Does improving diagnostic accuracy increase artificial intelligence adoption? A public acceptance survey using randomized scenarios of diagnostic methods

Yulin Hswen, Ismaël Rafaï, Antoine Lacombe, Bérengère Davin-Casalena, Dimitri Dubois, Thierry Blayac, Bruno Ventelou

ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

BRIEF REPORT

Does improving diagnostic accuracy increase artificial intelligence adoption? A public acceptance survey using randomized scenarios of diagnostic methods

Yulin Hswen[1,2*], Ismaël Rafaï[2], Antoine Lacombe[2], Bérengère Davin-Casalena[3], Dimitri Dubois[4], Thierry Blayac[4], and Bruno Ventelou[2]

[1]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, United States of America

[2]Aix-Marseille Univiversity, CNRS, AMSE, Marseille, France

[3]Observatoire Régional de la Santé, Provence-Alpes-Côte d'Azur, France

[4]CEE-M, Univ. Montpellier, CNRS, INRAe, Institut Agro, Montpellier, France

(This article belongs to the *Special Issue: Artificial intelligence for diagnosing brain diseases*)

Abstract

This study examines the acceptance of artificial intelligence (AI)-based diagnostic alternatives compared to traditional biological testing through a randomized scenario experiment in the domain of neurodegenerative diseases (NDs). A total of 3225 pairwise choices of ND risk-prediction tools were offered to participants, with 1482 choices comparing AI with the biological saliva test and 1743 comparing AI+ with the saliva test (with AI+ using digital consumer data, in addition to electronic medical data). Overall, only 36.68% of responses showed preferences for AI/AI+ alternatives. Stratified by AI sensitivity levels, acceptance rates for AI/AI+ were 35.04% at 60% sensitivity and 31.63% at 70% sensitivity, and increased markedly to 48.68% at 95% sensitivity (p <0.01). Similarly, acceptance rates by specificity were 29.68%, 28.18%, and 44.24% at 60%, 70%, and 95% specificity, respectively (P < 0.01). Notably, AI consistently garnered higher acceptance rates (45.82%) than AI+ (28.92%) at comparable sensitivity and specificity levels, except at 60% sensitivity, where no significant difference was observed. These results highlight the nuanced preferences for AI diagnostics, with higher sensitivity and specificity significantly driving acceptance of AI diagnostics.

Keywords: Artificial intelligence; AI diagnostics; Neurodegenerative diseases; Machine learning

***Corresponding author:**
Yulin Hswen
(yulin.hswen@ucsf.edu)

Citation: Hswen Y, Rafaï I, Lacombe A, *et al.* Does improving diagnostic accuracy increase artificial intelligence adoption? A public acceptance survey using randomized scenarios of diagnostic methods. *Artif Intell Health.* doi: 10.36922/aih.3561

Received: May 2, 2024

Accepted: September 27, 2024

Published Online: October 18, 2024

1. Background

The integration of artificial intelligence (AI) into health care brings the promise of revolutionizing diagnostic and prognostic capabilities, offering more precise, data-driven insights that can enhance patient outcomes.[1,2] By harnessing AI's ability to analyze large datasets, including electronic health records (EHRs) and digital data concerning consumer behaviors, health-care systems can potentially improve the early detection

and prediction of disease risks with greater accuracy and timeliness. This combination of clinical and behavioral data could enable more personalized diagnostic models.

However, despite the substantial potential AI offers in transforming health care, there remains significant hesitation toward its widespread adoption, particularly in AI-assisted diagnostics. Much of this reluctance stems from concerns about patient privacy and the risks associated with data surveillance.[3] Health-care professionals and patients alike worry that the use of AI in clinical settings could lead to breaches of sensitive information, unauthorized data access, and misuse of personal health data, which are the main factors undermining users' trust in AI-driven systems. This aversion to AI, fueled by privacy concerns, continues to be a major obstacle to its full acceptance in the health-care field.

