

Article

Not peer-reviewed version

Enhancing Biomedical Research with ResearchKit: Digitization of IPAQ and MMSE for Sedentary Behavior and Cognitive Impairment Analysis

Juan David López-Regalado , [Gerardo Marx Chávez-Campos](#) * , [Adriana del Carmen Téllez-Anguiano](#) , [Antony Morales-Cervantes](#)

Posted Date: 13 March 2025

doi: 10.20944/preprints202503.0932.v1

Keywords: ResearchKit; Biomarkers; Machine Learning; Clinical Trials



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Enhancing Biomedical Research with ResearchKit: Digitization of IPAQ and MMSE for Sedentary Behavior and Cognitive Impairment Analysis

Juan David López-Regalado, Gerardo Marx Chávez-Campos *,
Adriana del Carmen Téllez-Anguiano and Antony Morales-Cervantes

TecNM - Instituto Tecnológico de Morelia: División de Estudios de Posgrado e Investigación, Maestría en Ciencias en Ingeniería Electrónica, Michoacán, México.

* Correspondence: gmarx_cc@itmorelia.edu.mx

Abstract: Medical questionnaires and forms play a crucial role in diagnosing diseases and gathering essential patient information. Traditionally, these assessments are conducted through face-to-face interactions or phone calls, which can be time-consuming, costly, and inefficient. To address these challenges, digital adaptations of medical questionnaires have been developed, enabling faster and more accessible data collection across various platforms and applications. ResearchKit, an open-source framework introduced by Apple, allows researchers and developers to create robust mobile applications for medical research. By leveraging this technology, large-scale data collection can be conducted efficiently, facilitating real-time analysis while reaching a broader population. The availability of extensive datasets is essential for computational techniques such as Machine Learning, which require significant amounts of data for classification, pattern recognition, and predictive modeling in healthcare. This paper focuses on the digitization of medical questionnaires, specifically the International Physical Activity Questionnaire (IPAQ) and the Mini-Mental State Examination (MMSE), using ResearchKit. The transition from paper-based to digital forms significantly improves the efficiency of medical assessments, allowing healthcare professionals to analyze results remotely and optimize patient diagnosis and treatment plans. The findings emphasize the impact of digital health solutions in advancing medical research, reducing costs, and enhancing the accuracy and scalability of data-driven healthcare applications.

Keywords: ResearchKit; biomarkers; machine learning; clinical trials

1. Introduction

Clinical research plays a vital role in understanding diseases, developing treatments, and improving public health [1,2]. However, traditional data collection methods, such as face-to-face interviews, postal surveys, or telephone questionnaires, are time-consuming and costly [3]. Digital tools have transformed biomedical research by enabling real-time, large-scale data acquisition at lower costs [4].

Apple's ResearchKit, an open-source framework, facilitates mobile-based medical research by leveraging the iPhone's capabilities to collect biomarkers and administer digital surveys [5,6]. This innovation streamlines data collection, improving accessibility and efficiency while addressing limitations in traditional clinical research [7].

Meanwhile, Machine Learning (ML) offers advanced data analysis techniques for pattern recognition, classification, and predictive modeling, significantly enhancing biomedical informatics [8]. Although this study does not yet apply ML methods, it lays the groundwork for future ML applications by utilizing ResearchKit to gather extensive data for biomedical research [9].

This study focuses on the digital implementation of the International Physical Activity Questionnaire (IPAQ) and the Mini-Mental State Examination (MMSE) using ResearchKit—two widely used tools for assessing physical activity and cognitive function [10,11]. The broader research effort aims to

collect IPAQ and MMSE data alongside biomarker measurements to explore their correlations. Ultimately, the goal is to implement ML techniques to identify key parameters associated with sedentary behavior and cognitive impairment.

The **IPAQ** is a validated and reliable questionnaire for measuring physical activity levels in populations aged 15 to 69 years [10]. It assesses physical activity across various domains, including work, transportation, household activities, and leisure time, categorizing activity levels based on metabolic equivalents (METs) consumed during those activities [12].

Conversely, the **MMSE** is a widely used pen-and-paper test that evaluates cognitive function, with a maximum score of 30 points [13]. It assesses multiple cognitive domains, including orientation, concentration, attention, verbal memory, naming, and visuospatial skills. Its predictive utility can be determined through predefined thresholds or statistical methods such as logistic regression, with optimal cut-off points established based on sensitivity, specificity, and other relevant metrics [13].

