amse
école d'économie d'aix-marseille
aix-marseille school of economics

Working Papers / Documents de travail

When algorithms replace biologists: A Discrete Choice Experiment for the valuation of risk-prediction tools in Neurodegenerative Diseases

Ismaël Rafaï
Bérengère Davin-Casalena
Dimitri Dubois
Thierry Blayac
Bruno Ventelou

WP 2024 - Nr 32

When algorithms replace biologists: A Discrete Choice Experiment for the valuation of risk-prediction tools in Neurodegenerative Diseases

Ismaël Rafaï, PhD, Aix Marseille Univ, CNRS, AMSE, Marseille, France

Bérengère Davin-Casalena, PhD, Observatoire Régional de la Santé Paca, Marseille, France

Dimitri Dubois, PhD, CEE-M, Univ Montpellier, CNRS, INRAE, Institut Agro, Montpellier, France

Thierry Blayac, PhD, CEE-M, Univ Montpellier, CNRS, INRAE, Institut Agro, Montpellier, France

Bruno Ventelou, PhD, Aix Marseille Univ, CNRS, AMSE, Marseille, France[1]

bruno.ventelou@univ-amu.fr

[1] We acknowledge financial support from the French government under the "France 2030" investment plan managed by the French National Research Agency (reference: ANR-17-EURE-0020); Excellence Initiative of Aix-Marseille University - A*MIDEX; French National Research Agency (GRANT ANR-20-COVR-00 and ANR-21-JPW2-002).

Abstract.

Background. Earlier detection of neurodegenerative diseases may help patients plan for their future, achieve a better quality of life, access clinical trials and possible future disease modifying treatments. Due to recent advances in artificial intelligence (AI), a significant help can come from the computational approaches targeting diagnosis and monitoring. Yet, detection tools are still underused. We aim to investigate the factors influencing individual valuation of AI-based prediction tools.

Methods. We study individual valuation for early diagnosis tests for neurodegenerative diseases when Artificial Intelligence Diagnosis is an option. We conducted a Discrete Choice Experiment on a representative sample of the French adult public (N=1017), where we presented participants with a hypothetical risk of developing in the future a neurodegenerative disease. We ask them to repeatedly choose between two possible early diagnosis tests that differ in terms of (1) type of test (biological tests vs AI tests analyzing electronic health records); (2) identity of whom communicates tests' results; (3) sensitivity; (4) specificity; and (5) price. We study the weight in the decision for each attribute and how socio-demographic characteristics influence them.

Results. Our results are twofold: respondents indeed reveal a reduced utility value when AI testing is at stake (that is evaluated to 36.08 euros in average, IC = [22.13; 50.89]) and when results are communicated by a private company (95.15 €, IC = [82.01; 109.82]).

Conclusion. We interpret these figures as the shadow price that the public attaches to medical data privacy. The general public is still reluctant to adopt AI screening on their health data, particularly when these screening tests are carried out on large sets of personal data.

Key Points.

Individuals need to be compensated for having a AI-based diagnosis, if compared to non-invasive biological tests.

Individuals are also willing to pay for their results being read and interpreted by their family doctor, instead of themselves, the health insurance, or the test company.

The general public is reluctant to adopt AI screening on their personal data, particularly when these screening tests are carried out on extended sets of behavioral data.

1. Introduction

Neurodegenerative diseases (NDDs hereafter) are characterized by the progressive degeneration of the central nervous system's structure and function, causing physical and cognitive disability [1]. NDDs affect a large group of patients by encompassing various diseases that include Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD), and amyotrophic lateral sclerosis (ALS). AD and PD remain the first and second most common neurodegenerative diseases worldwide [2]. The most evident risk factor for developing this condition is aging [3] and, with the increase of the population average age, the prevalence of NDDs is remarkably increasing, currently affecting approximately 15% of the worldwide population [4]. This increase and the lack of effective treatments [5] lead to an enormous burden on patients and their caregivers [6] as well as healthcare systems, both in terms of direct and indirect costs [7-13]. One of the identified ways to improve patient outcomes and reduce the economic burden on healthcare systems and society is to diagnose ND earlier in a patient's lifetime [14-20]. It has been suggested that earlier detection (even imperfect, with sensitivity and specificity < 100%) may help patients plan for their future, achieve a better quality of life, and access clinical trials and possible future disease modifying treatments [21].

Due to recent advances in artificial intelligence (AI), a significant help can come from the computational approaches targeting diagnosis and monitoring [22-27]. AI-based early diagnosis tests, using genetic data, imaging data, and clinical data, have the potential to change the way we diagnose and manage NDDs [28]. AI-based diagnosis aims to identify patients at risk of developing neurodegenerative diseases earlier, before the onset of symptoms, using data from electronic records and have the potential to change the way we diagnose and manage ND [18, 29]. This would allow for earlier intervention, which can potentially improve outcomes for patients [30]. AI-based

diagnosis can be more efficient than traditional diagnosis methods, as it can process large amounts of data quickly and accurately, reducing the need for time-consuming tests [31]. Furthermore, as telemedicine has been shown to improve access to care in AD and PD [32], AI-based diagnosis can also help to address disparities in access to care and diagnosis by reducing the reliance on specialized expertise [33]. Even in the absence of highly effective treatment options, predictive tests for NDDs may be useful to help patients and families prepare for decisions that need to be made in the future, including advance care planning [34, 35]. Those with high levels of perceived risk -those who are worried about getting NDDs as well as those with more experience with the disease, including caregivers and blood relatives of NDDs patients - are expected to be among those most likely to pursue testing [36, 37]. However, the success of these tests will first depend on the willingness of individuals to adopt and use them [38, 39]. A recent paper showed that the general population is mainly distrustful of AI in medicine [40].

