrera, Valencia, Spain
[2]Community Pharmacist, Valencia, Spain
[3]Department of Pharmacy, Universidad CEU Cardenal Herrera, Valencia, Spain

Corresponding Author:
Juan Pardo Albiach
juaparal@uchceu.es

Email Addresses of Co-authors:
FJ Muñoz-Almaraz (malmaraz@uchceu.es)
MT Climent (m.climent.000@micof.es)
MD Guerrero (ma_dolores.guerrero@uchceu.es)
L Moreno (lmoreno@uchceu.es)

KEYWORDS:

Memory complaint, early detection, mild cognitive impairment, sleep duration, community pharmacist, risk factors, decision trees, statistical learning

SUMMARY:

This methodology produces decision trees that target population groups more prone to suffering from mild cognitive impairment and are useful for cost-effective selective screening of the disease.

ABSTRACT:

Mild cognitive impairment (MCI) is the first sign of dementia among elderly populations and its early detection is crucial in our aging societies. Common MCI tests are time-consuming such that indiscriminate massive screening would not be cost-effective. Here, we describe a protocol that uses machine learning techniques to rapidly select candidates for further screening via a question-based MCI test. This minimizes the number of resources required for screening because only patients who are potentially MCI positive are tested further.

This methodology was applied in an initial MCI research study that formed the starting point for the design of a selective screening decision tree. The initial study collected many demographic and lifestyle variables as well as details about patient medications. The *Short Portable Mental Status Questionnaire* (SPMSQ) and the *Mini-Mental State Examination* (MMSE) were used to

detect possible cases of MCI. Finally, we used this method to design an efficient process for classifying individuals at risk of MCI. This work also provides insights into lifestyle-related factors associated with MCI that could be leveraged in the prevention and early detection of MCI among elderly populations.

INTRODUCTION:

Population aging is increasing the prevalence of chronic and degenerative diseases, especially degenerative dementias, which are expected to affect more than 131 million people worldwide by 2050[1]. Among all the degenerative dementias, Alzheimer's disease (AD) is the most common with an overall prevalence in Europe of 6.88%[2]. Due to the ever-declining independence of AD patients, this group should start receiving support as soon as AD starts to manifest. Therefore, the early detection of prodromal signs of AD, such as mild cognitive impairment (MCI), is essential.

MCI is defined as an intermediate cognitive decline stage corresponding to normal aging and severe deterioration due to dementia[3]. According to estimates by Petersen et al.[4], the prevalence of MCI is 8.4% among people aged 65–69 years and reaches 25.2% for those aged over 80 years. MCI results in individuals experiencing more difficulties than expected in the execution of low-level cognitive skills, especially those related to memory and language, but does not interfere with the activities of daily living.

Screening is not synonymous with diagnosis; the diagnosis of MCI will always be a clinical task whereas screening methods can only inform us that a patient has a higher probability of suffering from this pathology and that there is a well-founded suspicion of MCI that should be confirmed clinically. Hence, primary healthcare workers (doctors, pharmacists, nurses, etc.) could benefit from the availability of simple screening methods (brief cognitive tests) that can be applied in minutes. Ideally, these would objectively identify patients with a high probability of suffering an MCI so that they can then be clinically tested by general or specialized physicians.

Given that the early detection of MCI is becoming an essential task within the context of public health, this work aimed to identify which characteristics are useful in the targeted identification of MCI in screening tests of elderly populations. These groups would then be more thoroughly tested for MCI in tests administered by primary health care providers. This methodology provides a decision tree with the appropriate algorithms for identifying the population groups to target.

Among these characteristics, age is one of the most consistent factors associated with the development of this pathology. Other relevant characteristics are related to demographics or lifestyle[5]. Among the latter, some studies have identified the duration of daytime or nighttime sleep as a risk factor that can lead to the diagnosis of MCI[5–9]. The prolonged consumption of medications such as benzodiazepines, consumed by an estimated 20%–25% of older adults[10,11], can also influence sleep hours and the development of MCI[12,13]. Indeed, prolonged treatments for chronic diseases may be important features useful in the pre-selection of individuals with a high risk of suffering from MCI.

Here, we developed data-based models that use automatic learning algorithms, a decision tree, and a predictive tool to increase the efficiency of the methodology for detecting MCI by discriminating which characteristics play an important role in the early detection of MCI. The resultant decision tree presented here was produced using a specific cohort of Spanish patients using community pharmacies. However, this method would also be useful among other populations with different characteristics.

This work was completed in collaboration with primary healthcare and specialized medical doctors. Community pharmacies were ideal for testing this algorithm because they are close to patients, have long opening hours, and are frequently visited and consulted. Degenerative dementias are complex conditions which are not always well understood by primary health care providers[14]. Therefore, becoming involved in the process will raise awareness of people suffering from MCI and dementias.

PROTOCOL:

The methodology applied in this study has been previously published[5] in work carried out at the University CEU Cardenal Herrera together with community pharmacies in the region of Valencia (Spain) associated with the Spanish Society of Family and Community Pharmacy (SEFAC). This current study was reviewed and approved by the Research Ethics Committee at the Universidad CEU Cardenal Herrera (approval no. CEI11/001) in March 2011. All individuals involved in the study gave their written informed consent to participation in accordance with the Declaration of Helsinki.

1. Selection of factors associated with mild cognitive impairment

1.1. Search for terms related to MCI for use in screening Cochrane Systematic Reviews (e.g., cognitive impairment, dementia, risk factors, etc.).

1.2. Search for terms for which there is some evidence of a relationship with cognitive deterioration or dementia published in the PubMed database; these include demographic factors (sex, age, education level, and economic status), social factors (cognitive and social activities), chronic pathologies (cholesterol, depression, hypertension, diabetes, and obesity), and lifestyle behaviors (alcohol consumption, smoking habit, diet, physical activity, and sleep hours).

1.3. Calculate the odds ratio for qualitative variables or Cohen's d effect size for quantitative variables[15]. Select the variables with larger effect sizes for cognitive deterioration or dementia for use in elaborating a questionnaire.

2. **Design of the questionnaires**

2.1. Design a questionnaire to collect information about the selected variables, following the guidelines provided by Nardi[16]. For instance, the variables used in Climent et al.[5] were demographic (age, weight, and height [measured with standardized procedures using calibrated

scales and stadiometers], sex, education level, and employment type), lifestyle (physical exercise, reading, time spent sleeping overnight and during the day, puzzles, games, TV consumption time, and tobacco and alcohol consumption), and chronic pathologies (hypertension, hyperlipidemia, and diabetes). In addition, record the presence or absence of depression, which is frequently associated with cognitive deterioration.

2.2. Design a pharmacotherapy follow-up sheet to report all the drugs consumed by the participants at the time of the interview, as in Climent et al.[5], which used Dader's method[17] to design this sheet.

3. Selection of tests for MCI screening

3.1. Determine all the tests used to screen for MCI that could be administered by primary healthcare workers (e.g., pharmacists). Reject any tests that must be administered by a specialist. Some of the tests that fulfill these conditions are the *Short Portable Mental State Questionnaire* (SPMSQ)[18], *Mini Mental State Examination* (MMSE)[19], *Memory Impairment Screen* (MIS)[20], *Picture Memory Impairment Screen* (PMIS)[21], *Montreal Cognitive Assessment* (MoCA)[22], *Saint Louis University Mental Status* (SLUMS)[23], and *Quick Mild Cognitive Impairment* (Qmci)[24]. An exhaustive review of each MCI test is available in Cullen et al.[25].

3.2. Search for a good estimation of the test sensitivities and specificities in the scientific literature.

3.3. Estimate the time required to administer these tests to healthy individuals.

3.4. Consider the basic patient characteristics required for completion of these tests. For example, a minimum education level may be necessary because many MCI tests are not suitable for illiterate participants. A set of MCI screening tests is usually applied to increase sensitivity; however, the minimum number of tests must be quickly administered by pharmacists if the final selective screening is intended for a large population. Climent et al.[5] assessed MCI using the MMSE and SPMSQ tests, with the latter being suitable for the large number of individuals who lived through the Spanish civil war who are illiterate.

3.4.1. Use a variant of the SPMSQ by Pfeiffer[18] was validated in Spanish by Martínez de la Iglesia[26]. This test has a maximum score of 10 and the cut-off point for establishing cognitive impairment is 3 or more errors (4 or more for illiterate individuals). This test takes between 8 and 10 minutes to complete.

3.4.2. Use a NORMACODERM version of the MMSE validated for Spanish speakers by Blesa[27] by adapting the original version by Folstein[19]. This screening test has a maximum score of 30 and is corrected according to the patients' years of schooling and ages. Participants who score less than or equal to 24 are considered as MCI cases. The MMSE is a measure of general cognitive function and includes orientation to time and place, written and spoken language, attention span, calculation, and memory. It was administered to all the participants in this study because it is a

very short test which takes only around 5 minutes to complete.

4. Subject recruitment

4.1. Find pharmacists willing to recruit non-institutionalized people to form the study population. The mentioned study by Climent et al[5] included people aged 65 years or more who went regularly to the pharmacy and who agreed to participate in this study. Exclude patients with any difficulty in performing these evaluation tests (e.g., because of blindness, deafness, etc.) or who were already being treated for dementia.

4.2. Provide the participating pharmacists with informed consent forms, which must be completed by every individual taking part in the study. This consent form specifies the title of the research, the objectives of the project, a comprehensible explanation of all the procedures that the participant would take part in, the absence of specific risks, the confidentiality of all the collected data, and the right to withdraw from the study for any reason at any time.

4.3. Train the pharmacists to administer structured personal interviews to the participants, which should last approximately half an hour per person. Collect data for 1 year and send all the forms to the researchers responsible for data protection in the study. Subsequently follow-up with the patients for 3 months.

4.4. Instruct the pharmacists how to identify a probable MCI case using MCI tests. Based on Climent et al.[5] we used SPMSQ scores of 4 or more points (for illiterate participants) or 3 or more points for the other participants and scores of 24 points or less were used in the corrected MMSE test.

4.5. Instruct pharmacists how to refer MCI cases to a medical specialist (a neurologist) for their clinical diagnosis—the last step in the flow chart used in this research study (**Figure 1**).

[Place Figure 1 here]

5. Pharmacist researcher training

5.1. Contact specialists to organize sessions for training the participating pharmacists in basic knowledge related to cognitive impairment and in managing its screening tools, for instance, the SPMSQ and MMSE.

5.2. Ensure that the participating pharmacists are aware of the procedures, data collection protocol, and all the possible issues related to data protection. Inform them that the project was approved by a Research Ethics Committee and of the importance of the consent form according to the Declaration of Helsinki.

NOTE: To perform the study described by Climent et al[5], workshops were held at the Official College of Pharmacists and the Cardenal-Herrera CEU University (UCH-CEU), and covered the

following: MCI and dementia; diagnostic approaches to MCI and management of the SPMQP and MMSE (taught by the Neurology Service at La Plana Hospital in Castellón); project presentation and explanation of the methodology by senior community pharmacist researchers; and health education and cognitive training by researchers from the Department of Pharmacy at the UCH-CEU University.

6. Study design

6.1. Calculate a sample size to assess the feasibility of the project. Because this was an observational study, a larger sample will produce more effective tools. There are two ways to determine sample size: one is based on the estimation of the prevalence and the other is more precise, taking into account effect sizes.

6.1.1. Calculate an accurate estimation of the prevalence of the condition in the population

$$□^2_{(1-\alpha/2)} \frac{□_0(1-□_0)}{□□□□□^2}$$

where α is the significance level, $□_0$ is the initial estimation and □□□□□ is the maximum error expected with a $100(1-\alpha)\%$ confidence.

6.1.2. According to the effect sizes found in the literature for each factor, use tools like the *pwr* package in R to estimate how much power is required to detect differences[15,28].

NOTE: For instance, in our study[5] we designed the first proposal with an error of 3% at 95% confidence and an initial estimation of the prevalence of MCI at 15% in the population aged 65 years or older, resulting in an estimated required sample size of 541 individuals.

7. Interdisciplinary communication network, pharmacists, primary healthcare physicians, and specialists

7.1. Design letters to communicate information about the project to the healthcare centers involved.

7.2. Explain to participating pharmacists how to inform their assigned physicians about the results of the screenings through a letter to the primary healthcare center.

7.3. Send written communications to the medical coordinators of the healthcare centers related to the participating pharmacies and to the Neurology Services of the hospitals to which they are assigned.

7.4. Contact participating neurologists to find out each patient's definitive diagnosis obtained via specific tests undertaken by specialized healthcare providers. Before this, primary healthcare providers should carry out the following protocol, as summarized by the clinical guidelines (**Figure 2**).

[Place Figure 2 here]

8. Statistical analysis and preprocessing

NOTE: Before applying machine-learning techniques a preparatory step is required to transform the original data into a new data set according to the final study objective and the procedures to be applied. For this transformation, several things should be considered, including the characteristics of the algorithms. This is because some of them are sensitive to a lack of variability or sharing of information across columns, although the algorithms used to generate decision trees are particularly robust against these problems. This initial phase aims to categorize qualitative variables and gather values with enough cases for each variable. For efficient screening it is important to choose variables whose acquisition is proven to be easy and accurate. Participants are selected by a short interview in which the algorithms used were constrained to a white-box model, making it easy to check the criteria used to decide if the individual should take the test. We suggest using the *rpart*[29] package in R software for these algorithms, and implementing recursive partitioning.

8.1. Collect all the forms from the participating pharmacies and convert them into a table in which every column is a variable and every participating individual is a row.

8.2. Assign an identification number to each participant. Save the identification number and contact information in a different document so that it is not used by the machine-learning algorithm.

8.3. Generate variables to classify whether each drug the patient takes corresponds to second or third ATC[30] (Anatomical Therapeutic Chemical) level codes, according to the active principal ingredients on the pharmacotherapy follow-up sheet.