Therefore, this research aims to enhance data collection using ResearchKit to enable more effective health assessments, especially in studying the relationship between sedentary behavior and cognitive impairment. The findings emphasize the potential of mobile health applications to advance public health initiatives and support personalized medicine.

2. Methods

A structured methodology is implemented to develop a robust digital platform for administering IPAQ and MMSE surveys using ResearchKit, encompassing application design, development, and an initial data collection phase to assess functionality and usability.

The following sections will cover the application design for general clinical trials and their implementation, including a basic consent survey, IPAQ and MMSE implementation, and examples of the resulting data for the clinical trial participant.

The methodology is divided into two sections: design and implementation. The design section outlines the application's key components and their relevance to this context, while the implementation section details its structure, functionality, and development process.

2.1. Application Design

The mobile application uses Swift language and Xcode as IDE, integrating ResearchKit as the primary framework for collecting participants' responses. The application follows a modular structure, see Figure 1, with the following key components.

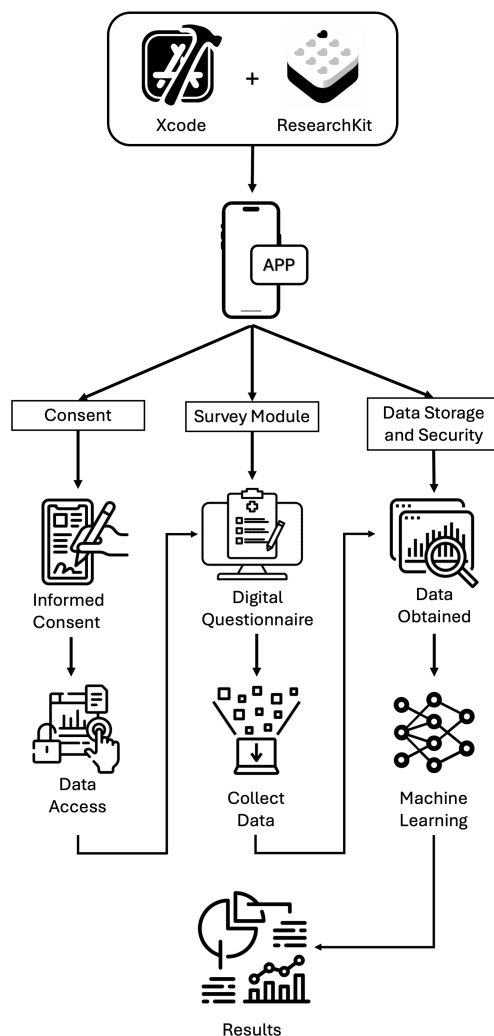


Figure 1. Diagram of development of a biomedical application.

2.1.1. Consent:

- **Informed Consent:** Ensures participants understand the study objectives, procedures, risks, and benefits before providing and signing digital consent.
- **Data Access:** With the participant's signed consent, the application is permitted to access only the data from the forms or questionnaires that the user contests.

2.1.2. Survey Module:

- **Digital Questionnaire:** The IPAQ and MMSE questionnaires are digitized using ResearchKit's pre-built survey and question functionalities to collect participant data efficiently.
- **Collect Data:** All responses obtained from the questionnaires serve as valuable study data, making it essential to store them securely for further analysis.

ResearchKit has functions for doing surveys and has many different formats to answer, existing two types of format to present surveys, Question and Form, Question is used to present only one question at a time per window and Form is used to present more questions in one window [7]. The function is needed to display the different questions and decide which answer format to use. Some of the most commonly used answer formats in this paper are:

- **ORKValuePickerAnswerFormat:** represents an answer format that lets participants use a value picker to choose from a fixed set of text choices [14].

- ORKTextAnswerFormat: represents the answer format for questions that collect a text response from the user [14].
- ORKTextChoiceAnswerFormat: represents an answer format that lets participants choose from a fixed set of text choices in a multiple or single choice question [14].
- ORKScaleAnswerFormat: represents an answer format that includes a slider control [14].