In this context of underuse of NDDs early detection tools [41], we aim to investigate the factors influencing individual valuation of AI-based prediction tools. As suggested by the literature just cited, we hypothesized that sensitivity and specificity of tests, confronted to the price, would be important attributes to assess. We added AI attributes (versus saliva bio-testing method) and the identity of the reader of the results (private company/family doctor/self-reading) in order to take into account the potential opportunities brought by computational approaches. The rest of the paper presents the survey methods and gives results and discussion.

2. Methods

2.1. Data collection

We conducted a discrete choice experiment (DCE) using a sample of participants from the general French adult public. The DCE is a survey method that allows researchers to investigate how people make choices by presenting them with hypothetical goods that vary in certain attributes [42, 43]. In short, the attributes are qualifying the performances of the good -here, a predictive test- and respondents are constrained to trade performance in one attribute for performance in another. For instance, DCE has been recently used to determine the relative importance of attributes of illness that influence the value placed on alleviating that illness [44], to focus on patient preferences for different treatment aspects of cancer [45] or spinal muscular atrophy [46], or to measure the benefit-risk preferences of US adults with heart failure for artificial intelligence (AI)-assisted echocardiograms [47]. In our own DCE design, we selected the forced-choice method (no exit option), a limited number of attributes that characterizes the testing method (5 attributes), and each respondent has had to make 5 consecutive decisions (5 scenarios presenting two options, A and B, that vary on attributes). We provide the details of DCE methodology in Appendix 1 and a sample of choices screen in Appendix 2.

The research has been performed in accordance with the Declaration of Helsinki. Each participant was informed through the consent form that the study would ask them about their testing behavior for various illnesses, both infectious and/or chronic. In October 2022, the institute ViaVoice recruited 1017 participants online with a targeted representativeness of the metropolitan French adult population, in terms of age, gender, socio-professional status and living area, using the quota method. The polling institute was in charge of anonymizing the data, and monitoring ethical procedures with respect to participants (e.g. their right to control their personal data). Ethical

approval was obtained in September 2022 by the Ethic Committee of the University, Number: 2022-10-20-009.

In the DCE section, participants were presented with an average hypothetical risk of developing in the future a neurodegenerative disease of 7% after the age of 65, and with the existence of tests able to predict their specific risk of developing the disease within 10 years. We randomize between participants with equal probability the curability of the disease: the symptoms were either "inevitable and incurable" (*Curable=0*) or "preventable and treatable" (*Curable=1*). See Supplementary information for the exact verbatim.

We then present to participants the 5 different attributes of the tests that were (Table 1): (1) the type of the test, (2) the reading of test results, (3) tests' sensitivity, (4) tests' specificity, (5) tests' price. Details about the possible type of tests were available to participants once at the beginning of the experiment, and at any time if they click on an "*info*" icon present on the decision screens.

[Table 1 here]

The type of the test (attribute 1) could be either a biological analysis (*Type = Bio*), an AI numerical analysis of health data (*Type = AI)* or an AI numerical analysis of health data, lifestyle habits and consumption patterns (*Type = AI+*). Reading of the results (attribute 2) can be made by their personal doctor (*Reading = doctor*), by the individuals themselves (*Reading = Self*), by the National Health Insurance (*Reading = NHI*), or by a private company selling the test (*Reading = Private*). Sensitivity (attribute 3) could take the values of either 60%, 70% or 95%. Test sensitivity was presented within each decision screen, as follows: "*[60% / 70% / 95%]* of the individuals who will develop the disease in the next 10 years would be correctly declared positive by the tests." Specificity (attribute 4) could take the values of either *60%*, *70%* or *95%*, but was communicated using type II error rate (1-specificity), since we thought this concept was easier for the general

population, as follows: "*[5% / 30% / 40%] of the individuals with no risk of developing the disease in the next 10 years are incorrectly declared positive by the test*". Price of the test (attribute 5) could amount to 0 euros, 20 euros or 90 euros, paid out of pocket by the person carrying out the test.

After the DCE, we collected the following characteristics for the participants: gender (48.48% women), age (M=52.54, SD=16.31), household income (M = 3.759€, SD = 2.312€), level of education (25% high school or less, 34% graduate), and cognitive abilities through the Cognitive Reflection Test-CRT [48]. Table on sociodemographics of the sample is presented in Appendix 3.

2.2. Statistical Methodology.

We assume that when facing the choice set k, the individual *i* chooses the test j=A or B, to maximize the following random utility function:

$$U\_ijk = \alpha\_j + X'\_jk \times (\beta + Curable\_i \times \beta\_C\ ) + e\_ijk$$

With $X'\_jk = (AI, AI + \_\ , Self, Company, NHI, Sensitivity, Specificity, Price)\_jk$, the vector attributes of the alternative *j* of scenario *k*; α_j a constant that equals to 0 if $j = B$; $Curable_i$=1 if the individual *i* was presented with a curable disease; β and β_c the vectors of marginal utilities associated with each attribute (according to disease curability), and e_{ijk} an error term independently distributed with an extreme value distribution (We also tested an alternative model without the assumption of linearity in the Price, Sensitivity and Specificity attributes (appendix A4). This alternative specification leads to qualitatively similar results and does not outperform the linear model (LR–test: $\chi^2(6)$=7.92,p=0.244)).