8.4. Perform an initial descriptive analysis.

8.4.1. For every ordinal variable, choose an adequate contrast for the variable. For categorical variables, select the value considered as the baseline.

8.4.2. For categorical variables, calculate a univariate logistic regression with a response variable for screening for MCI. Analyze the outcome of the regression with a contingency table, the *p*-value, sample odds ratio, and the 95% confidence interval of the odds ratio.

8.4.3. For quantitative variables, calculate the mean, standard deviation, coefficient of logistic regression, and the 95% confidence interval of their coefficients.

8.5. Reject variables with missing (unavailable) values, considering these variables difficult to accurately collect.

8.6. Select only variables for which there is at least one statistically significant category

($\alpha < 0.01$) according to the logistic regression analysis. The outcome of this step produces a reduced data set compared to the initial one.

9. **Algorithms to create a decision tree**

NOTE: Machine-learning algorithms must be properly parameterized to predict which individuals are likely to have a positive MCI test result. One of the main problems while screening for a condition is that the original data is expected to be imbalanced (i.e., few positive cases compared to the negative ones). To get models with balanced data we used a technique called down-sampling, or random sampling, to equalize the frequency with that of the lowest frequency class[31]. Efficient screening also requires reducing the number of false negatives as much as possible (i.e., increasing the sensitivity of the selection of participants suffering from MCI). One of the techniques used to achieve a greater sensitivity is the introduction of penalties in the calculation of Gini's impurity index (i.e., the index used by the algorithm to select the best split for the decision tree)[32].

9.1. Generate a training and test data set with 80% and 20% of the whole data set, respectively using the *createDataPartition* function in the caret library[33].

9.2. Apply the algorithms used to generate decision trees to the training data set. Use the *train again* function in the caret library[33]. The following steps are different parameters of the function; for instance, the tree used in this paper was generated with *rpart*[29] (*method="rpart"),* but other algorithms are available.

9.2.1. Select the 'down sampling' sampling method and introduce the *sampling = "down*" parameter into the caret.

9.2.2. Set the prior probabilities for both classes.

9.2.3. Provide a loss matrix with the Gini's impurity index penalties applied in order to focus on the increasing sensitivity.

9.2.4. For every parameter in the algorithm, choose an appropriate grid of values.

9.2.5. Use a cross-validation estimation of the receiver operating curve (ROC) values to select the best models within the parameter grid.

9.3. Calculate a confusion matrix and the area under the ROC curve (AUC) for the test set prediction to assess the true performance of the model.

REPRESENTATIVE RESULTS:

The participating pharmacies gathered data from 728 users and collected demographic variables in addition to the drugs prescribed to the participants. A univariate logistic regression was performed for all the variables[34]; the error bar graphs shown in **Figure 3** and **Figure 4** are

convenient graphical representations of the confidence interval of the odds ratio (for qualitative variables) and the confidence interval of the coefficient of the logistic regression (for quantitative variables). Variables with *p*-values exceeding 0.01 (sex, age, education level, reading habit, time spent sleeping, depression, and memory complaints) were selected and used to generate a white-box model based on a decision tree. This decision tree was generated using a training data set comprising 583 individuals as an input and was validated with a test set of a cohort of 145 participants.

After using the caret[33] library in R, the resultant tree assigned a probability of suffering MCI to each individual depending on their final node in the tree (depicted in **Figure 5**) as well as their answers to a few questions. To evaluate the forecasting capability of these probabilities, a ROC analysis of the test set was performed (**Figure 6**); its AUC was 0.763 and its 95% confidence interval was (0.6624, 0.8632). In addition to the probabilities, the tree shown in **Figure 5** also used very simple questions about how long the person sleeps and how often they read, to recommend (with a sensitivity of 0.76 and specificity of 0.70) whether patients should take the MCI tests.

Using this decision tree and short interview to select users at risk of MCI we were able to significantly reduce the number of patients requiring MCI tests (administration is quite time-consuming). This reduction can be estimated by using data in the test set and interpreting the confusion matrix of the observed and predicted classes shown in **Table 1**. In this work, 55 out of 145 participants in the test set were identified by the decision tree for further MCI testing, (representing a reduction of 62% of users taking the tests) while also selecting most of the individuals (19 out of 25) who were positive for MCI.

FIGURE AND TABLE LEGENDS:

Figure 1. Flowchart of the research study and the proposed selective screening. The left side represents the initial study whose data were analyzed with machine-learning techniques to propose the selective screening for early detection of MCI shown in the right panel. This figure was modified from Climent[34].

Figure 2. Protocol for primary healthcare action. An example of primary healthcare actions that should be considered for early MCI detection before the patient is referred for a medical diagnosis by specialists.

Figure 3. Example of the variables selected during preprocessing. A 99% confidence interval of the odds ratio was calculated and is represented as an error bar. The base value for the logistic regression is indicated below the name of the variable at the top of every panel. For every value of the variable, an error bar represents the confidence interval of the odds ratio of taking that value versus taking the base value. Because the variables used to generate the tree were selected, the confidence intervals do not include the value 0 for some values as these showed significant differences. The scale of the vertical axis is logarithmic to help in comparisons across groups.

Figure 4. Example of non-selected variables during preprocessing. A 99% Confidence Interval of

the odds ratio was calculated and is represented with an error bar. The base value for the logistic regression is indicated below the name of the variable at the top of every panel. For every value of the variable, an error bar represents the confidence interval of the odds ratio of taking that value versus taking the base value. In contrast with the previous figure, all the confidence intervals of the selected variables include the value 0, since no significant differences were found to be included to generate the tree. The scale of the vertical axis is logarithmic to help comparison across groups.

Figure 5. Proposed partition tree for selection of pharmacy users. The following tree shows the selection algorithm for MCI tests for individuals aged over 65 years. The text at the top of the box corresponds to the recommendation of taking the MCI screening tests, the two numbers below are the estimated probability of a negative or positive MCI testing outcome, respectively. The value at the bottom of the box is the percentage of individuals with these characteristics in the training set. The warmer the color of the box, the more likely the MCI tests was positive. The top node corresponds to the question about whether the participant has a memory complaint. If the individual does not have a memory complaint, the tree leads to the left branch and the ensuing questions ask about the individual's sex; patients with a memory complaint are asked about the amount of time they sleep per day. This figure was modified from Climent[34].

Figure 6. Receiver operating curves for the partition tree and sensitivity and specificity of the final decision in the test set. The graph represents the ROC curve of the probabilities assigned by the partition tree algorithm in the test set. The red surface corresponds to the AUC and the blue point on the curve shows the sensitivity and specificity of the final recommendation made by the tree.

Table 1. Confusion matrix. Confusion matrix of the predicted and observed values in the test set which were used to validate the proposed model.

DISCUSSION:

After searching for terms associated with MCI in Cochrane studies in the PubMed database, a specific questionnaire was created for this study that used the most evident variables with a proven association with MCI. Demographic, lifestyle, and social factors, as well as the patient's pharmacotherapy and some relevant pathologies were also recorded. Additionally, the SPMSQ and MMSE MCI tests were also selected. Importantly, the SPMSQ was not affected by participants' level of schooling. Pharmacists were trained to administer this study and communication with primary and specialized care was assured via letters informing them of this work. Only specialized healthcare providers could definitively make a diagnosis if MCI was suspected as a result of these tests.

In conclusion, in this study we screened for MCI among a population with a low prevalence of the condition (17%). We designed a set of selection criteria for use with machine-learning techniques, which increased the percentage of MCI positives up to more than 30% among the selected users. Consequently, these tools help increase the screening efficiency and substantially reduce the cost of mass screening among the population group selected by the decision tree.

A limitation of this method is that the decision tree may become invalid in this specific cohort as the population changes and thus, will likely require periodic updates. For instance, many individuals in this population were illiterate, but the number of illiterate individuals aged over 65 years will decrease in the future. These demographic changes will affect the variables related to reading and will require future recalibration of the decision tree.

Remarkably, this data-driven model provided information about the most important variables (from among hundreds) in the construction of a concise yet informative and efficient model. Constructing a decision tree provides insight into the best variables to focus on and is both a cost-effective way to help select people for whom further MCI testing is recommended and furthers our knowledge of these populations in this context.

To increase the future percentage detection rate of MCI, we will require new cost-effective techniques that can assure increased effectiveness. This protocol is time-consuming and is difficult for pharmacists to integrate into their daily work. Thus, other tests such as the MoCA[22] or SLUMS[23] (both with adequate sensitivity and specificity) could be considered for fast the detection of MCI in the future.

A systematic evaluation of the trade-off between specificity and test duration should improve the effectiveness of the set of MCI tests used for screening. Moreover, relevant quantitative variables included in the study should have a wide range so that an efficient cut-off can be selected for them; a narrow range would exclude a large portion of the population from early detection. For instance, the age variable (which is always considered an important criteria in MCI diagnoses) was not considered relevant in this decision tree because the recruitment criteria (age over 65 years) was too conservative; inclusion of younger individuals in a future study would allow the optimal age for starting MCI screening to be calculated.

ACKNOWLEDGMENTS:
This work was made possible by the support of the Know Alzheimer Foundation and help from the multimedia production service at the Universidad CEU Cardenal Herrera, especially Enrique Giner. We would like to recognize the work of all the participating pharmacies (SEFAC), and the collaborating doctors from the Society of Primary Care Doctors (SEMERGEN) and Neurology Society (SVN) who helped with the MCI diagnoses, especially Vicente Gassull, Rafael Sánchez, and Jordi Pérez. Finally, we thank all those who agreed to take part in this study.

DISCLOSURES:
The authors have nothing to disclose.

REFERENCES:
1. Prince, M., Comas-Herrera, A., Knapp, M., Guerchet, M., Karagiannidou, M. *World Alzheimer report 2016: improving healthcare for people living with dementia: coverage, quality and costs now and in the future*. (Alzheimer's Disease International (ADI), .: London, 2016).at <https://scholar.googleusercontent.com/scholar.bib?q=info:mEGpcpLHEIMJ:scholar.google.co

m/&output=citation&scisig=AAGBfm0AAAAAWp5n1MqJdSh73Ri_wiiFuiIjNddB-KMa&scisf=4&ct=citation&cd=-1&hl=en> (2016).
2. Niu, H., Álvarez-Álvarez, I., Guillén-Grima, F., Aguinaga-Ontoso, I. Prevalence and incidence of Alzheimer's disease in Europe: A meta-analysis. *Neurología (English Edition)* . **32** (8), 523–532 (2017).
3. Petersen, R. C. et al. Mild Cognitive Impairment: Clinical Characterization and Outcome. *Archives of Neurology*. **56** (3), 303 (1999).
4. Petersen, R. C. et al. Practice guideline update summary: Mild cognitive impairment: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. *Neurology*. **90** (3), 126–135 (2018).
5. Climent, M. T., Vilaplana, A. M., Arnedo, A., Vilar, J., Moreno-Royo, L. Life styles associated cognitive impairment. Study from the community pharmacy [Estilos de vida asociados a deterioro cognitivo. Estudio preliminar desde la farmacia comunitaria]. *Revista de Investigacion Clinica*. **65** (6), 500–509 (2013).
6. Faubel, R. et al. Usual sleep duration and cognitive function in older adults in Spain. *Journal of Sleep Research*. **18** (4), 427–435 (2009).
7. Benito-León, J., Louis, E. D., Bermejo-Pareja, F. Cognitive decline in short and long sleepers: A prospective population-based study (NEDICES). *Journal of Psychiatric Research*. **47** (12), 1998–2003 (2013).
8. Ramos, A. R. et al. Association between Sleep Duration and the Mini-Mental Score: The Northern Manhattan Study. *Journal of Clinical Sleep Medicine*. **9** (7), 669 (2013).
9. Gabelle, A. et al. Excessive Sleepiness and Longer Nighttime in Bed Increase the Risk of Cognitive Decline in Frail Elderly Subjects: The MAPT-Sleep Study. *Frontiers in Aging Neuroscience*. **9**, 312 (2017).
10. Fernández, M. et al. [Prevalence of dementia in the elderly aged above 65 in a district in the Basque Country]. *Revista de Neurologia*. **46** (2), 89–96 (2008).
11. Velert Vila, J. et al. Suitability of the use of benzodiazepines prescribed by the pharmacist in the elderly. A doctor-pharmacist collaboration study. *Atención Primaria*. **44** (7), 402–410 (2012).
12. Ranstam, J. et al. Impaired cognitive function in elderly men exposed to benzodiazepines or other anziolytics. *The European Journal of Public Health*. **7** (2), 149–152 (1997).
13. Airagnes, G., Pelissolo, A., Lavallée, M., Flament, M., Limosin, F. Benzodiazepine Misuse in the Elderly: Risk Factors, Consequences, and Management. *Current Psychiatry Reports*. **18** (10), 89 (2016).
14. Alacreu, M. et al. Importance of Increasing Modifiable Risk Factors Knowledge on Alzheimer's Disease Among Community Pharmacists and General Practitioners in Spain. *Frontiers in Pharmacology*. **10**, 860 (2019).
15. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. (Routledge: 2013).
16. Nardi, P. M. *Doing Survey Research: A Guide To Quantitative Methods*. (Routledge: 2018).
17. Sabater Hernández, D., Silva Castro, M. M., Faus Dáder, M. J. *Método Dáder: guía de seguimiento farmacoterapéutico*. (Grupo de Investigación en Atención Farmacéutica (GIAF): <Grupo de Investigación en Atención Farmacéutica (CTS-131). Universidad de Granada.> (2017).
18. Pfeiffer, E. A Short Portable Mental Status Questionnaire for the Assessment of Organic Brain Deficit in Elderly Patients†. *Journal of the American Geriatrics Society*. **23** (10), 433–441