By examining the resistance to AI, previous research found that generative chatbot AI faces a hesitant adoption.[4] A review of 7912 articles aimed at identifying predictors of AI adoption revealed that perceived usefulness, performance expectancy, trust, and effort were key factors influencing the willingness to use AI in health care.[4] The review also emphasized that no amount of AI could fully replace the value of human interaction or ensure cultural sensitivity. In another study related to AI use in health care,[5] this reluctance was shown to be more pronounced among individuals with limited proficiency in Internet or computer technologies. A noted source of concern stems from the uncertainty surrounding the data sources that power these AI models, leading to skepticism about the reliability and accuracy of the health information they generate. In addition, users express unease over the lack of transparency in how these models operate and the inherent complexity of AI systems. These factors contribute to fears of miscommunication, misinterpretation of health symptoms, and the potential for inaccurate diagnoses. In another related survey, trust in AI adoption was found to be closely linked to regulatory oversight, with performance and communication also playing critical roles in users' willingness to embrace AI applications in health care.[6]

A survey conducted in Sweden showed that only 20% of health-care professionals used AI-based systems in their work, with "trust" emerging as the most critical factor in their willingness to adopt these technologies.[7] A review of 42 studies examining health-care professionals' acceptance of AI revealed widespread concerns, particularly regarding AI's potential for errors, sensitivity, and timely access. In addition, the perceived loss of professional autonomy and challenges in integrating AI into existing clinical workflows were consistently identified as a significant barrier to adoption.[8] These findings highlight that healthcare workers, such as patients, exhibit a degree of AI adoption hesitancy, particularly in its utilization in diagnostics.

In contrast to previous studies, which primarily rely on surveys, this study aims to broaden the existing literature on AI adoption hesitancy by testing AI adoption through randomized scenario-based experiments. This approach allows for a more nuanced understanding of how individuals respond to AI in varied controlled contexts.

2. Methods

This study evaluated the public acceptability of AI-based diagnostic tools and the accuracy trade-offs required to integrate EHRs and digital data in the domain of neurodegenerative diseases (NDs). A survey was conducted on a representative sample of the French adult population (n = 1017) using a quota non-probability sampling method (quotas were on age, gender, socio-professional status, and living area). This collection of data was part of the larger Discrete Choice Experiment[9-11] aiming at unveiling the trade-offs surrounding the decision-making by individuals about neurodegenerative testing. Before agreeing with study participation, all subjects were given comprehensive information regarding the study's purpose, procedures, potential risks, and benefits. The study protocol was reviewed and approved by the Ethics Committee of Aix-Marseille University (approval number: 2022-10-20-009). Written consent was obtained from each of the subjects to participate in this study.

The 1017 participants were exposed to a set of alternative scenarios of testing methods to predict the hypothetical 10-year risk of developing an ND that affects an average of 7% of the population after the age of 65.[12] Through the pool institut ViaVoice, participants were confidentially randomized to scenarios depicting various levels of AI-based diagnostic integration and non-AI traditional laboratory saliva test. The researchers were blinded to participants' identities. The three scenarios of tests included: (1) non-AI diagnostics using a laboratory test with a salivary sample, (2) AI diagnostics incorporating EHRs, defined as "AI," and (3) AI diagnostics incorporating EHR and digital consumer data from mobile devices, thereafter, defined as "AI+." To assess the impact of diagnostic accuracy on participants' preferences, the attributes of sensitivity (true positive rate) at 60%, 70%, or 95%, and 1-specificity (false positive rate) at 5%, 30%, or 40% were also varied. An example of the randomized scenario is shown in Figure 1.