We thus estimate by maximum likelihood the probability of choosing an option *j,* using a conditional logit function, for the general population. We also do the analysis without the interaction with disease curability for several subsamples, according to individuals' gender, age (*low* if age≤34 and *high* if age≥64), education (*low* if high school or less; and *high* if master or more), income (*low* if household income ≤ 2k€ per month and *high* if ≥ 5k €), and CRT (*high* if CRT ≥2 and *low* if CRT ≤1).

Based on model specification, we estimate marginal willingness to pay at the mean (WTP), for each attribute described in X' (except *Price*), by dividing each attribute's estimated marginal utility (e.g. β_{AI} , for AI) by the estimated marginal utility of money (β_{Price}). Confidence intervals of the WTP estimates were calculated according to Krinsky and Robb's simulation method [49], using the R package "*support.CEs*" (version 0.3-0).

3. Results

Table 2 presents the regression results for the general population. Both columns represent results from the same regression on the general population. The first column shows the marginal effect of each attribute on utility when the disease is not curable, while the second column shows the additional effect of each attribute when the disease is curable.

[Table 2 here]

As predicted, we observe positive marginal utility for sensitivity and specificity, and a negative marginal utility for price. We observe a disutility of AI tests if compared to biological tests. Who was reading the test was also important: individuals exhibit an aversion for self-reading, reading by the National Health Insurance or a reading by a private company (if compared with a reading

by a family doctor). All those effects were significant at the 0.001 level. Note that reading by a private company was significantly worse than self-reading and reading by the NHI ($p< 0.001$).

The interaction with curability (column 2) was only significant for this "reading" attribute: curability adds a negative value for a reading by a private company (*p=0.018*), if compared with family doctor reading. Aversion to self-reading was marginally strengthened (*p=0.085*). The non-significance of the other coefficients on column 2 suggests that the framing of curability does not influence much the relative valuation of specificity, sensitivity, price, and tests type.

The DCE makes it possible to study how attributes can be traded with each other, using the monetary attribute as a common standard; Figure 1 is giving this information, for the whole sample and for stratifications on a selection of relevant individual variables: gender, income, etc. (as in [50], money-equivalent values are estimated for a change/improvement between attribute levels within the range used in the survey). Note that the design of the DCE study, using forced-choice method, does not allow to study the **levels** of WTPs in relation to the individual variables, but we can study **variations** of WTP -under changes in the test characteristics- in relation to individual variables.

[Figure 1 here]

Concerning the type of the test, and if compared to a biological test, individuals are on average willing to pay 36.08€ (IC = [22.13 ; 50.89]) to avoid an AI test based on health records and are willing to pay 61.32€ (IC = [41.12 ; 83.07]) to avoid it when it also exploits consumption and lifestyle personal data provided by another company (AI+). We did not observe statistical differences between subpopulations.

Concerning the reading of the results and compared to a reading made by their family doctor, individuals are willing to pay 49.97€ (IC = [39.15 ; 61.29]) for not reading the results themselves, 46.68€ (IC = [34.13 ; 60.45]) for the results not communicated directly by the NHI and 95.15€ (IC = [82.01 ; 109.82]) for the results not being read by a private company commercializing the test.

Finally, concerning tests' accuracy, on average, individuals are willing to pay 2.30€ (IC = [1.99; 2.65]) to increase sensitivity by 1% point and 2.60€ (IC = [2.26; 2.99]) to increase specificity by 1% point. Individuals with higher income are willing to pay more for increasing sensitivity (M = 3.80€, IC = [2.86; 5.05]) if compared with individuals with lower income (M = 1.61€, IC = [1.11; 2.18]). The willingness to pay for specificity is also higher for individuals with higher cognitive abilities (M = 3.59€, IC = [2.93; 4.39]), income (M = 3.58€, IC = [2.64; 4.86]), and education (M = 3.13€, IC = [2.54; 3.86]), if compared with individuals with lower cognitive abilities (M = 1.98€, IC = [1.59; 2.42]), income (M = 1.81€, IC = [1.29; 2.43]), and education (M = 1.59€, IC = [0.98; 2.31]).

4. Discussion.

Our results suggest that individuals need to be compensate for having a AI-based diagnosis, if compared to non-invasive biological tests, and that both sensitivity and specificity positively impact the probability of opting for a test. Individuals are also willing to pay for their results being read and interpreted by their family doctor, instead of themselves, the public health insurance, or the company selling the tests. The difference in valuation between a reading made by their family doctor and by the company selling the tests is strengthened when the disease is curable.

We conducted stratified analysis to investigate how attribute values depend on individuals' sociodemographic data. We found that, while individual characteristics (demographic and socioeconomic status) do not impact much the valuation for the type of test or the identity of who reads the test, they could impact on specificity and sensitivity monetary trade-off. For example, individuals with higher income are more willing to pay for improvements in sensitivity than other parts of the society, and specificity is more important for individuals with higher levels of education, income, or cognitive abilities.

We were surprised that the curability of the disease only plays a significant role in the valuation of the who-is-the-reader attribute. The result that concerns the "reader" can be explained: when a therapy exists, individuals value more the expertise of their family doctor more (e.g., to prescribe the adapted treatment). Absence of influence on other attributes is surprising, particularly for sensitivity, since theory predicts that willingness for appropriate testing should be higher when the disease is curable [51]. Public knowledge about the neurodegenerative diseases remains poor [52]; we might suppose that respondents have "automatically" associated neurodegenerative diseases with no curability [53], as well as for cancer [54].