(1975).
19. Folstein, M. F., Folstein, S. E., McHugh, P. R. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. **12** (3), 189–98 (1975).
20. Buschke, H. et al. Screening for dementia with the memory impairment screen. *Neurology*. **52** (2), 231–238 (1999).
21. Verghese, J. et al. Picture-based memory impairment screen for dementia. *Journal of the American Geriatrics Society*. **60** (11), 2116–2120 (2012).
22. Nasreddine, Z. S. et al. The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society*. **53** (4), 695–699 (2005).
23. Kaya, D. et al. The Saint Louis University Mental Status Examination Is Better than the Mini-Mental State Examination to Determine the Cognitive Impairment in Turkish Elderly People. *Journal of the American Medical Directors Association*. **17** (4), 370.e11-370.e15 (2016).
24. O’Caoimh, R. et al. Comparison of the quick mild cognitive impairment (Qmci) screen and the SMMSE in screening for mild cognitive impairment. *Age and Ageing*. **41** (5), 624–629 (2012).
25. Cullen, B., ONeill, B., Evans, J. J., Coen, R. F., Lawlor, B. A. A review of screening tests for cognitive impairment. *Journal of Neurology, Neurosurgery & Psychiatry*. **78** (8), 790–799 (2007).
26. Martínez de la Iglesia, J. et al. Adaptación y validación al castellano del cuestionario de Pfeiffer (SPMSQ) para detectar la existencia de deterioro cognitivo en personas mayores e 65 años. *Medicina Clínica*. **117** (4), 129–134 (2001).
27. Blesa, R. et al. Clinical validity of the ‘mini-mental state’ for Spanish speaking communities. *Neuropsychologia*. **39** (11), 1150–1157 (2001).
28. Champely, S., Ekstrom, C., et al. pwr: Basic Functions for Power Analysis <https://cran.r-project.org/web/packages/pwr/index.html> (2018).
29. Therneau, T., Atkinson, B., Ripley, B. rpart: Recursive Partitioning and Regression Trees. R package version 4.1--13 <https://cran.r-project.org/web/packages/rpart> (2018).
30. WHO Collaborating Centre for Drugs Statistics Methodology ATC/DDD Index 2018 <https://www.whocc.no/atc_ddd_index/> (2018).
31. Kuhn, M., Johnson, K. *Applied Predictive Modeling*. **26**, (Springer New York: New York, NY, 2013).
32. Therneau, T., Atkinson, E. An introduction to recursive partitioning using the RPART routines. *r.789695.n4.nabble.com* at <http://r.789695.n4.nabble.com/attachment/3209029/0/zed.pdf> (2019).
33. Kuhn, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*. **28** (5), 1–26 (2008).
34. Climent, M. T., Pardo, J., Muñoz-Almaraz, F. J., Guerrero, M. D., Moreno, L. Decision Tree for Early Detection of Cognitive Impairment by Community Pharmacists. *Frontiers in Pharmacology*. **9** (October), 1–12 (2018).