3. Statistical analysis

Of the 5085 scenarios randomly proposed, we selected the pairs (3225) that display a comparison between AI (or AI+)

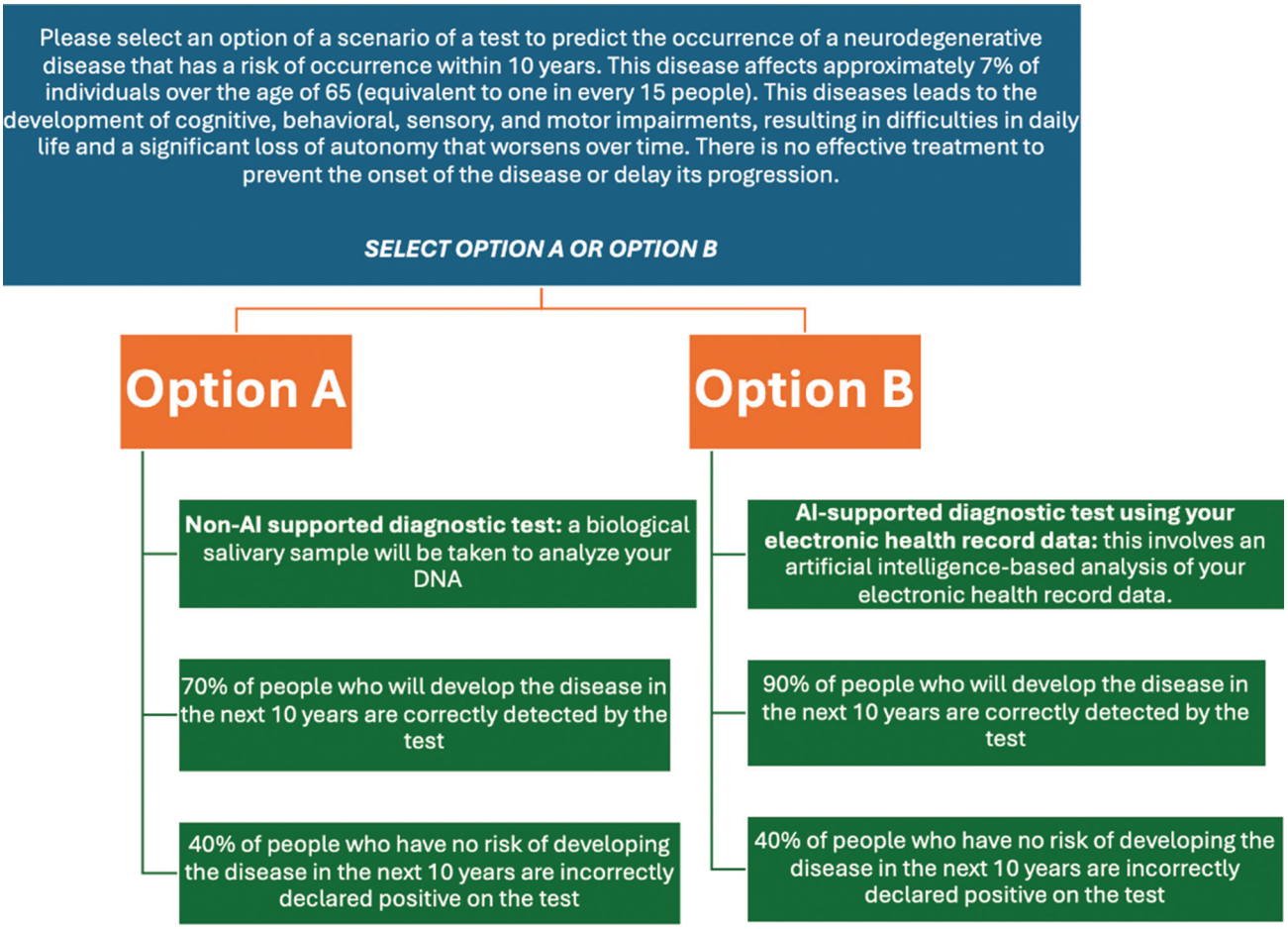

Figure 1. Example of a randomized scenario

and saliva test as the benchmark for further analysis. Then, descriptive analyses were conducted to compare the proportions of acceptance of the AI and/or AI+ option with those of the saliva test option. The differences in acceptance rates between the AI-versus-saliva-test option and the AI+-versus-saliva-test option were analyzed using pairwise z-tests, whereas the differences in proportions between three levels of sensitivity and specificity per type of pairs offered (*i.e.*, AI vs. saliva and AI+ vs. saliva) were compared using Chi-squared tests.[13,14]