Last, considering the reluctance to have their tests read and interpreted by a private company and the extra loss associated with the use of algorithms analysis on extended (commercial) datasets, we can put the result in few words: the public do not dislike algorithm *per se*, they mainly have a problem of privacy concern with their medical data. The literature showed that patients are open towards the usage of AI in healthcare [55], but there is also a patients' concern over information, privacy, and safety [56], including for neurodegenerative diseases such Alzheimer [57]. Given personal medical information is among the most private and legally protected forms of data, future scaling-up of commercial healthcare AI will first have to face serious privacy challenges [58].

Although the analyzes undertaken in this study used large representative samples and rigorous methods, some practical limitations should be recognized before generalizing the results. The first is the question of the intelligibility of the decision task (choice of tests) and the scenarios presented in the DCE survey. This is a weakness, true for all DCE studies which include a large number of attributes, and which we tried to limit by being very restrictive on the number of attributes (5 is an acceptable figure, considering of literature, e.g. [59]): but we must recognize that we did not ask a question of intelligibility at the end of the DCE part. In the same vein, certain individual variables may appear to us today to be missing: family history about neurodegenerative diseases or individual mental health, for example, which could affect the choice of tests. As the questionnaire was followed by other parts (dedicated to infectious diseases), we had to limit ourselves on the number of questions to ask. The second limitation could lie in the choice of the "forced-choice methodology"; in the appendix we provide a review of the literature on this point (appendix A1).

In conclusion, this DCE study shows, for the general public, a reluctance to adopt AI screening on their health data, particularly when these screening tests are carried out on large sets of personal data. They mainly have a confidentiality problem with their medical data. If it wishes to increase the use of these methods, the public authority should consider better guaranteeing the confidentiality and honest use of personal medical information.

Declarations

Funding/support: French government under the “France 2030” investment plan managed by the French National Research Agency (reference: ANR-17-EURE-0020); Excellence Initiative of Aix-Marseille University - A*MIDEX; French National Research Agency (GRANT ANR-20-COVR-00 and ANR-21-JPW2-002).

Conflict of interest. The authors have no conflict of interest to declare.

Authors contributions. All authors designed the study ; **TB** defined the scenarios for the DCE, **DD** coded the experiment ; **IR** made the statistical analysis ; **IR**, **BDC** and **BV** wrote the first draft of the article; **All authors** reviewed the article.

References.

[1] Van Schependom J, D'haeseleer M. (2023). "Advances in neurodegenerative diseases." *Journal of clinical medicine*, *12*(5), 1709.

[2] Erkkinen MG, Kim MO, Geschwind MD. (2018). "Clinical neurology and epidemiology of the major neurodegenerative diseases." *Cold Spring Harbor perspectives in biology*, *10*(4), a033118.

[3] Hou Y, Dan X, Babbar M, et al. (2019). Ageing as a risk factor for neurodegenerative disease. *Nature Reviews Neurology*, 15(10), 565-581.

[4] Feigin VL, Vos T, Nichols E, et al. (2020). The global burden of neurological disorders: translating evidence into policy. *The Lancet Neurology*, 19(3), 255-265.

[5] Peplow PV, Martinez B, Gennarelli TA. (2022). "Prevalence, needs, strategies, and risk factors for neurodegenerative diseases." *Neurodegenerative Diseases Biomarkers: Towards Translating Research to Clinical Practice*, 3-8.

[6] Fakeye MBK, Samuel LJ, Drabo EF, et al. (2023). "Caregiving-related work productivity loss among employed family and other unpaid caregivers of older adults." *Value in Health*, *26*(5), 712-720.

[7] Achtert K, Kerkemeyer L. (2021). The economic burden of amyotrophic lateral sclerosis: a systematic review. *The European Journal of Health Economics*, 22(8), 1151-1166.

[8] Cantarero-Prieto D, Leon PL, Blazquez-Fernandez C, et al. (2020). The economic cost of dementia: a systematic review. *Dementia*, 19(8), 2637-2657.

[9] Dauphinot V, Potashman M, Levitchi-Benea M, et al. (2022). Economic and caregiver impact of Alzheimer's disease across the disease spectrum: a cohort study. *Alzheimer's Research & Therapy*, 14(1), 34.

[10] Jönsson L, Tate A, Frisell O, et al. (2023). The costs of dementia in Europe: an updated review and meta-analysis. *Pharmacoeconomics*, 41(1), 59-75.

[11] Meijer E, Casanova M, Kim H, et al. (2022). Economic costs of dementia in 11 countries in Europe: Estimates from nationally representative cohorts of a panel study. *The Lancet Regional Health–Europe, 20.*

[12] Rodríguez-Santana I, Mestre T, Squitieri F, et al. (2023). Economic burden of Huntington disease in Europe and the USA: Results from the Huntington's Disease Burden of Illness study. *European Journal of Neurology*, 30(4), 1109-1117.

[13] Yang W, Hamilton JL, Kopil C, et al. (2020). Current and projected future economic burden of Parkinson's disease in the US. *npj Parkinson's Disease*, 6(1), 15.

[14] DeKosky ST, Marek K. (2003). "Looking backward to move forward: early detection of neurodegenerative disorders." *Science*, *302*(5646), 830-834.

[15] Zahra W, Rai SN, Birla H, et al. (2020). "The global economic impact of neurodegenerative diseases: Opportunities and challenges." *Bioeconomy for sustainable development*, 333-345.

[16] Logroscino G, Urso D, Savica R. (2022). "Descriptive epidemiology of neurodegenerative diseases: What are the critical questions?" *Neuroepidemiology*, *56*(5), 309-318.