2.1.3. Data Storage and Security:

- Data Obtained: All collected data is securely stored and encrypted in local storage to comply with data privacy regulations. Cloud storage platforms such as Firebase or MongoDB can also enhance data management, ensuring scalability, secure access, and real-time synchronization across multiple devices.
- Machine Learning Applications: The stored data undergoes an anonymization stage to ensure participants' privacy before any processing. Identifiable information is removed or masked, aligning with data protection regulations and ethical research standards. Following anonymization, the data can be processed using advanced computational techniques, including Machine Learning.

2.1.4. Results:

Applying a Machine Learning approach is an effective method for extracting valuable insights, identifying trends, and making predictions. In this context, it allows for the evaluation of physical activity levels and cognitive impairment. However, rather than evaluating participants in isolation, ML enables the analysis of these factors in relation to other participants, which facilitates a more comprehensive understanding of their correlations and trends within the dataset.

This enables the development of models capable of identifying patterns, making predictions, and performing classifications based on the collected data, contributing to deeper insights and enhanced analysis while maintaining participant confidentiality.

2.2. Implementation

The implementation of this application begins with the integration of the ResearchKit framework into an empty Xcode project. This process requires downloading the framework from its official GitHub repository, maintained by its original developers at [ResearchKit](#).

Following a series of structured steps outlined on the site, the framework is imported into the project, enabling access to its core functionalities, including *informed consent*, *surveys*, and *active tasks*. The configuration process includes setting up dependencies, linking the framework, and customizing modules to align with the specific requirements of the IPAQ and MMSE assessments.

2.2.1. Consent:

One of the most critical stages in the application process is obtaining user consent, as it is necessary to proceed with the subsequent steps. This consent stage provides essential information about the study, including how the collected data will be used. Depending on the specific application requirements, it is structured into multiple sections, such as a welcome screen, data collection details, privacy policies, data usage, time commitment, study tasks, and other relevant aspects.

Creating a comprehensive and well-structured consent form is crucial, ensuring applicants fully understand the study and its procedures before agreeing to participate. Each section must clearly outline the necessary information, allowing users to review all aspects before making an informed decision. Figure 2 illustrates a consent review screen displaying the key components of the study that participants should read¹.

¹ The full code implementation is explained in detail at the repository: [Digitization Forms](#).

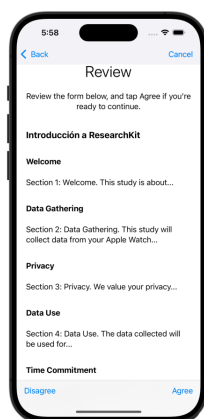


Figure 2. Consent information review screen.

Moreover, supplementary information is necessary upon granting consent, including the participant's name and, most crucially, their signature. The absence of a signature likely prevents further access, as it indicates that the participant has not formally agreed to the collection of their data. Figure 3 illustrates the interface where users must provide their signature to proceed with the study.

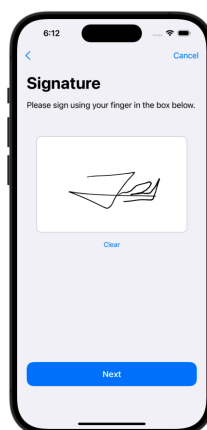


Figure 3. Signature interface screen.

After the participant correctly completes and signs the consent form, they will gain access to the second part, which consists of the digitized questionnaires.

2.2.2. Digital Questionnaire:

The second component of this study involves the digitization of questionnaires. Digitizing these questionnaires enhances efficiency by streamlining data collection and improving accessibility.

This study focuses on two widely used assessments: the International Physical Activity Questionnaire (IPAQ) and the Mini-Mental State Examination (MMSE) [15,16]. These tools are crucial in evaluating physical activity levels and cognitive impairment, facilitating more effective and standardized patient assessments.

2.2.3. IPAQ:

The IPAQ consists of seven questions assessing the frequency, duration, and intensity of physical activity (moderate and vigorous) performed over the past seven days, as well as walking and sitting time on a typical workday. The responses are stored in METs (Metabolic Equivalents of Task), a unit that estimates energy expenditure.

To calculate the total METs, each activity type is assigned a specific value—walking: 3.3 METs, moderate physical activity: 4 METs, vigorous physical activity: 8 METs-. The total MET score is

determined by multiplying the assigned MET value by the duration of the activity (in minutes per day) and the number of days per week the activity is performed [17].