4. Results

Contingency tables describing proportions of agreement between the AI and/or AI+ alternative with the saliva test alternative are presented in Table 1. From the 3225 pairs of AI/AI+ against saliva tests that were offered, 1482 are linked to a choice of AI test versus saliva test, whereas 1743 are associated with a choice of AI+ test versus saliva test. Only 36.68% of the answers were pro-AI/AI+ (45.82% for the AI-vs.-saliva-test pairs and 28.92% for the AI+-vs.-saliva-test pairs; proportions were significantly different at a 1% threshold). Figure 2 shows the proportion of AI adoption (vs. saliva test) across different levels of accuracy. Upon stratifying the answers by sensitivity level, we found that 35.04% of the answers were in favor of AI or AI+ when AI's sensitivity was 60% and 31.63% when AI's sensitivity was 70%, and this proportion increased to 48.68% when AI's sensitivity was 95% (Chi-squared test showed significant difference at a 1% threshold). With respect to specificity, 29.68%, 28.18%, and 44.24% of the answers favored AI/AI+ test over saliva test when specificity levels were 60%, 70%, and 95%, respectively (Chi-squared test showed significant differences at a 1% threshold). Finally, when we compared AI-versus-saliva-test option and AI+-versus-saliva-test option per sensitivity or specificity levels, we found significantly higher acceptance rates in the AI group than in the AI+ group from the same sensitivity or specificity level (except when sensitivity is 60%, where we found no significant differences between rates of acceptance when AI-vs.-saliva-test or AI+-vs.-saliva-test options were offered).

Table 1. The proportion of individuals choosing the AI and/or AI+ alternative over the saliva test per sensitivity and specificity levels

	AI versus saliva test	AI+ versus saliva test	AI or AI+ versus saliva test
n	1482	1743	3225
Proportion of yes to AI/AI+	$45.82\%_{+++}$	$28.92\%_{+++}$	36.68%
Sensitivity			
Proportion of yes to AI/AI+ when sensitivity=60%	36.95%***	32.77%**	35.04%***
Proportion of yes to AI/AI+ when sensitivity=70%	$44.33\%^{***}_{+++}$	$26.43\%^{**}_{+++}$	31.63%***
Proportion of yes to AI/AI+ when sensitivity=95%	$59.39\%^{(***)}_{(+++)}$	$30.04\%^{(**)}_{(+++)}$	48.68%***
Specificity			
Proportion of yes to AI/AI+ when specificity=60%	$41.24\%_{+++}$	$18.2\%^{(***)}_{(+++)}$	29.68%***
Proportion of yes to AI/AI+ when specificity=70%	$47.67\%_{+++}$	$12.73\%^{(***)}_{(+++)}$	28.18%***
Proportion of yes to AI/AI+ when specificity=95%	$47.04\%_{+++}$	$41.94\%^{(***)}_{(++)}$	44.24%***

Notes: (1) Chi-square statistical test of difference of acceptance rate per type of scenario offered (AI vs. saliva test or AI+ vs. saliva test): *P<0.1, **P<0.05, and ***P<0.01; (2) Pairwise z-test of difference in proportions per sensitivity or specificity level (60%, 70%, or 95%) between AI-versus-saliva-test and AI+-versus-saliva-test: ^{+}P<0.1, ^{++}P<0.05, ^{+++}P<0.01.
Abbreviations: AI: Artificial intelligence using electronic health records (EHRs) data; AI+: Artificial intelligence using electronic health records (EHRs) data and digital consumer data.