[17] Rasmussen J, Langerman H. (2019). "Alzheimer's disease–why we need early diagnosis." *Degenerative neurological and neuromuscular disease*, 123-130.

[18] Nedelec T, Couvy-Duchesne B, Monnet F, et al. (2022). "Identifying health conditions associated with Alzheimer's disease up to 15 years before diagnosis: an agnostic study of French and British health records." *The Lancet Digital Health*, *4*(3), e169-e178.

[19] Zucchella C, Sinforiani E, Tamburin S, et al. (2018). “The multidisciplinary approach to Alzheimer's disease and dementia. A narrative review of non-pharmacological treatment.” *Frontiers in neurology*, *9*, 1058.

[20] Schapira AH, Olanow CW. (2008). “Drug selection and timing of initiation of treatment in early Parkinson's disease.” *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *64*(S2), S47-S55.

[21] Ford E, Milne R, Curlewis K. (2023). Ethical issues when using digital biomarkers and artificial intelligence for the early detection of dementia. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 13(3), e1492.

[22] Borchert RJ, Azevedo T, Badhwar A, et al. (2023). Artificial intelligence for diagnostic and prognostic neuroimaging in dementia: A systematic review. *Alzheimer's & Dementia*, 19(12), 5885-5904.

[23] Javed AR, Saadia A, Mughal H. et al. (2023) Artificial Intelligence for Cognitive Health Assessment: State-of-the-Art, Open Challenges and Future Directions. *Cogn Comput* 15, 1767–1812. https://doi.org/10.1007/s12559-023-10153-4

[24] Myszczynska MA, Ojamies PN, Lacoste AM, et al. (2020). Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nature Reviews Neurology*, 16(8), 440-456.

[25] Tăuţan AM, Ionescu B, Santarnecchi E. (2021). “Artificial intelligence in neurodegenerative diseases: A review of available tools with a focus on machine learning techniques.” *Artificial Intelligence in Medicine*, *117*, 102081.

[26] Termine A, Fabrizio C, Strafella C, et al. (2021). « Multi-Layer Picture of Neurodegenerative Diseases: Lessons from the Use of Big Data through Artificial Intelligence." *Journal of Personalized Medicine*, *11*(4), 280.

[27] Thesmar D, Sraer D, Pinheiro L, et al. (2019). Combining the power of artificial intelligence with the richness of healthcare claims data: opportunities and challenges. *PharmacoEconomics*, 37, 745-752.

[28] Feng T. (2023). Applications of Artificial Intelligence to Diagnosis of Neurodegenerative Diseases. *Studies in Health Technology and Informatics*, 308, 648-655.

[29] Shusharina N, Yukhnenko D, Botman S, et al. (2023). "Modern Methods of Diagnostics and Treatment of Neurodegenerative Diseases and Depression." *Diagnostics*, *13*(3), 573.

[30] Bakirtzis C, Boziki MK, Grigoriadis N. (2023). Prevention, Intervention and Care of Neurodegenerative Diseases. *Healthcare,* 11(16), p. 2349.

[31] Khanna NN, Maindarkar MA, Viswanathan V, et al. (2022) "Economics of Artificial Intelligence in Healthcare: Diagnosis vs. Treatment." *Healthcare*, 10, 2493. https://doi.org/10.3390/healthcare10122493

[32] Adams JL, Myers TL, Waddell EM, et al. (2020). Telemedicine: a valuable tool in neurodegenerative diseases. *Current geriatrics reports*, 9, 72-81.

[33] Sirilertmekasakul C, Rattanawong W, Gongvatana A, et al. (2023). The current state of artificial intelligence-augmented digitized neurocognitive screening test. *Frontiers in Human Neuroscience*s, 17, 1133632.

[34] Neumann PJ, Hammitt JK, Mueller C, et al. (2001). Public attitudes about genetic testing for Alzheimer's disease. *Health Affairs*, 20(5), 252-264.

[35] Sheffrin M, Stijacic Cenzer I, Steinman MA. (2016). Desire for predictive testing for Alzheimer's disease and impact on advance care planning: a cross-sectional study. *Alzheimer's research & therapy*, 8, 1-7.

[36] Wikler EM, Blendon RJ, Benson JM. (2013). "Would you want to know? Public attitudes on early diagnostic testing for Alzheimer's disease." *Alzheimer's research & therapy*, *5*(5), 1-11.

[37] Angelidou IA, Makri M, Beyreuther K, et al. (2023). Attitudes toward pre-symptomatic screening for Alzheimer's dementia in five European countries: a comparison of family members of people with Alzheimer's dementia versus non-family members. *Frontiers in Genetics*, 14, 1305107.

[38] Young AT, Amara D, Bhattacharya A, et al. (2021). "Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review." *The Lancet Digital Health*, *3*(9), e599-e611.

[39] Nichol BA, Hurlbert AC, Read JC. (2022). "Predicting attitudes towards screening for neurodegenerative diseases using OCT and artificial intelligence: Findings from a literature review." *Journal of Public Health Research*, *11*(4), 22799036221127627.

[40] Yakar D, Ongena YP, Kwee TC, et al. (2022). "Do people favor artificial intelligence over physicians? A survey among the general population and their view on artificial intelligence in medicine." *Value in Health*, *25*(3), 374-381.

[41] Weimer DL, Sager MA. (2009). "Early identification and treatment of Alzheimer's disease: social and fiscal outcomes". *Alzheimer's & Dementia*, *5*(3), 215-226.

[42] Clark MD, Determann D, Petrou S, et al. (2014). "Discrete choice experiments in health economics: a review of the literature." *Pharmacoeconomics*, *32*, 883-902.