Figure 1 new

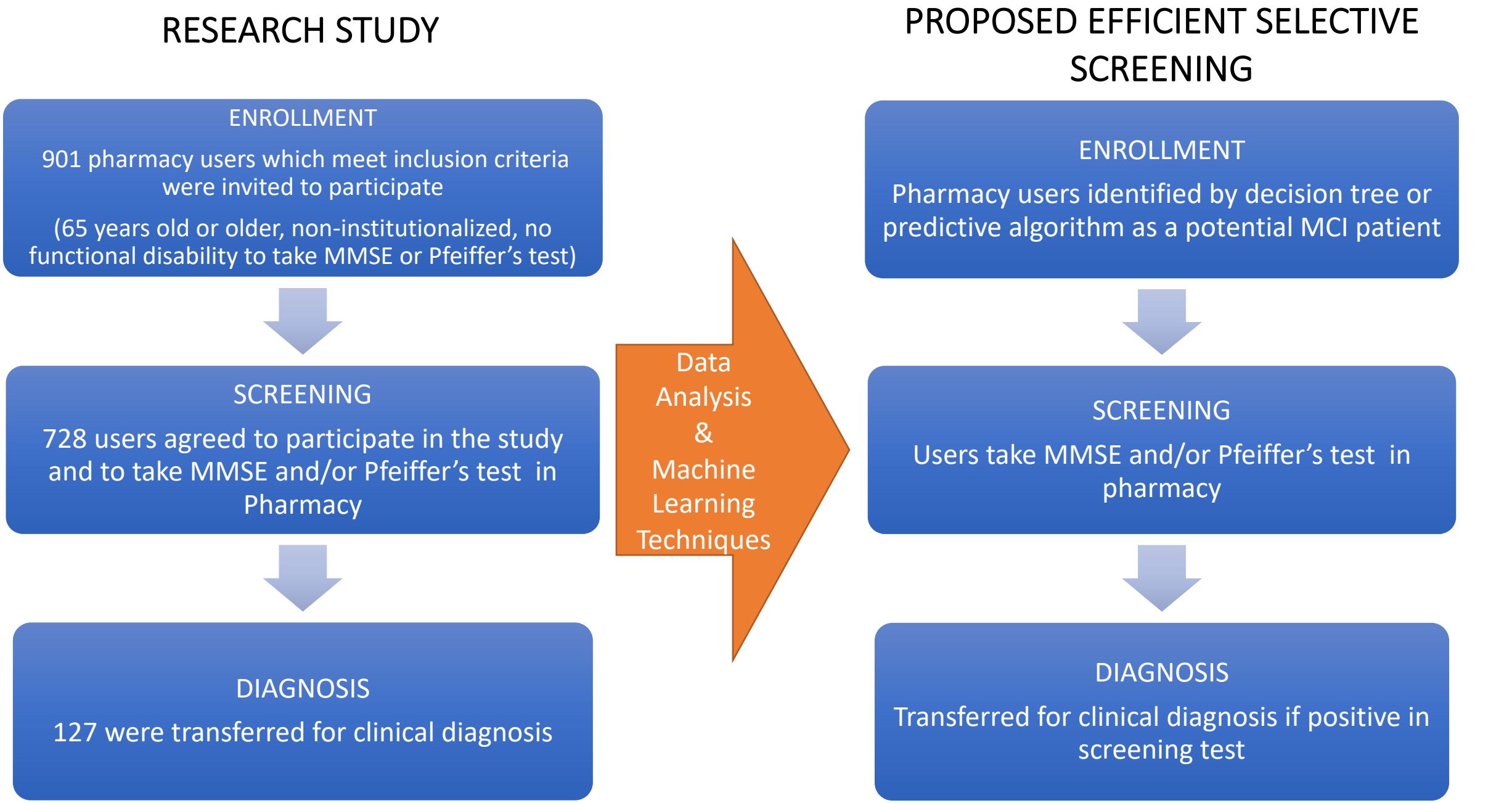

Figure 2

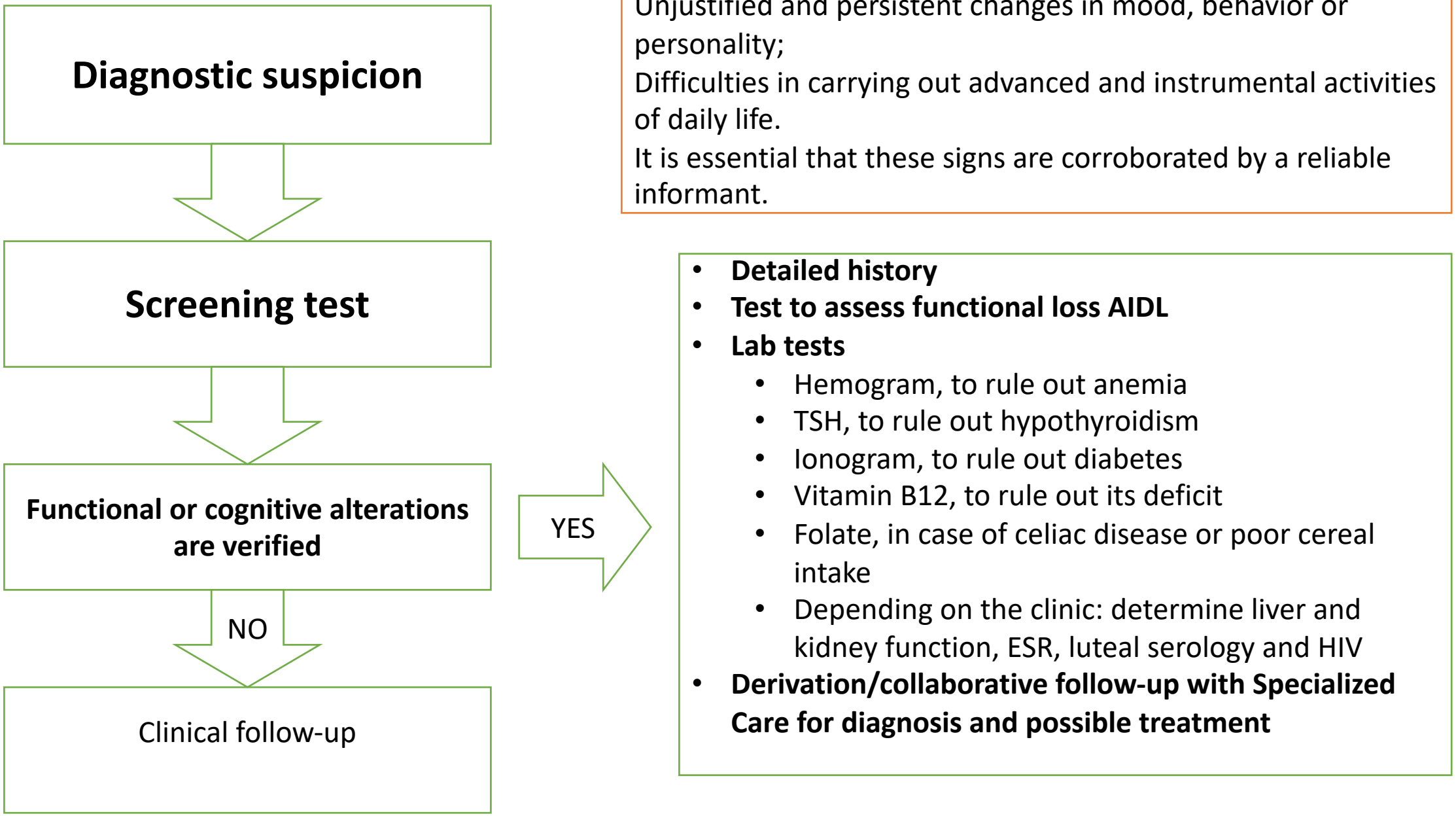

Figure 3 new

Click here to access/download;Figure;Figure3.pdf

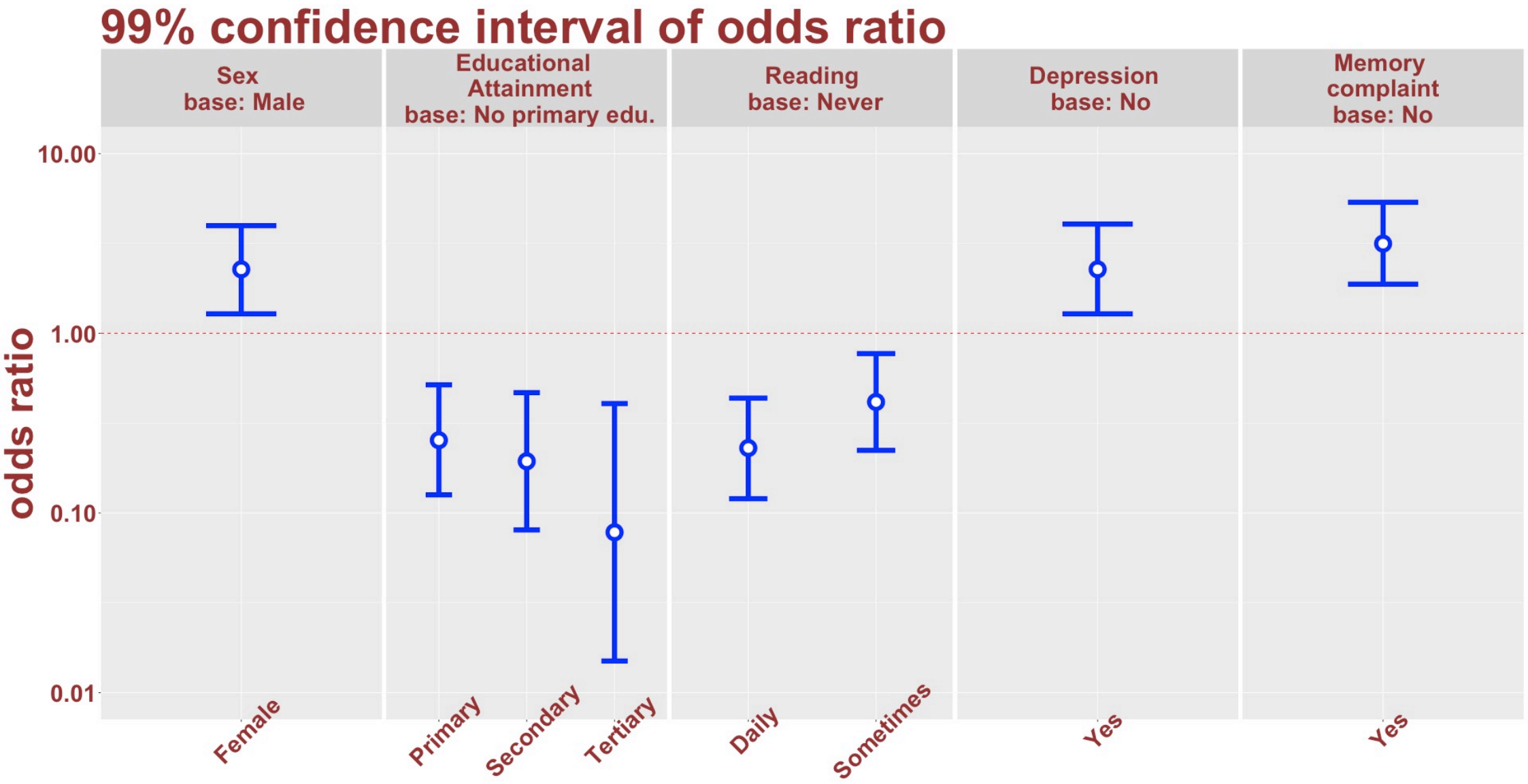

Figure 4 new

Click here to access/download;Figure;Figure4.pdf

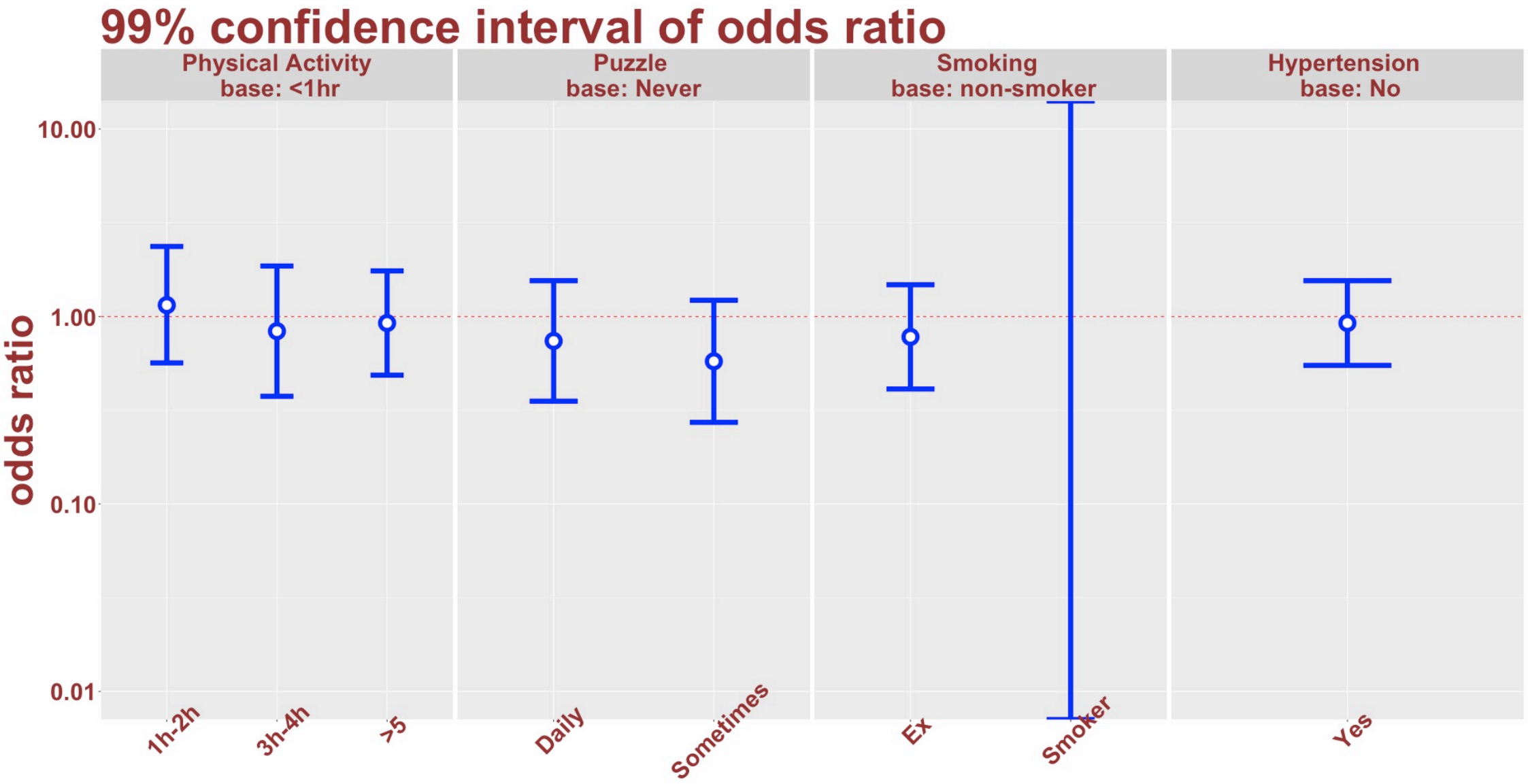

Figure 5 new

Click here to access/download;Figure;Figure5.pdf

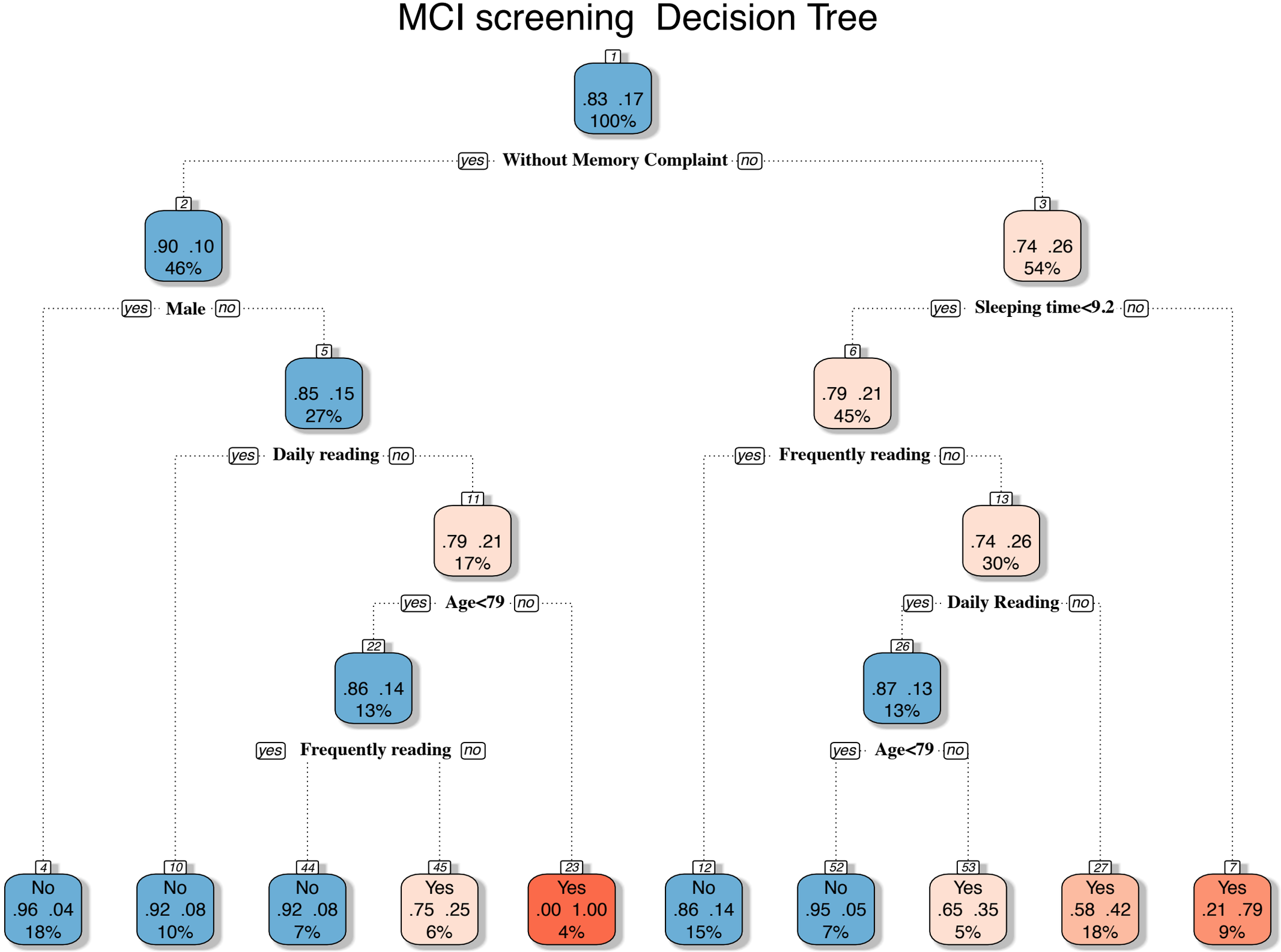

Figure 6

Click here to access/download;Figure;Figure6.pdf

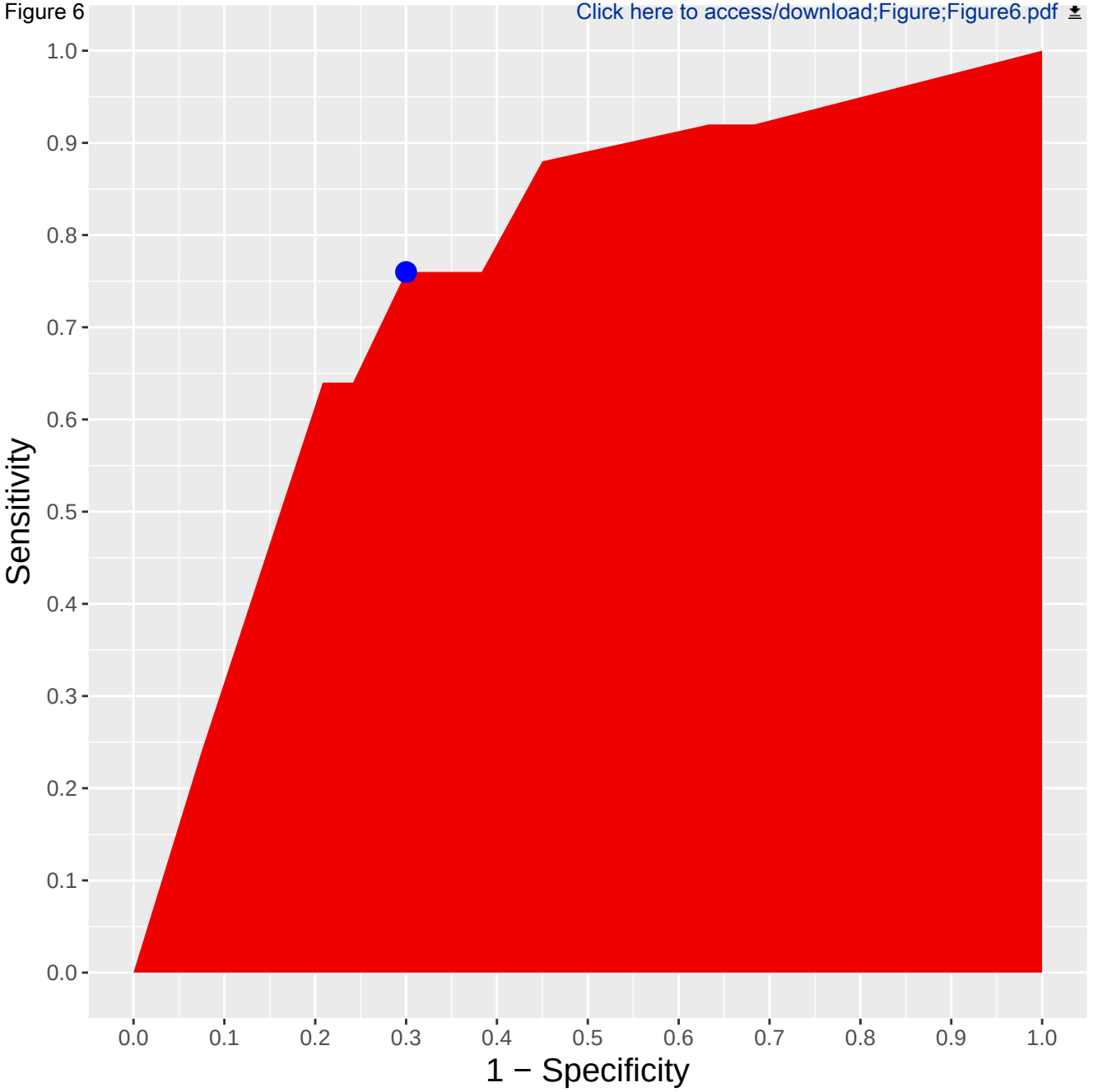

Table 1

		Reference	
		No	Yes
Prediction	No	84	6
	Yes	36	19

Table of Materials

Click here to access/download;Table of Materials;JoVE_Table_of_Materials_review.xlsx

Name of Material/Equipment	Company	Catalog Number	Comments/Description
caret	Max Kuhn		R package
rpart	Terry Therneau, Beth Atkinson, Brian Ripley		R package
SPMSQ in Spanish	Farmaceuticoscomunitarios.org		http://farmaceuticoscomunitarios.org/anexos/vol11_n1/ANEX(
SPMSQ in English	geriatrics.stanford.edu		https://geriatrics.stanford.edu/culturemed/overview/assessme
MMSE in Spanish	Farmaceuticoscomunitarios.org		http://farmaceuticoscomunitarios.org/anexos/vol11_n1/ANEX(
MMSE in English	www.oxfordmedicaleducation.com		http://www.oxfordmedicaleducation.com/geriatrics/mini-men

O1.pdf

ent/assessment_toolkit/spmsq.html

O2.pdf

tal-state-examination-mmse/

1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article:	A machine learning approach to design an efficient selective screening of Mild Cognitive Impairment
Author(s):	Francisco J Muñoz-Almaraz, Maria Teresa Climent, María Dolores Guerrero, Lucrecia Moreno, Juan Pardo

Item 1: The Author elects to have the Materials be made available (as described at http://www.jove.com/publish) via:

☑ Standard Access ☐ Open Access

Item 2: Please select one of the following items:

☑ The Author is **NOT** a United States government employee.

☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. **Defined Terms.** As used in this Article and Video License Agreement, the following terms shall have the following meanings: "**Agreement**" means this Article and Video License Agreement; "**Article**" means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; "**Author**" means the author who is a signatory to this Agreement; "**Collective Work**" means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; "**CRC License**" means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode; "**Derivative Work**" means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; "**Institution**" means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; "**JoVE**" means MyJove Corporation, a Massachusetts corporation and the publisher of The Journal of Visualized Experiments; "**Materials**" means the Article and / or the Video; "**Parties**" means the Author and JoVE; "**Video**" means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. **Background.** The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. **Grant of Rights in Article.** In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and(c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

4. **Retention of Rights in Article.** Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. **Grant of Rights in Video – Standard Access.** This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. **Grant of Rights in Video – Open Access.** This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this **Section 6** is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. **Government Employees.** If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. **Protection of the Work.** The Author(s) authorize JoVE to take steps in the Author(s) name and on their behalf if JoVE believes some third party could be infringing or might infringe the copyright of either the Author's Article and/or Video.

9. **Likeness, Privacy, Personality.** The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

10. **Author Warranties.** The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

11. **JoVE Discretion.** If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole

1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

12. **Indemnification.** The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in JoVE or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

13. **Fees.** To cover the cost incurred for publication, JoVE must receive payment before production and publication of the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

14. **Transfer, Governing Law.** This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement is required per submission.

CORRESPONDING AUTHOR

Name:	Francisco Javier Muñoz-Almaraz
Department:	Matemáticas, física y ciencias de la computación
Institution:	Universidad CEU Cardenal Herrera
Title:	Dr.
Signature:	
Date:	January 3, 2019

Please submit a **signed** and **dated** copy of this license by one of the following three methods:

1. Upload an electronic version on the JoVE submission site
2. Fax the document to +1.866.381.2236
3. Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02140

Dear Dr. Nguyen,

We are pleased to submit a new version of our video and our manuscript. We appreciate all the editorial comments to improve our work. The answer to every comment is in red. We have also attached with this letter the track changes in the manuscript.

Editorial comments:

1. Please employ professional copy-editing as the language is not publication grade. I have gone through the manuscript a bit but there are still sections that are unclear.

We have hired a professional to review the language and we are expecting that now all the ideas in our manuscript are clear enough. We think that some initially obscure expressions are now more clearly explain. It has been thoroughly revised and many sentences changed, but the content is not changed

2. Figure 3/4: Where are the base values indicated in the Figure?

We have written a more appropriate legend of Figure 3 and 4 to explain that is the odds ratio of every value versus the base value.

3. The explanations of the figures in the figure legends needs revision to be more clear.

In addition to Figure 3 and 4, we have modified all the legends with a more explanatory text. In our opinion, these new legends describe better the graphs than the previous ones.

4. Figure 5 seems off. Is the tree flipped? The line for Yes for "Without Memory Complaint" leads to the No box. This goes for every decision in the tree. This is used in the video as well.

The “No” in the box corresponds to a non-final recommendation of taking or not taking MCI test. We have realized that these “partial” recommendations are not properly explained in the video and we have decided to remove them from video and manuscript, leaving only the final recommendation for the final nodes.

5. Please change the numbers of the protocol sections in the video: 2.2 Design of the questionnaires in the video should be 2. Design of the questionnaires. Please do not number the major sections of the manuscript in the video (Introduction, protocol, Representative Results). Please use the numbering from the written manuscript as I have changed it to fit our publication standard.

The sections in the video have been modified according to the new titles of the sections in the document attached with your e-mail.

Please submit a revised high-resolution video here:
https://www.dropbox.com/request/d9KDHGwq4lVvveEFItbL?oref=e

The video has been already uploaded.

Sincerely,

Javier Muñoz PhD

TITLE:

machine learning approach to design an efficient selective screening of mild cognitive impairment

AUTHORS AND AFFILIATIONS:

Francisco J Muñoz-Almaraz[1], María Teresa Climent[2], María Dolores Guerrero[3], Lucrecia Moreno[3], Juan Pardo[1]

[1]Embedded Systems and Artificial Intelligence Group, Universidad CEU Cardenal Herrera, Valencia, Spain

[2]Community Pharmacist, Valencia, Spain

[3]Department of Pharmacy, Universidad CEU Cardenal Herrera, Valencia, Spain

Corresponding Author:

Juan Pardo Albiach

juaparal@uchceu.es

Email Addresses of Co-authors:

FJ Muñoz-Almaraz (malmaraz@uchceu.es)

MT Climent (m.climent.000@micof.es)

MD Guerrero (ma_dolores.guerrero@uchceu.es)

L Moreno (lmoreno@uchceu.es)

KEYWORDS:

Memory complaint, early detection, mild cognitive impairment, sleep duration, community pharmacist, risk factors, decision trees, statistical learning

SUMMARY:

This methodology produces decision trees which target population groups more prone to suffering from mild cognitive impairment that are useful for cost-effective selective screening of the disease.

Formatted: Space After: 12 pt

Deleted: a

Deleted: ¶

Formatted: Spanish

Deleted: ¶

Formatted: Spanish

Formatted: Spanish

Formatted: Spanish

Field Code Changed

Field Code Changed

Formatted: Spanish

Formatted: Spanish

Deleted: ¶

Deleted: ¶

Deleted: that

Deleted: a

Deleted: mild cognitive impairment,

Deleted: which targeting the population groups more prone to suffering from …

ABSTRACT:

Mild cognitive impairment (MCI) is a first sign of dementia among elderly populations and so its early detection is crucial in our aging societies. Common MCI tests are time-consuming meaning that indiscriminate massive screening would not be cost-effective. Here we describe a protocol which uses machine learning techniques to rapidly select candidates for further screening via a question-based MCI test. This minimizes the number of resources required for screening because only patients who are potentially MCI positive are tested further.

This methodology was applied in an initial MCI research study which formed the starting point for the design of a selective screening decision tree. The initial study collected many demographic and lifestyle variables as well as details about patient medications. The *Short Portable Mental Status Questionnaire* (SPMSQ) and the *Mini-Mental State Examination* (MMSE) were used to detect possible cases of MCI. Finally, we used this method to design an efficient process for classifying individuals at risk of MCI. This work also provides insights into lifestyle-related factors associated with MCI which could be leveraged in the prevention and early detection of MCI among elderly populations.

INTRODUCTION:

Population aging is increasing the prevalence of chronic and degenerative diseases, especially degenerative dementias, which are expected to affect more than 131 million people worldwide by 2050[1]. Among all the degenerative dementias, Alzheimer's disease (AD) is the most common with an overall prevalence in Europe of 6.88%[2]. Due to the ever declining independence of AD patients, this group should start receiving support as soon as AD starts to manifest. Therefore, the early detection of prodromal signs of AD, such as mild cognitive impairment (MCI), is essential.

MCI is defined as an intermediate cognitive decline stage corresponding to normal aging and severe deterioration due to dementia[3]. According to estimates by Petersen et al.[4], the prevalence of MCI is 8.4% among people aged 65–69 years and reaches 25.2% for those aged over 80 years. MCI results in individuals experiencing more difficulties than expected in the execution of low-level cognitive skills, especially those related to memory and language, but does not interfere with the activities of daily living.

Screening is not synonymous with diagnosis; the diagnosis of MCI will always be a clinical task whereas screening methods can only inform us that a patient has a higher probability of suffering from this pathology and that there is a well-founded suspicion of MCI that should be confirmed clinically. Hence, primary healthcare workers (doctors, pharmacists, nurses, etc.) could benefit from the availability of simple screening methods (brief cognitive tests) that can be applied in minutes. Ideally, these would objectively identify patients with a high probability of suffering a MCI so that they can then be clinically tested by general or specialized physicians.

Given that the early detection of MCI is becoming an essential task within the context of public health, this work aimed to identify which characteristics are useful in the targeted identification

Deleted: ¶

Deleted: Early detection of …m…ld cognitive impairment (MCI),…is a first sign of dementia among the …lderly populations and so its,…early detection of …s crucial in our aging societies. Common MCI tests are time-consuming and someaning that , implying that an …ndiscriminate massive screening would not be cost-effective. Here we describe In this paper, a protocol which uses based on the application of of machine learning techniques is described …o rapidly carry out a fast selection…of …andidates for further screening via a a question-based to take an …CI test by means of few questions. This Consequently, …inimizes the number of resources required for screening is minimized,…because since only patients who are potentially MCI positives…in MCI … [1]

Deleted: ¶
The …his methodology is …as applied to …n an initial MCI … [2]

Formatted: Font: Italic

Deleted: Portable Mental State Questionnaire

Formatted: Font: Italic

Deleted: *mental Mental State Examination* (MMSE) were employed used for the…o detection…of …ossible cases of MCI. Finally, we used this method to generate …esign an… [3]

Deleted: ¶

Deleted: for

Deleted: Alzheimer's Alzheimer's disease (AD) is the most … [4]

Deleted: ; its…n overall prevalence in Europe is …f 6.88%[2]. … [5]

Deleted: they this group should start get…eceiving support … [6]

Deleted: g ofa … [7]

Deleted: ¶

Deleted: stage between a cognitive decline stage … [8]

Deleted: one deterioration due to dementia[3]. According to … [9]

Deleted: for among people ages …ged 65-…69 years,…and … [10]

Deleted: of age. MCI results in individuals experiencing … [11]

Deleted: mainly related to memory and language, but it … [12]

Deleted: ¶ … [13]

Deleted: to with diagnosis. … The … [14]

Deleted: and screening methods can only inform us that … [15]

Deleted: must

Deleted: care there is a need for all …ealthcare workers … [16]

Deleted: to use rapid and

Deleted: , for …bjectively identification …dentify of … [17]

Deleted: mild cognitive deteriorationMCI so that they can … [18]

Deleted: Consequently, early detection of MCI is becoming … [19]

Deleted: in society….This …his work aims …imed to … [20]

of MCI in screening tests of elderly populations. These groups would then be more thoroughly tested for MCI in tests administered by primary health care providers. This methodology provides a decision tree with the appropriate algorithms for identifying the population groups to target.

Among these characteristics, age is one of the most consistent factors associated with the development of this pathology. Other relevant characteristics are related to demographics or lifestyle[5]. Among the latter, some studies have identified the duration of daytime or nighttime sleep as a risk factor that can lead to the diagnosis of MCI[5–9]. The prolonged consumption of medications such as benzodiazepines, consumed by an estimated 20%–25% of older adults[12,13], can also influence sleep hours and the development of MCI[10,11]. Indeed, prolonged treatments for chronic diseases may be important features useful in the pre-selection of individuals with a high risk of suffering from MCI.

Here we developed data-based models which use automatic learning algorithms, a decision tree, and a predictive tool to increase the efficiency of the methodology for detecting MCI by discriminating which characteristics play an important role in the early detection of MCI. The resultant decision tree presented here was produced using a specific cohort of Spanish patients using community pharmacies. However, this method would also be useful among other populations with different characteristics.

This work was completed in collaboration with primary healthcare and specialized medical doctors. Community pharmacies were ideal for testing this algorithm because they are close to patients, have long opening hours, and are frequently visited and consulted. Degenerative dementias are complex conditions which are not always well understood by primary health care providers[14]. Therefore, becoming involved in the process will raise awareness of people suffering from MCI and dementias.

PROTOCOL:

The methodology applied in this study has been previously published[5] in work carried out at the University CEU Cardenal Herrera together with community pharmacies in the region of Valencia (Spain) associated with the Spanish Society of Family and Community Pharmacy (SEFAC). This current study was reviewed and approved by the Research Ethics Committee at the Universidad CEU Cardenal Herrera (approval no. CEI11/001) in March 2011. All individuals involved in the study gave their written informed consent to participation in accordance with the Declaration of Helsinki.

1. Selection of factors associated with mild cognitive impairment

2.1. Search for terms related to MCI for use in screening Cochrane Systematic Reviews, e.g., cognitive impairment, dementia, risk factors, etc.

2.2. Search for terms for which there is some evidence of a relationship with cognitive deterioration or dementia published in the PubMed database; these include demographic

Deleted: the …lderly populations. These groups would who can then be more thoroughly … [21]

Deleted: suffering from MCI …or and screening this population group with an …CI in tests administered by primary health care providers. The … … [22]

Deleted: produces … decision tree with the appropriate algorithms for to … … [23]

Deleted: ¶

Deleted: one of the most relevant, since it is …ne of the most consistent factors that is … … [24]

Deleted: the this pathologydisease… Other relevant characteristics are related to demographics or related to … [25]

Deleted: se latterlifestyle variables relevant to cognitive decline, some studies have identified the duration of daytime or nighttime daily and/or night duration of … … [26]

Deleted: P

Deleted: can also influence sleep hours and the development of MCI[10,11]. It is …y an estimated that between … [27]

Deleted: general, these

Deleted: could be important features useful in the pre-t … [28]

Deleted: who could be

Deleted: ¶
The application of d

Deleted: , based on automatic learning algorithms, a a … [29]

Deleted: a a predictive tool has been developed …o increase the efficiency of the detection …ethodology for detecting MCI by , … … [30]

Deleted: inging which characteristics play an important role for …n the early detection of MCI. Theis… … [31]

Deleted: by using it with is the outcome of applying this methodology to the pharmacy users to … specific cohort of … [32]

Deleted: the …his method would also presented here is…e useful for … … [33]

Deleted: ¶
All tThis work has been…as completed carried out …n … [34]

Deleted: ¶

Deleted: ***The present …ethodology applied in this study has been previously is illustrated by means of a previous … [35]

Formatted: Superscript

Deleted: [5] in work …arried out at …¶ … [36]

Deleted: ¶

Formatted: Font: Bold

Deleted: MCI

Deleted: ¶ … [37]

Deleted: ¶ … [38]

factors (sex, age, education level, and economic status), social factors (cognitive and social activities), chronic pathologies (cholesterol, depression, hypertension, diabetes, and obesity), and lifestyle behaviors (alcohol consumption, smoking habit, diet, physical activity, and sleep hours).