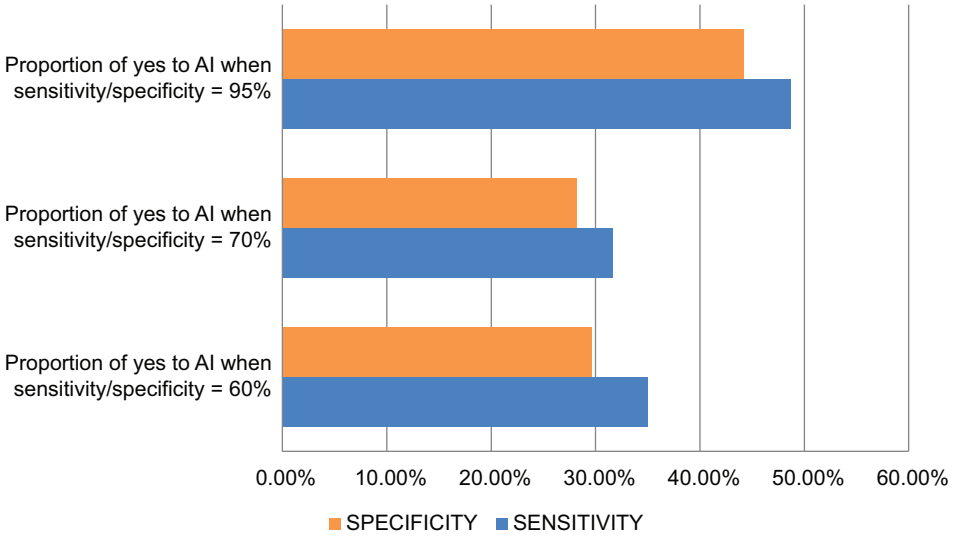

Figure 2. The proportion of artificial intelligence adoption (vs. saliva test) across different levels of accuracy

5. Discussion

The findings of this study provide valuable insights into public acceptance of AI-based diagnostic alternatives compared to conventional saliva tests. By analyzing 3225 pairwise choices, we observed that only 36.68% of participants preferred AI/AI+ alternatives over traditional saliva tests, specifically with a significant preference for AI (45.82%) compared to AI+ (28.92%).

The results strongly showed the influence of AI sensitivity (how well the AI test can identify true positives) on acceptance rates. As AI sensitivity increased from 60% to 95%, support for AI/AI+ diagnostics rose considerably, reaching 48.68% at 95% sensitivity. A similar pattern emerged with specificity (how well an AI test can identify true negatives), where acceptance rates increased from 29.68% at 60% specificity to 44.24% at 95% specificity.

These findings underscore the importance of enhancing diagnostic accuracy in fostering public trust in AI-based tools. Public acceptance of AI diagnostics is closely tied to accuracy levels, and these results suggest that AI tools must meet or exceed a 95% performance threshold to achieve meaningful levels of AI acceptance.

In addition, AI consistently outperformed AI+ across all levels of sensitivity, with the exception of 60% sensitivity, where no significant difference in preference between AI and AI+ was found. This outcome may indicate a hesitancy toward the integration of digital consumer data such as those from mobile phones versus EHR data alone. However, when accuracy approaches a sensitivity level of 95%, the public appears more willing to consider the use of these digital consumer data resources, reflecting a trust deficit that can be mitigated by increased diagnostic performance.

6. Study limitations

It is important to emphasize that, to minimize the biases of physical invasiveness[15] while striving to level the playing field in comparison to AI testing methods, we deliberately chose a salivary test for this study. As a result, our estimates of the public's preference for biological tests may be in fact lower if AI testing was compared to more physically invasive procedures such as brain imaging, cerebrospinal fluid analysis, or blood tests.[16,17] This decision likely shaped the participants' responses, as the less invasive nature of the salivary test may have led them to favor it over more physically invasive testing methods. As a result, the reluctance toward AI diagnostics observed in this study may be less significant when compared to scenarios involving more invasive testing procedures.