[43] de Bekker-Grob EW, Ryan M, Gerard K. (2012). “Discrete choice experiments in health economics: a review of the literature.” *Health economics*, *21*(2), 145-172.

[44] Morrell L, Buchanan J, Rees S, et al. (2021). What aspects of illness influence public preferences for healthcare priority setting? A discrete choice experiment in the UK. *Pharmacoeconomics*, 39, 1443-1454.

[45] Jiang S, Ren R, Gu Y, et al. (2023). Patient preferences in targeted pharmacotherapy for cancers: a systematic review of discrete choice experiments. *Pharmacoeconomics*, 41(1), 43-57.

[46] Lo SH, Lawrence C, Martí Y, et al. (2022). Patient and caregiver treatment preferences in type 2 and non-ambulatory type 3 spinal muscular atrophy: a discrete choice experiment survey in five European countries. *Pharmacoeconomics*, 40(Suppl 1), 103-115.

[47] Poulos C, Gebben D, Peay H, et al. (2023). « PCR199 Benefit-Risk Preferences of Patients for the Use of Artificial Intelligence and Ultrasound Imaging in Different Settings in Echocardiography.” *Value in Health*, *26*(6), S349.

[48] Frederick S. (2005). “Cognitive reflection and decision making.” *Journal of Economic perspectives*, *19*(4), 25-42.

[49] Krinsky I, Robb AL. (1986) “On Approximating the Statistical Properties of Elasticities”. *The Review of Economics and Statistics*, 68, 715–719

[50] Hauber AB, González JM, Groothuis-Oudshoorn CG, et al. (2016). Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR conjoint analysis good research practices task force. *Value in health*, 19(4), 300-315.

[51] Picone G, Sloan F, Taylor D. (2004). “Effects of risk and time preference and expected longevity on demand for medical tests”. *Journal of Risk and Uncertainty*, *28*, 39-53.

[52] Cations M, Radisic G, Crotty M, et al. (2018). "What does the general public understand about prevention and treatment of dementia? A systematic review of population-based surveys." *PLoS One*, 13(4), e0196085.

[53] Jennekens FG. (2014). A short history of the notion of neurodegenerative disease. *Journal of the History of the Neurosciences*, 23(1), 85-94.

[54] Robb KA, Simon AE, Miles A, et al. (2014). "Public perceptions of cancer: a qualitative study of the balance of positive and negative beliefs". *BMJ open*, 4(7), e005434.

[55] Fritsch SJ, Blankenheim A, Wahl A, et al. (2022) "Attitudes and perception of artificial intelligence in healthcare: A cross-sectional survey among patients." *Digital Health*. 8;8:20552076221116772. doi: 10.1177/20552076221116772.

[56] Richardson JP, Smith C, Curtis S, et al. (2021). "Patient apprehensions about the use of artificial intelligence in healthcare." *NPJ digital medicine*, 4(1), 140.

[57] Mirkin S, Albensi BC. (2023). "Should artificial intelligence be used in conjunction with Neuroimaging in the diagnosis of Alzheimer's disease?" *Frontiers in Aging Neuroscience*, 15, 1094233.

[58] Murdoch B. (2021). "Privacy and artificial intelligence: challenges for protecting health information in a new era." *BMC Medical Ethics*, 22(1), 1-5.

[59] Johnson FR, Lancsar E, Marshall D, et al. (2013). Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value in health*, 16(1), 3-13.

Table 1: DCE attributes.

Attributes	**Possible values**
(1) Type	Bio vs AI vs AI+
(2) Reading	Doctor vs NHI vs Self vs Company
(3) Sensitivity	60% vs 70% vs 95%
(4) Specificity	60% vs 70% vs 95%
(5) Price	0€ vs 20€ vs 90€

Table 2- Regression result of attributes on utility

	Direct effect	Interaction with Curability
Test Type (Ref= Biological)		
AI	-0.480*** (0.120)	-0.073 (0.122)
AI+	-0.874*** (0.162)	-0.0025 (0.120)
Test reading (Ref= Family Doctor)		
Self	-0.581*** (0.114)	-0.267· (0.155)
NHI	-0.545*** (0.111)	-0.238 (0.147)
Private Company	-1.186*** (0.112)	-0.358* (0.164)
Sensitivity	0.031*** (0.003)	0.002 (0.004)
Specificity	0.038*** (0.003)	-0.001 (0.004)
Price	-0.014*** (0.001)	-0.001 (0.001)
N	10,170	
Likelihood Ratio	2214***	
Wald	1316***	

Note: ·: p<0.1; *: p<0.05; **: p<0.01; ***: p<0.001. *Standard errors in parenthesis.*

Appendix A1. DCE survey development

Survey design forms the backbone of any DCE (Bridges et al (2011), Johnson et al (2013)). This appendix gives detailed information on the development of the DCE survey that we used in this study.

A.1.1 The choice task

Each respondent was presented with successive scenarios featuring a varying set of attributes for the test. To build the set of attributes, we relied on a prior literature review: Huang et al. (2014)[2]notably emphasized 'accuracy', and 'anonymity'. Neumann et al. (2001)[3] suggested that people value testing for personal and financial reasons ('price'), but also express the need to counsel tested persons about accuracy and implications of test information ('doctor reading'). We refined these attributes, adding our main point 'AI versus saliva test'. We end up with 5 attributes: (1) the type of the test, (2) the reading of test results, (3) tests' sensitivity, (4) tests' specificity, (5) tests' price.