2.3. Calculate the odds ratio for qualitative variables or Cohen's d effect size for quantitative variables[15]. Select the variables with larger effect sizes for cognitive deterioration or dementia for use in elaborating a questionnaire.

2. Design of the questionnaires

2.1. Design a questionnaire to collect information about the selected variables, following the guidelines provided by Nardi[16]. For instance, the variables used in Climent et al.[5] were demographic (age, weight, and height [measured with standardized procedures using calibrated scales and stadiometers], sex, education level, and employment type), lifestyle (physical exercise, reading, time spent sleeping overnight and during the day, puzzles, games, TV consumption time, and tobacco and alcohol consumption), and chronic pathologies (hypertension, hyperlipidemia, and diabetes). In addition, the presence or absence of depression, which is frequently associated with cognitive deterioration, was also recorded.

2.2 Design a pharmacotherapy follow-up sheet to report all the drugs consumed by the participants at the time of the interview, as in Climent et al.[5] which used Dader's method[17] to design this sheet.

3. Selection of tests for MCI screening

3.1. Determine all the tests used to screen for MCI that could be administered by primary healthcare workers (e.g., pharmacists). Reject any tests which must be administered by a specialist. Some of the tests that fulfill these conditions are the *Short Portable Mental State Questionnaire* (SPMSQ)[18], *Mini Mental State Examination* (MMSE)[19], *Memory Impairment Screen* (MIS)[20], *Picture Memory Impairment Screen* (PMIS)[21], *Montreal Cognitive Assessment* (MoCA)[22], *Saint Louis University Mental Status* (SLUMS)[23], and *Quick Mild Cognitive Impairment* (Qmci)[24]. An exhaustive review of each MCI test is available in Cullen et al.[25].

3.2. Search for a good estimation of the test sensitivities and specificities in the scientific literature.

3.3. Estimate the time required to administer these tests to healthy individuals.

Deleted: condition

Deleted: ¶
Calculate the odds ratio for qualitative variables and …r Cohen's d effect size for quantitative variables[15]. Select the variables with a …arger effect sizes forof…cognitive deterioration or dementia for use in to …laborate… ... [39]

Formatted: Space After: 12 pt, Add space between paragraphs of the same style, No bullets or numbering, Tab stops: Not at 2,54 cm

Formatted: Space After: 12 pt

Formatted: Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 2 + Alignment: Left + Aligned at: 0 cm + Indent at: 1,27 cm, Tab stops: Not at 1,27 cm

Formatted: Space After: 12 pt, Add space between paragraphs of the same style

Deleted: of …he selected associated …ariables, following the guidelines provided by Nardi[16]. For instance, the variables For instanceAccording to Maite [5]Here, the chosenused variables by…n Climent et al.[5] ... [40]

Formatted: Indent: Left: -0,63 cm, Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 1,27 cm + Indent at: 1,9 cm, Tab stops: Not at 2,54 cm

Formatted: Add space between paragraphs of the same style, No bullets or numbering

Deleted: ¶
Design a pharmacotherapy follow-up sheet to reportwhich contains all the For treatment …rugs consumed by the participants at the time of the interview, as in , design a pharmacotherapy follow-up sheet. We followed In Maite [5] Climent et al.[5] Here,…which who had …sed we use the Dader's method[17] to was used for the …esign of thise…pharmacotherapy follow-up … ... [41]

Formatted: Space After: 12 pt, Add space between paragraphs of the same style, No bullets or numbering, Tab stops: Not at 2,54 cm

Formatted: Space After: 12 pt

Formatted ... [42]

Deleted: ¶
Determine Find …ll the testse…used to screening…tests of ... [43]

Formatted ... [44]

Deleted: E

Formatted ... [45]

Deleted: , etc… In Cullen et al.[25],…Aa ... [46]

Deleted: ¶
Search for a good estimation of the test sensitivity ... [47]

Deleted: ¶

3.4. Consider the basic patient characteristics required for completion of these tests. For example, a minimum education level may be necessary because many MCI tests are not suitable for illiterate participants. A set of MCI screening tests is usually applied to increase sensitivity, however, the minimum number of tests must be quickly administered by pharmacists if the final selective screening is intended for a large population. Climent et al.[5] assessed MCI using the MMSE and SPMSQ tests, with the latter being suitable for the large number of individuals who lived through the Spanish civil war who are illiterate.

3.4.1. A variant of the SPMSQ by Pfeiffer[18] was validated in Spanish by Martínez de la Iglesia[26]. This test has a maximum score of 10 and the cut-off point for establishing cognitive impairment is 3 or more errors (4 or more for illiterate individuals). This test takes between 8 and 10 minutes to complete.

3.4.2. The NORMACODERM version of the MMSE was validated for Spanish speakers by Blesa[27] by adapting the original version by Folstein[19]. This screening test has a maximum score of 30 and is corrected according to the patients' years of schooling and ages. Participants who score less than or equal to 24 are considered as MCI cases. The MMSE is a measure of general cognitive function and includes orientation to time and place, written and spoken language, attention span, calculation, and memory. It was administered to all the participants in this study because it is a very short test which takes only around 5 minutes to complete.

4. Subject recruitment

4.1 Find pharmacists willing to recruit non-institutionalized people to form the study population. The mentioned study by Climent et al[5] included people aged 65 years or more who went regularly to the pharmacy and who agreed to participate in this study. We excluded patients with any difficulty in performing these evaluation tests (e.g., because of blindness, deafness, etc.) or who were already being treated for dementia.

4.2 Provide the participating pharmacists with informed consent forms which must be completed by every individual taking part in the study. This consent form specifies the title of the research, the objectives of the project, a comprehensible explanation of all the procedures that the participant would take part in, the absence of specific risks, the confidentiality of all the collected data, and the right to withdraw from the study for any reason at any time.

4.3 Train the pharmacists to administer structured personal interviews to the participants, which should last approximately half an hour per person. Collect data for 1 year and send all the forms to the researchers responsible for data protection in the study. Subsequently follow-up the patients for 3 months.

4.4 Instruct the pharmacists how to identify a probable MCI case using MCI tests. Based on Climent et al.[5] we used SPMSQ scores of 4 or more points (for illiterate participants) or 3 or more points for the other participants and scores of 24 points or less were used in the

Deleted: ¶
Take into account…onsider the basic patient characteristics ... [48]

Formatted ... [49]

Deleted: illiterate patients because a …arge number of individuals who lived through the Spanish civil war who are illiterateone of which is adequate for illiterates due to the large number of illiterates who lived through the Spanish civil war… Namely,
Namely, ... [50]

Deleted: ¶
NOTE: For instance, for an elderly Spanish population with a large number of illiterates who lived through the Spanish civil war, we assessed MCI by means of two different tests widely used in memory clinic, one of which is adequate for illiterates. ¶
¶ ... [51]

Formatted: Font: Italic

Deleted:) by Pfeiffer[18] was validated in Spanish by Martínez de la Iglesia[26]. This test It …as a maximum score of 10 with …nd the a …ut-off point for establishing cognitive impairment is with … or more errors (4 or more for illiterate ... [52]

Deleted: ¶
The MMSE …ORMACODERM version of the Alternatively, use ... [53]

Formatted: Font: Italic

Deleted: of by Folstein[19]. This screening test has a maximum score of 30 and is with …orrections…d according ... [54]

Formatted: Space After: 12 pt, Add space between paragraphs of the same style

Formatted: Indent: First line: 0 cm, Add space between paragraphs of the same style

Deleted: ¶

Formatted ... [55]

Formatted: Add space between paragraphs of the same style, No bullets or numbering

Deleted: ¶
¶

Formatted: Font: (Asian) Calibri

Deleted: from the study population. The mentioned study by Climent et al The mentioned study by Climemt included ... [56]

Formatted ... [57]

Deleted: Find ¶
Provide the participating pharmacists with informed consent ... [58]

Deleted: on to administer structured personal interviews with to the participants, which should last approximately half ... [59]

Formatted: Font: (Asian) Times New Roman

Formatted ... [60]

Deleted: ¶
¶ ... [61]

corrected MMSE test.

4.5 Instruct pharmacists how to refer MCI cases to a medical specialist (a neurologist) for their clinical diagnosis—the last step in the flow chart used in this research study (**Figure 1**).

[Place Figure 1 here]

5. Pharmacist researcher training

5.1. Contact specialists to organize sessions for training the participating pharmacists in basic knowledge related to cognitive impairment and in managing its screening tools, for instance, the SPMSQ and MMSE.

5.2. Ensure that the participating pharmacists are aware of the procedures, data collection protocol, and all the possible issues related to data protection. Inform them that the project was approved by a Research Ethics Committee and of the importance of the consent form according to the Declaration of Helsinki.

NOTE: To perform the study described by Climent et al[5], workshops were held at the Official College of Pharmacists and the Cardenal-Herrera CEU University (UCH-CEU), and covered the following: MCI and dementia; diagnostic approaches to MCI and management of the SPMQP and MMSE (taught by the Neurology Service at La Plana Hospital in Castellón); project presentation and explanation of the methodology by senior community pharmacist researchers; and health education and cognitive training by researchers from the Department of Pharmacy at the UCH-CEU University.

6. Study design

6.1 Calculate a sample size to assess the feasibility of the project. Because this was an observational study, a larger sample will produce more effective tools. There are two ways to estimate this; one is more simple and the other is more precise.

6.1.1. Calculate an accurate estimation of the prevalence of the condition in the population

$$z^2_{(1-\alpha/2)} \frac{p_0(1-p_0)}{error^2}$$

where α is the significance level, p_0 is the initial estimation and *error* is the maximum error expected with a $100(1-\alpha)\%$ confidence.

6.1.2. According to the effect sizes found in the literature for each factor, use tools like the *pwr* package in R to estimate how much power is required to detect differences[15,28].

NOTE: For instance, in our study[5] we designed the first proposal with an error of 3% at 95% confidence and an initial estimation of the prevalence of MCI at 15% in the population aged 65 years or older, resulting in an estimated required sample size of 541 individuals.

Deleted: ¶

Formatted: Font: (Asian) Times New Roman, Font colour: Text 1

Deleted: ¶
¶

Formatted: Font: (Asian) Times New Roman, Font colour: Text 1

Deleted: Specify to the pharmacists on…how to refer MCI cases to a the …edical specialist (a neurologist) for their clinical diagnosis—, which is …he last step in of …he flow chart used in flow of the … ... [62]

Deleted: ¶

Deleted: ¶

Deleted: ¶
¶ ... [63]

Formatted: Font: (Asian) Times New Roman, Font colour: Text 1

Deleted: With that purpose, contact specialists from the different fields related to the study. ¶
¶
Have these training sessions provide enough information to perform all the procedures described in the previous step. ¶
¶
¶

Formatted ... [64]

Deleted: nting pharmacists are aware of the procedures, the data collect ... [65]

Formatted: Font colour: Text 1

Deleted: about data protection. Inform them that the project has been… ... [66]

Deleted: ¶
NOTE: To perform the study described by Climent et alAs described elsewhereTo perform the study··[5],Here,…workshops were held at the Official College of Pharmacists and at in …he Cardenal-Herrera CEU University ... [67]

Deleted: ¶

Formatted ... [68]

Formatted ... [69]

Deleted: ¶ ... [70]

Deleted: ¶

Formatted: Space After: 12 pt

Deleted: t

Deleted: ¶ ... [71]

Formatted: Font: Italic

Deleted: enough required power ... [72]

Deleted: ¶ ... [73]

7. Interdisciplinary communication network, pharmacists, primary healthcare physicians, and specialists

7.1. Design letters to communicate information about the project to the healthcare centers involved.

7.2. Explain to participating pharmacists how to inform their assigned physicians about the results of the screenings through a letter to the primary healthcare center.

7.3. Send written communications to the medical coordinators of the healthcare centers related to the participating pharmacies and to the Neurology Services of the hospitals to which they are assigned.

7.4. Contact participating neurologists to find out each patient's definitive diagnosis obtained via specific tests undertaken by specialized healthcare providers. Before this, primary healthcare providers should carry out the following protocol, as summarized by the clinical guidelines (**Figure 2**).

[Place Figure 2 here]

8. Statistical analysis and preprocessing

NOTE: Before applying machine-learning techniques a preparatory step is required to transform the original data into a new data set according to the final study objective and the procedures to be applied. For this transformation, several things should be considered, including the characteristics of the algorithms. This is because some of them are sensitive to a lack of variability or sharing of information across columns, although the algorithms used to generate decision trees are particularly robust against these problems. This initial phase aims to categorize qualitative variables and gather values with enough cases for each variable. For efficient screening it is important to choose variables whose acquisition is proven to be easy and accurate. Participants are selected by a short interview in which the algorithms used were constrained to a white-box model, making it easy to check the criteria used to decide if the individual should take the test. We suggest using the *rpart*[29] package in R software for these algorithms, and implementing recursive partitioning.

8.1. Collect all the forms from the participating pharmacies and convert them into a table in which every column is a variable and every participating individual is a row.

8.2. Assign an identification number to each participant. Save the identification number and contact information in a different document so that it is not used by the machine-learning algorithm.

8.3. Generate variables to classify whether each drug the patient takes corresponds to second or third ATC[30] (Anatomical Therapeutic Chemical) level codes, according to the active principal ingredients on the pharmacotherapy follow-up sheet.