Public perceptions of AI adoption are also likely to differ significantly across geographic regions, influenced by varying cultural, economic, and social factors that shape attitudes toward technology. Although previous studies have shown similar AI hesitancies, this study was conducted in France and national differences could result in diverse levels of trust, familiarity, and comfort with AI, thereby affecting how AI technologies are embraced across different nations. Consequently, this variability poses a potential limitation to the generalizability of this study's findings. Factors such as regional regulatory environments, access to technology, socioeconomic disparities, and historical experiences with digital tools could further amplify these discrepancies in AI acceptance. Therefore, our findings must be considered within the diverse global contexts where AI technologies may be implemented. This underscores the importance of future research to examine AI adoption across a broader range of geographic and cultural settings, ensuring greater applicability and relevance.

7. Contributions of this study

Unlike earlier research that has largely focused on survey-based methods, this study expands the body of knowledge on AI adoption by conducting investigations on AI acceptance through randomized, scenario-driven experiments. Using this approach, we can capture a more detailed perspective on how people react to AI in diverse and controlled situations, addressing the broader challenge of AI hesitancy and the complexity of its acceptance in real-world settings. Our findings significantly enhance the current body of research by providing empirical evidence on the threshold of diagnostic accuracy required for AI-driven technologies to achieve widespread public acceptance. By quantifying these levels of accuracy, we offer a framework for understanding the public's expectations of AI in health-care settings. This research not only underscores the importance of reliability and accuracy in AI diagnostics but also highlights the nuanced factors influencing public trust and adoption. In addition, it sheds light on how varying degrees of accuracy can shape public perceptions, offering insights for developers, policymakers, and health-care professionals aiming to bridge the gap between technological advancements and public readiness for AI integration. These insights are particularly valuable in addressing AI hesitancy and ensuring the ethical implementation of AI in health care.

8. Conclusion

Our findings carry important implications for the development and implementation of AI diagnostics in health care. Public hesitation persists as a significant barrier, especially when AI tools are perceived as lacking sufficient accuracy or integrating excessive amounts of personal data. Our results emphasize the critical need for AI developers and health-care providers to prioritize transparency, accuracy, and usability in AI diagnostic technologies. Moreover, educating the public about the potential benefits of AI diagnostics, particularly diagnostic accuracy, could further alleviate concerns and promote broader acceptance.

This study highlights the nuanced preferences of the public for AI diagnostics, with higher sensitivity and specificity acting as key drivers of acceptance. While AI holds considerable potential to transform health-care diagnostics, addressing the public's concerns about accuracy and complexity will be essential to its successful adoption.

Acknowledgments

None.

Funding

The project leading to this publication has received funding from the French government under the "France 2030" investment plan managed by the French National Research Agency (reference: ANR-17-EURE-0020) and from Excellence Initiative of Aix-Marseille University – A*MIDEX. This research also received support from the French National Research Agency (GRANT ANR-20-COVR-00 and ANR-21-JPW2-002), as well as funding from the National Institute of Health T32 grant (5T32MD015070-05).

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: All authors
Formal analysis: Bruno Ventelou, Yulin Hswen, Ismaël Rafaï, Antoine Lacombe
Investigation: Bruno Ventelou, Yulin Hswen, Ismaël Rafaï, Antoine Lacombe
Methodology: Bruno Ventelou, Yulin Hswen, Ismaël Rafaï, Antoine Lacombe
Writing–original draft: Bruno Ventelou, Yulin Hswen
Writing–review & editing: Ismaël Rafaï, Antoine Lacombe

Ethics approval and consent to participate

The study protocol was reviewed and approved by the Ethics Committee of Aix-Marseille University (approval number: 2022-10-20-009). Written consent was obtained from each of the subjects to participate in this study.

Consent for publication

Written consent was obtained from each of the subjects to publish their data and/or images.