Note that the curability of the disease which is an important driver of testing decision[4] (Myers et al. 1990; Weller et al. 1995) has been included as a general context opening the questionnaire. We preferred to "prefix" the nature of the disease by respondent, to be able to focus on the characteristics of the test.

A.1.2 Scenarios combination

The full factorial design was generated and provided 324 possible combinations of all attribute levels (4 attributes at 3 levels and 1 attribute at 4 levels, 3^4 * 4^1=324). Each one corresponds to one of two possible options in a choice scenario. However, this number is too high to consider working with the complete design. We have chosen to reduce the number of possible combinations by considering a D-efficient fractional design that takes into account main effects and first order interactions. This leads to 70 possible combinations of attribute levels which, by random pair-wise matching, generate 35 scenarios of choice between two options. The D-efficiency of the fractional design generated is 82%.

For each person interviewed, 5 scenarios were drawn at random, from the 35 possible.

A.1.3 Forced choice methodology

We opted for the 'forced choice' for two reasons: i) according to the literature, this method seems acceptable, and even preferable (see references below); ii) the task was hard cognitively for respondent (they have to compare options with several attributes and figures, like sensitivity and specificity), and was not incentivized, although other parts of the questionnaire were gamified and incentivized; this would expose to a high risk of numerous "do not know" of this online study.

---

[2] Huang, M. Y., Huston, S. A., & Perri, M. (2014). Consumer preferences for the predictive genetic test for Alzheimer disease. *Journal of genetic counseling*, 23(2), 172-178

[3] Neumann, P. J., Hammitt, J. K., Mueller, C., Fillit, H. M., Hill, J., Tetteh, N. A., & Kosik, K. S. (2001). Public attitudes about genetic testing for Alzheimer's disease. *Health Affairs*, 20(5), 252-264.

[4] Myers, R. E., Trock, B. J., Lerman, C., Wolf, T., Ross, E., & Engstrom, P. F. (1990). Adherence to colorectal cancer screening in an HMO population. *Preventive medicine*, *19*(5), 502-514.

Weller, D. P., Owen, N., Hiller, J. E., Willson, K., & Wilson, D. (1995). Colorectal cancer and its prevention: prevalence of beliefs, attitudes, intentions and behaviour. *Australian Journal of Public Health*, *19*(1), 19-23.

In the literature[5], many papers put forward the 'forced choice'. Jonker et al. (2021) stated that 'a dual-response option format' can provide (potentially useful) preference information for respondents who otherwise consistently choose the opt-out option. Unfortunately, there is still a lack of evidence about the quality of these additional choice data. Moreover, the different framing of the opt-out options could induce slightly different choice behavior, resulting in different preferences assessment and different adoption-rates estimates. Similarly, Veldwijk et al. reported that 'including an opt-out reduces efficiency with respect to power' (2013), and that 'since respondents seem to learn from answering forced choice tasks, a dual response design might result in higher data quality compared to offering a direct opt-out option (2014). In their study, Determan et al. (2019) found that 'the use of different unforced choice formats affects marginal utilities and welfare estimates and hence the conclusions that will be drawn from the DCE to inform health care decision making'. Last, Campbell & Erdem (2019) showed that 'mis-specifying the opt-out effect has repercussions for marginal willingness to pay estimation and the forecasting of market shares'.

[5] Jonker, M. F., de Bekker-Grob, E. W., Veldwijk, J., Goossens, L. M. A., & Rutten-Van Molken, M. (2021). Improved External Validity of DCE Uptake Predictions Based on a Dual-Response None Option Format?. The Patient, 14(6), 867-867.

Veldwijk, J., Lambooij, M. S., de Bekker-Grob, E., Smit, H. A., & de Wit, G. A. (2013). The effect of including an opt-out option in discrete choice experiments. Value in Health, 16(3), A46.

Veldwijk, J., Lambooij, M. S., de Bekker-Grob, E. W., Smit, H. A., & de Wit, G. A. (2014). The effect of including an opt-out option in discrete choice experiments. PloS one, 9(11), e111805.

Determann, D., Gyrd-Hansen, D., de Wit, G. A., de Bekker-Grob, E. W., Steyerberg, E. W., Lambooij, M. S., & Bjørnskov Pedersen, L. (2019). Designing unforced choice experiments to inform health care decision making: implications of using opt-out, neither, or status quo alternatives in discrete choice experiments. Medical Decision Making, 39(6), 681-692.

Campbell, D., & Erdem, S. (2019). Including opt-out options in discrete choice experiments: issues to consider. The Patient-Patient-Centered Outcomes Research, 12, 1-14.

Appendix A3. Participants sociodemographics

Table A1. Socio-demographics of the sample

Variable	Distribution or Mean (Sd)
Gender	48.48% female 51.43% male 0.10% other
Age	M = 52.5 years, SD = 16.3
Income	M = 3.759€, SD = 2.312€
Education	High school or lower : 25.27% Undergraduate level : 40.41% Graduate level or above : 34.32%