Deleted: ¶

Deleted: ¶
Design communication …etters to communicate for the healthcare center …nforming… ... [74]

Deleted: ¶
Explain to participant …articipating pharmacists how to inform their assigned physicians about of … ... [75]

Deleted: ¶
Send written communicationsletters ... [76]

Deleted: ¶
Contact participant …articipating neurologists to find out know about the…ach patient's definitive diagnosis obtained via , through …pecific tests perform …ndertaken by for specialized healthcare providers. Before this, PreviouslyHere, PreviouslyHere, the primary healthcare providers doctorfamily doctor w…hould carry outas expected to perform …he following protocol, as summarizedindicated…by the clinical guidelines (**Figure 2**). ... [77]

Formatted: Space After: 12 pt

Deleted: ¶

Deleted: ¶
NOTE: A preparatory step …b…fore applying machine machine-learning techniques a preparatory step is required to transform the original data into a new data set according to the final study objective and the procedures to be applied. For this transformation, several things should be considered, including . Consider …he characteristics of the algorithms. This is because ; …ome of them algorithms …re sensitive to the a lack of variability or sharing of information across columns. … although tT…e algorithms used to generate decision trees are particularly robust against to …hese problems. and t…he…s focus in this …nitial phase aims is about how to categorize qualitative variables and,…gathering…values with enough cases for each valuevariable. For efficient screening it is important On the other hand,to choose variables whose acquisition has been shownis proven to be easy and accurate and effortless for an an efficient screening… The selection of p…articipants are selected by is done with … short interview in which , and the algorithms to be …sed are …ere constrained to a white- ... [78]

Formatted: Font: Italic

Deleted: ¶
Collect all the forms from the participant …articipating pharmacies and convert themit…into a table where …n which every column is a variable and every participan…ing ... [79]

Deleted: ¶
Assign an identification number to each participant user… Save the identification number and contact information in a different document a…o that it is not employed …sed by the ... [80]

Deleted: ¶
Generate variables to classify whether each drug or not …he patient is …aking…s corresponds a drug according …o 2nd second and …r 3rd …hird level of …TC[30] (Anatomical ... [81]

8.4. Perform an initial descriptive analysis.

8.4.1. For every ordinal variable, choose an adequate contrast for the variable. For categorical variables, select the value considered as the baseline.

8.4.2. For categorical variables, calculate a univariate logistic regression with a response variable for screening for MCI. Analyze the outcome of the regression with a contingency table, the *p*-value, sample odds ratio, and the 95% confidence interval of the odds ratio.

8.4.3. For quantitative variables, calculate the mean, standard deviation, coefficient of logistic regression, and the 95% confidence interval of their coefficients.

8.5. Reject variables with missing (unavailable) values, considering these variables difficult to accurately collect.

8.6. Select only variables for which there is at least one statistically significant category ($\alpha \leq 0.01$) according to the logistic regression analysis. The outcome of this step produces a reduced data set compared to the initial one.

9. Algorithms to create a decision tree

NOTE: Machine-learning algorithms must be properly parameterized to predict which individuals are likely to have a positive MCI test result. One of the main problems while screening for a condition is that the original data is expected to be imbalanced (i.e., few positive cases compared to the negative ones). To get models with balanced data we used a technique called down-sampling, or random sampling, to equalize the frequency with that of the lowest frequency class[31]. Efficient screening also requires reducing the number of false negatives as much as possible (i.e., increasing the sensitivity of the selection of participants suffering from MCI). One of the techniques used to achieve a greater sensitivity is the introduction of penalties in the calculation of Gini's impurity index (i.e., the index used by the algorithm to select the best split for the decision tree)[32].

9.1. Generate a training and test data set with 80% and 20% of the whole data set, respectively using the *createDataPartition* function in the caret library[33].

9.2. Apply the algorithms used to generate decision trees to the training data set. Use the *train again* function in the caret library[33]. The following steps are different parameters of the function; for instance, the tree used in this paper was generated with *rpart*[29] (*method="rpart"),* but other algorithms are available.

9.2.1. Select the 'down sampling' sampling method and introduce the *sampling = "down"* parameter into the caret.

9.2.2. Set the prior probabilities for both classes.

Deleted: ¶

Deleted: ¶
For every ordinal variable, choose an adequate contrast for the variable. For other … ... [82]

Deleted: ¶
For categorical variables, calculate a univariate logistic regression with a response variable for the …creening for MCI. Analyze the outcome of the regression with a contingency table, the *p*-value, sample odds ratio, and a ... [83]

Deleted: ¶
For quantitative variables, calculate the mean, the …tandard deviation, and the …oefficient of logistic regression, and in addition to the ... [84]

Deleted: ¶
Reject variables with missing values …not unavailable),…values, considering that …hese variables are difficult to collect … ... [85]

Deleted: ¶
Select only variables for which where …here is at least one statistically significant category (α … < … 0.01) according to the logistic regression analysis. The outcome of this step produces is … ... [86]

Formatted: Space After: 12 pt

Formatted: Indent: Left: 0 cm, First line: 0 cm, Space After: 12 pt, Add space between paragraphs of the same style, Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 2 + Alignment: Left + Aligned at: 1,27 cm + Indent at: 1,9 cm, Tab stops: Not at 1,27 cm

Deleted: ¶
NOTE: MParameterizing machine …chine-learning algorithms must be properly parameterized is an essential step …o get a prediction…which of …ndividuals who …re likely to have a be positive in the …CI test result. During the screening of a condition, Oo…e of the main problems while screening for a condition is that the original data is expected to be imbalanced (i.e., few positive cases in …omparison…d with ... [87]

Deleted: ¶
Generate a training and a …est data set with 80% and 20% of the whole data set, respectively using with … ... [88]

Deleted: ¶
Apply the algorithms used to generate decision trees to the training data set. Use the function … ... [89]

Formatted: Font: Italic

Deleted: . ; For …or instance, the resultant …ree used in the th ... [90]

Formatted: Font: Italic

Deleted: ¶
Select the 'down sampling' sampling method and introduce the : down sampling. The parameter to introduce in caret is ... [91]

Deleted: ¶

9.2.3. Provide a loss matrix with the Gini's impurity index penalties applied in order to focus on the increasing sensitivity.

9.2.4. For every parameter in the algorithm, choose an appropriate grid of values.

9.2.5. Use a cross-validation estimation of the receiver operating curve (ROC) values to select the best models within the parameter grid.

9.3. Calculate a confusion matrix and the area under the ROC curve (AUC) for the test set prediction to assess the true performance of the model.

REPRESENTATIVE RESULTS:

The participating pharmacies gathered data from 728 users and collected demographic variables in addition to the drugs prescribed to the participants. A univariate logistic regression was performed for all the variables[34]; the error bar graphs shown in **Figure 3** and **Figure 4** are convenient graphical representations of the confidence interval of the odds ratio (for qualitative variables) and the confidence interval of the coefficient of the logistic regression (for quantitative variables). Variables with *p*-values exceeding 0.01 (sex, age, education level, reading habit, time spent sleeping, depression, and memory complaints) were selected and used to generate a white-box model based on a decision tree. This decision tree was generated using a training data set comprising 583 individuals as an input and was validated with a test set of a cohort of 145 participants.

After using the caret[33] library in R, the resultant tree assigned a probability of suffering MCI to each individual depending on their final node in the tree (depicted in **Figure 5**) as well as their answers to a few questions. To evaluate the forecasting capability of these probabilities, a ROC analysis of the test set was performed (**Figure 6**); its AUC was 0.763 and its 95% confidence interval was (0.6624, 0.8632). In addition to the probabilities, the tree shown in **Figure 5** also used very simple questions about how long the person sleeps and how often they read, to recommend (with a sensitivity of 0.76 and specificity of 0.70) whether patients should take the MCI tests.

Using this decision tree and short interview to select users at risk of MCI we were able to significantly reduce the number of patients requiring MCI tests (administration is quite time-consuming). This reduction can be estimated by using data in the test set and interpreting the confusion matrix of the observed and predicted classes shown in **Table 1**. In this work, 55 out of 145 participants in the test set were identified by the decision tree for further MCI testing (representing a reduction of 62% of users taking the tests) while also selecting most of the individuals (19 out of 25) who were positive for MCI.

FIGURE AND TABLE LEGENDS:

Figure 1. Flowchart of the research study and the proposed selective screening. The left side represents the initial study whose data were analyzed with machine-learning techniques to

Deleted: ¶
Provide a loss matrix with the penalties applied to the ... [92]

Deleted: ¶
For every parameter of ... [93]

Deleted: ¶

Deleted: ¶

Deleted: ¶

Deleted: participant …articipating pharmacies gathered data from 728 users,…and collecting …ollected demographic variables in addition to the drugs prescribed to the participants users… A univariate logistic regression is …as performed for all the variables[34]. … The …he error bar graphs shown in **Figure 3** and **Figure 4** are convenient graphical representations offor…the confidence interval of the odds ratio (for qualitative variables) and the confidence interval of the coefficient of the logistic regression (for quantitative variables). The v…ariables with a … ... [94]

Deleted: larger thanexceeding 0.01 (sex, age, education levelal attainment… reading habit, time spent sleeping time, depression, and memory complaints) are …ere selected selected and used to generate a white …hite-box model based on a decision tree. Thise…decision tree is …as generated employing …sing the … training data set as an input,…which consists of…omprising 583 individuals,…as an input and is …as validated with the … test set of a cohort of with 145 users… ... [95]

Formatted: Font: Italic

Formatted: Space After: 12 pt

Deleted: assigns assigned a probability of suffering MCI for to each individual depending on their final node in the tree (as … ... [96]

Formatted ... [97]

Deleted: based ontheir answers to a few questions to the participant. To With the purpose of …valuateing…the forecasting capability of these probabilities, a ROC analysis of the test set wai… performed,…(consisting of the ROC curve displayed in **Figure 6** ... [98]

Formatted: Font: Not Bold

Deleted: whose …UC is …as 0.763. …and its The …5% confidence interval of AUC is …as (0.6624, 0.8632). In addition to the probabilities, the tree shown in **Figure 5** also recommends whether or notwhether users should take the MCI tests with…ses… very simple questions about how long the user …erson is …leeping… and how often they are reading… to recommend (with a sensitivity of 0.76 and specificity of 0.70) whether patients should take the clinical ... [99]

Deleted: ¶
As a result,U u…ing this decision tree and a …hort interview ... [100]

Deleted: ¶

Deleted: flow on the …eft side represents the initial study whose data are …ere analyzed with machine … ... [101]

propose the selective screening for early detection of MCI shown in the right panel. This figure was modified from Climent[34].

Figure 2. Protocol for primary healthcare action. An example of primary healthcare actions that should be considered for early MCI detection before the patient is referred for a medical diagnosis by specialists.

Figure 3. Example of the variables selected during preprocessing. A 99% confidence interval of the odds ratio was calculated and is represented as an error bar. The base value for the logistic regression is indicated below the name of the variable at the top of every panel. For every value of the variable, an error bar represents the confidence interval of the odds ratio of taking that value versus taking the base value. Because the variables used to generate the tree were selected, the confidence intervals do not include the value 0 for some values as these showed significant differences. The scale of the vertical axis is logarithmic to help in comparisons across groups.

Figure 4. Example of non-selected variables during preprocessing. A 99% Confidence Interval of the odds ratio was calculated and is represented with an error bar. The base value for the logistic regression is indicated below the name of the variable at the top of every panel. For every value of the variable, an error bar represents the confidence interval of the odds ratio of taking that value versus taking the base value. In contrast with the previous figure, all the confidence intervals of the selected variables include the value 0, since no significant differences were found to be included to generate the tree. The scale of the vertical axis is logarithmic to help comparison across groups.

Figure 5. Proposed partition tree for selection of pharmacy users. The following tree shows the selection algorithm for MCI tests for individuals aged over 65 years. The text at the top of the box corresponds to the recommendation of taking the MCI screening tests, the two numbers below are the estimated probability of a negative or positive MCI testing outcome, respectively. The value at the bottom of the box is the percentage of individuals with these characteristics in the training set. The warmer the color of the box, the more likely the MCI tests was positive. The top node corresponds to the question about whether the participant has a memory complaint. If the individual does not have a memory complaint, the tree leads to the left branch and the ensuing questions ask about the individual's sex; patients with a memory complaint are asked about the amount of time they sleep per day. This figure was modified from Climent[34].

Figure 6. Receiver operating curves for the partition tree and sensitivity and specificity of the final decision in the test set. The graph represents the ROC curve of the probabilities assigned by the partition tree algorithm in the test set. The red surface corresponds to the AUC and the blue point on the curve shows the sensitivity and specificity of the final recommendation made by the tree.

Table 1. Confusion matrix. Confusion matrix of the predicted and observed values in the test set which were used to validate the proposed model.