Availability of data

Data used in the study can be obtained from the corresponding author upon reasonable request.

References

1. Reyna MA, Nsoesie EO, Clifford GD. Rethinking algorithm performance metrics for artificial intelligence in diagnostic medicine. *JAMA*. 2022;328(4):329-330.

   doi: 10.1001/jama.2022.10561

2. Aggarwal R, Sounderajah V, Martin G, *et al*. Diagnostic accuracy of deep learning in medical imaging: A systematic review and meta-analysis. *NPJ Digit Med*. 2021;4(1):65.

   doi: 10.1038/s41746-021-00438-z

3. Laux J, Wachter S, Mittelstadt B. Trustworthy artificial intelligence and the European Union AI act: On the conflation of trustworthiness and acceptability of risk. *Regul Gov*. 2024;18(1):3-32.

   doi: 10.1111/rego.12512

4. Choung H, David P, Ross A. Trust in AI and its role in the acceptance of AI technologies. *Int J Hum Comput Interact*. 2023;39(9):1727-1739.

   doi: 10.1080/10447318.2022.2050543

5. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *Digit Health*. 2019;5:2055207619871808.

   doi: 10.1177/205520761987180

6. Esmaeilzadeh P. Use of AI-based tools for healthcare purposes: A survey study from consumers' perspectives. *BMC Med Inform Decis Mak*. 2020;20:1-19.

   doi: 10.1186/s12911-020-01191-1

7. Floruss J, Vahlpahl N. *Artificial Intelligence in Healthcare: Acceptance of AI-based Support Systems by Healthcare Professionals*. Jönköping University. Master Thesis; 2020. Available from: https://www.diva-portal.org/smash/get/diva2:1433298/fulltext01.pdf [Last accessed on 2024 Oct 11].

8. Lambert SI, Madi M, Sopka S, *et al*. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. *NPJ Digit Med*. 2023;6(1):111.

   doi: 10.1038/s41746-023-00874-z

9. De Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: A review of the literature. *Health Econ*. 2012;21(2):145-172.

   doi: 10.1002/hec.1697

10. Szinay D, Cameron R, Naughton F, Whitty JA, Brown J, Jones A. Understanding uptake of digital health products: Methodology tutorial for a discrete choice experiment using the bayesian efficient design. *J Med Internet Res*. 2021;23(10):e32365.

    doi: 10.2196/32365

11. Clark MD, Determann D, Petrou S, Moro D, De Bekker-Grob EW. Discrete choice experiments in health economics: A review of the literature. *Pharmacoeconomics*. 2014;32:883-902.

    doi: 10.1007/s40273-014-0170-x

12. Alawode DO, Heslegrave AJ, Ashton NJ, *et al*. Transitioning from cerebrospinal fluid to blood tests to facilitate diagnosis and disease monitoring in Alzheimer's disease. *J Int Med*. 2021;290(3):583-601.

    doi: 10.1111/joim.13332

13. Greenwood PE, Nikulin MS. *A Guide to Chi-squared Testing*. Vol. 280. United States: John Wiley and Sons; 1996.

14. Campbell I. Chi-Squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. *Stat Med*. 2007;26(19):3661-3675.

doi: 10.1002/sim.2832

15. Freeman D, Lambe S, Yu LM, *et al*. Injection fears and COVID-19 vaccine hesitancy. *Psychol Med*. 2023;53(4):1185-1195.

doi: 10.1017/S0033291721002609

16. McLenon J, Rogers MA. The fear of needles: A systematic review and meta-analysis. *J Adv Nurs*. 2019;75(1):30-42.

doi: 10.1111/jan.13818

17. Von Wedel P, Hagist C. Physicians' preferences and willingness to pay for artificial intelligence-based assistance tools: A discrete choice experiment among german radiologists. *BMC Health Serv Res*. 2022;22(1):398.

doi: 10.1186/s12913-022-07769-x