Appendix A4. Alternative econometric specification

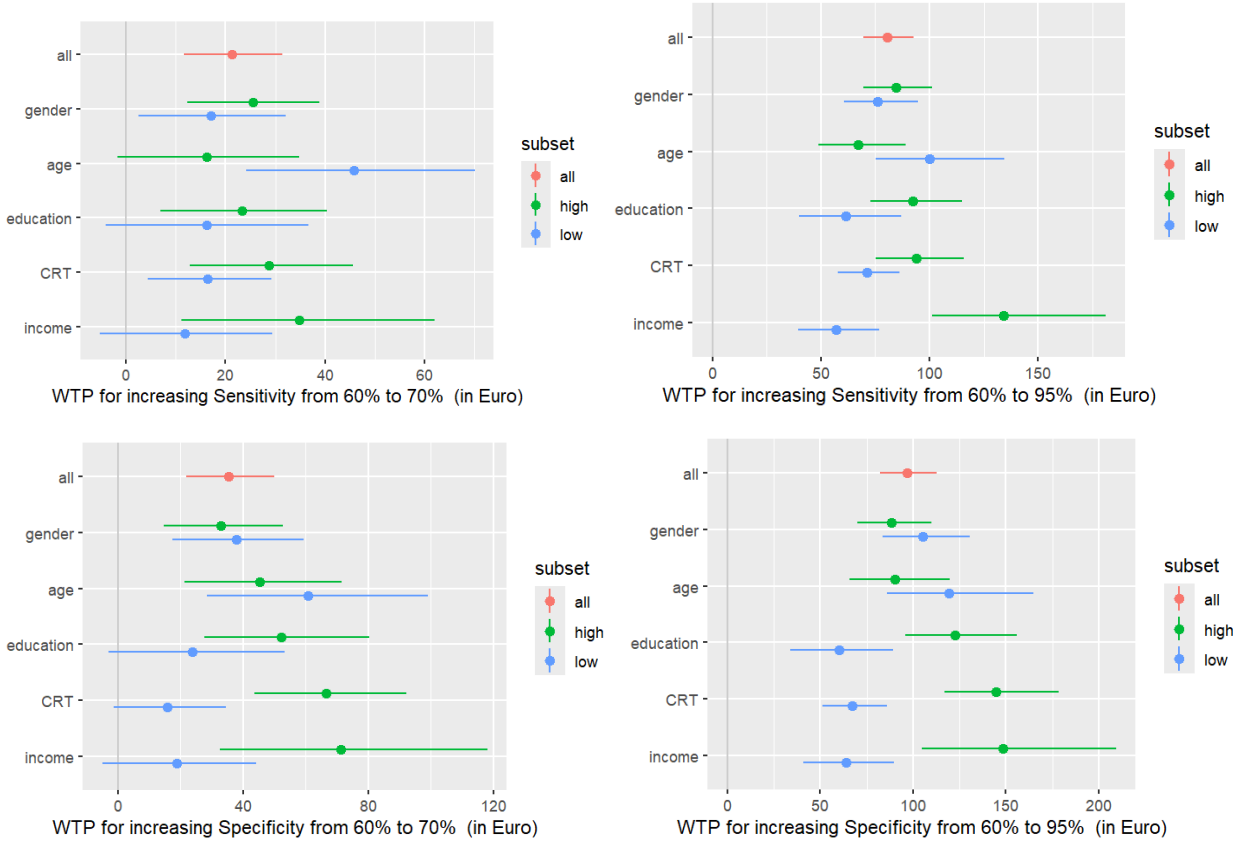
We tested an alternative econometric specification which does not assume linearity in the effect of the Price, Sensitivity and Specificity attributes. The alternative model lead to qualitatively similar results without significantly outperforming the model presented in the main paper (Likelihood Ratio - test: $\hat{\chi}(6) = 7.92, p = 0.244$).

Table A2. Alternative model without linearity assumption in the Price, Sensitivity and Specificity attributes

	Direct effect	Interaction with Curability
Test Type (Ref= Biological)		
AI	-0.493*** (0.127)	-0. 002 (0.135)
AI+	-0.854*** (0.187)	0.014 (0.152)
Test reading (Ref= Family Doctor)		
Self	-0.588*** (0.120)	-0.188 (0.166)
NHI	-0.614*** (0.130)	-0.086 (0.161)
Private Company	-1.200*** (0.115)	-0.293 · (0.169)
Sensitivity (Ref = 60%)		
70%	0.259* (0.102)	0.081 (0.140)
95%	1.158*** (0.106)	0.013 (0.147)
Specificity (Ref=60%)		
70%	0.433** (0.142)	0.163 (0.196)
95%	1.321*** (0.133)	0.099 (0.186)
Price (Ref = 0€)		
20€	-0.459*** (0.101)	0.225 (0.143)

90€	-1.338*** (0.100)	0.054 (0.139)
N	10,170	
Likelihood Ratio	2194***	
Wald	1287***	

As a robustness check for the subpopulation analysis, we also computed WTP for increasing Sensitivity and Specificity from 60% to 70% and 95%, with a model that assumes linearity in Price and non-linearity in Specificity and Sensitivity. The following figures present the estimated WTP for increasing Sensitivity and Specificity from 60% to 70% and 95%, for subsamples according to gender, age, education, cognitive abilities and income.



We obtain qualitatively similar results. We observe no difference in terms of willingness to pay for the Reading and Type attributes and observe the same group differences in the willingness to pay for increasing sensitivity and specificity. More precisely, individuals with higher income are willing to pay more for increasing sensitivity from 60% to 95% (M = 133.94€, IC = [101.01; 181.04]) if compared with individuals with lower income (M = 56.78€, IC = [39.30; 76.72]). The willingness to pay for increasing specificity from 60% to 95% is also higher for individuals with

higher cognitive abilities (M = 144.62€, IC = [117.20; 178.28]), income (M = 148.31€, IC = [104.79; 209.35]), and education (M = 122.84€, IC = [95.98; 155.71]), if compared with individuals with lower cognitive abilities (M = 67.42€, IC = [51.22; 85.84]), income (M = 64.14€, IC = [41.21; 89.81]), and education (M = 60.07€, IC = [34.04; 89.45]).