Deleted: a selective screening for early detection of MCI shown in the , …ight panel. This figure has been… ... [102]

Deleted: ¶
Figure 2. Protocol for of action in …rimary healthcare action. An example of primary healthcare actions that should to be considered for early MCI detection After a user tests positive for MCI…before,… …he patientuser…is referred to for a medical diagnosis to…y specialists.. …¶
¶ ... [103]

Deleted: ¶ ... [104]

Deleted: ¶
Figure 4. Example of non-selected variables during preprocessing. A 99% Confidence Interval of the odds ratio is was calculated and is represented with an error bar. The base value for the logistic regression is indicated below the name of the variable at the top of every panel. For every value of the variable, an error bar represents the confidence interval of the odds ratio of taking that value versus taking the base value. In contrast with the previous figure, all tT…e confidence intervals of the selected variables include the value 0, since no significant differences were found to be included to generate the tree. . The value taken as the base for the logistic regression is indicated with the base below the name of the variable. For the other values of the variables, an error bar is represented comparing with the base value. … ... [105]

Deleted: ¶
Figure 5. Proposed partition tree for selection of pharmacy users. TFor the users older than 65 years old, t…e following tree shows the selection algorithm of selection …or MCI tests for individuals aged over 65 years. The textvalue…at the top of the box corresponds to the recommendation offrom…taking the MCI screening tests, the two numbers below are the estimated probability of being … negative and or positive in …CI testing outcome, respectively. The value at the bottom of the box is the percentage of individuals with these characteristics in the training set. The warmer the color of the box, the more likely the MCI tests is …as positive. The top node corresponds to the question about whether of the participant has absence of … memory complaint. A …f the individual does not have a positive answer …emory complaint, the tree leads heads …o the left branch and the followed ensuinfollowin… by …uestions ask about on …he user's …ndividual's sex; Users…atients with a whereas a memory complaint negative answer …re asks …sked about ... [106]

Deleted: ¶
Figure 6. R…eceiver operating OC …urves for the partition tree and sensitivity and specificity of the final decision in the test set. The graph represents the ROC curve of the probabilities assigned by the partition tree algorithm in the test set. The red surface corresponds to the AUC and the blue point on the curve shows the sensitivity and specificity ... [107]

Deleted: ¶
Table 1. Confusion matrix. Confusion matrix of the predicted and observed values in the test set,… ... [108]

DISCUSSION:

After searching for terms associated with MCI in Cochrane studies in the PubMed database, a specific questionnaire was created for this study which used the most evident variables with a proven association with MCI. Demographic, lifestyle, and social factors, as well as the patient's pharmacotherapy and some relevant pathologies were also recorded. Additionally, the SPMSQ and MMSE MCI tests were also selected. Importantly, the SPMSQ was not affected by participants' level of schooling. Pharmacists were trained to administer this study and communication with primary and specialized care was assured via letters informing them of this work. Only specialized healthcare providers could definitively make a diagnosis if MCI was suspected as a result of these tests.

In conclusion, in this study we screened for MCI among a population with a low prevalence of the condition (17%). We designed of a set of selection criteria for use with machine-learning techniques which increased the percentage of MCI positives up to more than 30% among the selected users. Consequently, these tools help increase the screening efficiency and substantially reduce the cost of mass screening among the population group selected by the decision tree.

A limitation of this method is that the decision tree may become invalid in this specific cohort as the population changes and thus, will likely require periodic updates. For instance, many individuals in this population were illiterate, but the number of illiterate individuals aged over 65 years will decrease in the future. These demographic changes will affect the variables related to reading and will require future recalibration of the decision tree.

Remarkably, this data-driven model provided information about the most important variables (from among hundreds) in the construction of a concise yet informative and efficient model. Constructing a decision tree provides insight into the best variables to focus on and is both a cost-effective way to help select people for whom further MCI testing is recommended and furthers our knowledge of these populations in this context.

To increase the future percentage detection rate of MCI, we will require new cost-effective techniques that can assure increased effectiveness. This protocol is time-consuming and is difficult for pharmacists to integrate into their daily work. Thus, other tests such as the MoCA[22] or SLUMS[23] (both with adequate sensitivity and specificity) could be considered for fast the detection of MCI in the future.

A systematic evaluation of the trade-off between specificity and test duration should improve the effectiveness of the set of MCI tests used for screening. Moreover, relevant quantitative variables included in the study should have a wide range so that an efficient cut-off can be selected for them; a narrow range would exclude a large portion of the population from early detection. For instance, the age variable (which is always considered an important criteria in MCI diagnoses) was not considered relevant in this decision tree because the recruitment criteria (age over 65 years) was too conservative; inclusion of younger individuals in a future study would allow the optimal age for starting MCI screening to be calculated.

Deleted: ¶

Deleted: on in Cochrane studies and …n the PubMed database, a specific questionnaire was created for this study was created with…hich used the most evident variables that with a were …roven to be …ssociated…on with MCI. Demographic, lifestyle, and social factors, as well as the patient's pharmacotherapy and some relevant pathologies were also recorded. Additionally, the SPMSQ and MMSE MCI tests were also the MCI tests …elected. Importantly, the SPMSQ wai… not affected by the …articipants' level of schooling. Pharmacists were trained to take part inadminister this study and communication with primary and and specialized care was assured through …ia informing letters informing them of this work. Only specialized healthcare providers can …ould definitively make a diagnosis if MCI wai… ... [109]

Deleted: ¶
As a conclusion of this study… in this study we screened for ... [110]

Deleted: ¶
A limitation of the …his method is that the decision tree may become incould be …alid in for …his a …pecific cohort as the population changes and thus, will likely require , by …eriodic ally updating …pdatesthe decision tree… For instance, many individuals in the …his population were illiterate, but the number of illiterate individuals aged over s among elder than 65 years will is going to …ecrease according to Spanish demographyin the future. These demographic changes will are affecting…the variables related to reading and will require future recalibration of the decision tree is needed. ... [111]

Deleted: ¶
Remarkably, this a …ata-driven model utilization …rovides… information about knowledge on which information …he most important variables (from among hundreds) in the is the most important to construction of a reduced model from from hundreds of variables and to be …oncise yet informative and efficient model. CThe c…nstruction…g of…a decision tree provides insight intoon…which …he best variables to focus on and is both a cost-effective way to help . This selects…people for whom further MCI testing is recommended and for an MCI test in a cost-effective way, butway but …urthers our knowledge of these ... [112]

Deleted: ¶
In the future, in order tTo increase the future percentage of ... [113]

Formatted: Space After: 12 pt

Deleted: employed used for screening. For the study design, Moreover, relevant quantitative variables included in in the study should have a wide range so that to be able to select an efficient cut-off can be selected for them, … a narrow range in the study …ould exclude a large portion of the population from an …arly detection. For instance, the age variable (which is always considered an important criteria in MCI diagnoses) age in the decision tree i…as not considered relevant in this decision tree because due to …he recruiting recruitment criteria (age over 65 years old… was being excessively …oo conservative;, whereas age is always ... [114]

ACKNOWLEDGMENTS:

This work was made possible by the support of the Know Alzheimer Foundation and help from the multimedia production service at the Universidad CEU Cardenal Herrera, especially Enrique Giner. We would like to recognize the work of all the participating pharmacies (SEFAC), and the collaborating doctors from the Society of Primary Care Doctors (SEMERGEN) and Neurology Society (SVN) who helped with the MCI diagnoses, especially Vicente Gassull, Rafael Sánchez, and Jordi Pérez. Finally, we thank all those who agreed to take part in this study.

DISCLOSURES:

The authors have nothing to disclose.

REFERENCES:

1. Prince, M., Comas-Herrera, A., Knapp, M., Guerchet, M., Karagiannidou, M. *World Alzheimer report 2016: improving healthcare for people living with dementia: coverage, quality and costs now and in the future*. Alzheimer's Disease International (ADI), London (2016).
2. Niu, H., Álvarez-Álvarez, I., Guillén-Grima, F., Aguinaga-Ontoso, I. Prevalence and incidence of Alzheimer's disease in Europe: A meta-analysis. *Neurología (English Edition)*. **32** (8) 523–532 (2017).
3. Petersen, R. C. et al. Mild Cognitive Impairment: Clinical Characterization and Outcome. *Archives of Neurology*. **56** (3), 303 (1999).
4. Petersen, R. C. et al. Practice guideline update summary: Mild cognitive impairment: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. *Neurology*. **90** (3) 126–135 (2018).
5. Climent, M. T., Vilaplana, A. M., Arnedo, A., Vilar, J., Moreno-Royo, L. Life styles associated cognitive impairment. Study from the community pharmacy [Estilos de vida asociados a deterioro cognitivo. Estudio preliminar desde la farmacia comunitaria]. *Revista de Investigacion Clinica*. **65** (6), 500–509 (2013).
6. Faubel, R., López-García, E., Guallar-castillón, P., Graciani, A., Banegas, J. R., Rodríguez-Artalejo, F. Usual sleep duration and cognitive function in older adults in Spain. *Journal of Sleep Research*. **18** (4), 427–435 (2009).
7. Benito-León, J., Louis, E. D., Bermejo-Pareja, F. Cognitive decline in short and long sleepers: A prospective population-based study (NEDICES). *Journal of Psychiatric Research*. **47** (12) 1998–2003 (2013).
8. Ramos, A. R. et al. Association between Sleep Duration and the Mini-Mental Score: The Northern Manhattan Study. *Journal of Clinical Sleep Medicine*. **9** (7) 669 (2013).
9. Gabelle, A., Gutierrez, L.-A. et al. Excessive Sleepiness and Longer Nighttime in Bed Increase the Risk of Cognitive Decline in Frail Elderly Subjects: The MAPT-Sleep Study. *Frontiers in Aging Neuroscience*. **9**, 312 (2017).
10. Ranstam, J., Merlo, J., Blennow, G., Hanson, B. S., Östergren, P.-O., Melander, A. Impaired cognitive function in elderly men exposed to benzodiazepines or other anziolytics. *The European Journal of Public Health*. **7** (2), 149–152 (1997).
11. Airagnes, G., Pelissolo, A., Lavallée, M., Flament, M., Limosin, F. Benzodiazepine Misuse

Deleted: ¶
Deleted: has
Deleted: been
Deleted: with
Deleted: the
Deleted: of
Deleted: want
Deleted: pharmacies
Deleted: participants
Deleted: and
Deleted: who have collaborated with the diagnosis
Deleted: of Primary Care
Deleted:
Deleted: want to
Deleted: e
Deleted: people
Deleted: ¶
Deleted: ¶
Formatted: Spanish

in the Elderly: Risk Factors, Consequences, and Management. *Current Psychiatry Reports*. **18** (10), 89 (2016).
12. Fernández, M. et al. [Prevalence of dementia in the elderly aged above 65 in a district in the Basque Country]. *Revista de Neurologia*. **46** (2), 89–96 (2008).
13. Velert Vila, J., Velert Vila, M. del M., Salar Ibáñez, L., Avellana Zaragoza, J. A., Moreno Royo, L. Suitability of the use of benzodiazepines prescribed by the pharmacist in the elderly. A doctor-pharmacist collaboration study. *Atención Primaria*. **44** (7), 402–410 (2012).
14. Alacreu, M. et al. Importance of Increasing Modifiable Risk Factors Knowledge on Alzheimer's Disease Among Community Pharmacists and General Practitioners in Spain. *Frontiers in Pharmacology*. **10**, 860 (2019).
15. Cohen, J. *Statistical power analysis for the behavioral sciences*. Routledge (2013).
16. Nardi, P. M. *Doing survey research: A guide to quantitative methods*. Routledge (2018).
17. Sabater Hernández, D., Silva Castro, M. M., Faus Dáder, M. J. *Método Dáder: guía de seguimiento farmacoterapéutico*. (Grupo de Investigación en Atención Farmacéutica (GIAF): 2007).
18. Pfeiffer, E. A Short Portable Mental Status Questionnaire for the Assessment of Organic Brain Deficit in Elderly Patients†. *Journal of the American Geriatrics Society*. **23** (10), 433–441 (1975).
19. Folstein, M. F., Folstein, S. E., McHugh, P. R. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. **12** (3), 189–198 (1975).
20. Buschke, H. et al. Screening for dementia with the memory impairment screen. *Neurology*. **52** (2), 231–238 (1999).
21. Verghese, J. et al. Picture-based memory impairment screen for dementia. *Journal of the American Geriatrics Society*. **60** (11), 2116–2120 (2012).
22. Nasreddine, Z. S. et al. The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society.* **53** (4), 695–699 (2005).
23. Kaya, D. et al. The Saint Louis University Mental Status Examination Is Better than the Mini-Mental State Examination to Determine the Cognitive Impairment in Turkish Elderly People. *Journal of the American Medical Directors Association*. **17** (4), 370.e11-370.e15 (2016).
24. O'Caoimh, R. et al. Comparison of the quick mild cognitive impairment (Qmci) screen and the SMMSE in screening for mild cognitive impairment. *Age and Ageing*. **41** (5), 624–629 (2012).
25. Cullen, B., ONeill, B., Evans, J. J., Coen, R. F., Lawlor, B. A. A review of screening tests for cognitive impairment. *Journal of Neurology, Neurosurgery, & Psychiatry*. **78** (8), 790–799 (2007).
26. Martínez de la Iglesia, J. et al. Adaptación y validación al castellano del cuestionario de Pfeiffer (SPMSQ) para detectar la existencia de deterioro cognitivo en personas mayores e 65 años. *Medicina Clínica*. **117** (4), 129–134 (2001).
27. Blesa, R. et al. Clinical validity of the 'mini-mental state' for Spanish speaking communities. *Neuropsychologia*. **39** (11), 1150–1157 (2001).
28. Champely, S. et al. pwr: Basic Functions for Power Analysis. https://cran.r-project.org/web/packages/pwr/index.html (2018).

Formatted: Spanish

Formatted: Spanish

29. Therneau, T., Atkinson, B., Ripley, B. rpart: Recursive Partitioning and Regression Trees. R package version 4.1--13. https://cran.r-project.org/web/packages/rpart (2018).
30. WHO Collaborating Centre for Drugs Statistics Methodology ATC/DDD Index 2018. https://www.whocc.no/atc_ddd_index/ (2018).
31. Kuhn, M., Johnson, K. *Applied Predictive Modeling*. **26**, (Springer New York: New York, NY, 2013).
32. Therneau, T., Atkinson, E. An introduction to recursive partitioning using the RPART routines. *r.789695.n4.nabble.com*, http://r.789695.n4.nabble.com/attachment/3209029/0/zed.pdf (2019).
33. Kuhn, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*. **28** (5), 1–26 (2008).
34. Climent, M. T., Pardo, J., Muñoz-Almaraz, F. J., Guerrero, M. D., Moreno, L. Decision Tree for Early Detection of Cognitive Impairment by Community Pharmacists. *Frontiers in Pharmacology*. **9**, 1–12 (2018).

Deleted: ¶