Figure 4 presents an example of these questions. Due to space limitations, only the questions related to vigorous physical activity are shown. Additionally, the implementation includes windows providing explanations for the questions and welcome and thank-you screens.

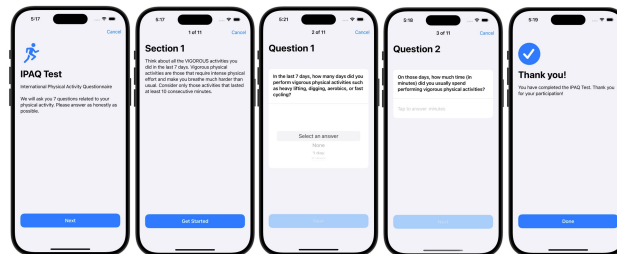


Figure 4. An overview of the IPAQ questionnaire.

2.2.4. MMSE:

The Mini-Mental State Examination (MMSE) is a widely used tool for evaluating adult cognitive impairment. It is a brief, quantitative test designed to detect early cognitive deficits and assess their severity. The MMSE evaluates various cognitive domains, including orientation, attention, calculation, memory, reading, and visuospatial skills.

The interpretation of the MMSE is based on the total score (0–30), which categorizes cognitive function into four groups: normal cognitive function (24–30 points), mild cognitive impairment (19–23 points), moderate cognitive impairment (14–18 points), and severe cognitive impairment (less than 14 points) [18,19].

Figure 5 illustrates various types of windows that ResearchKit can generate, including question screens that assist in completing the test, indications about upcoming questions, as well as welcome and thank-you views. These elements are designed to provide a user-friendly interface for participants.

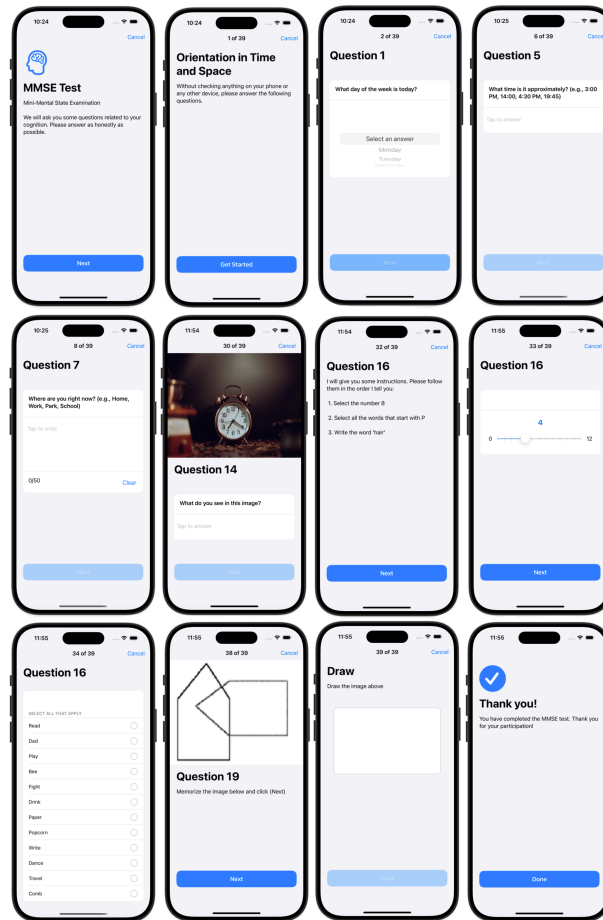


Figure 5. MMSE digital questions

2.2.5. Data Obtained:

At the end of each test, all responses can be stored on platforms such as Firebase, MongoDB, or other database services. Additionally, data can be saved locally. Participants are assured that their data remains confidential and can only be accessed by the developer or data scientist for research or diagnostic purposes.

Regarding data scientists, the collected information is stored using a predefined identifier, which is designed and programmed based on the developer's implementation preferences. The obtained data can be processed to generate statistical insights that contribute to developing new methodologies, treatments, diagnoses, and predictions, among other applications.

3. Results

All questionnaires can be digitized, although some limitations may arise due to question types and formatting variations. However, if the questionnaire primarily consists of written questions, it can be fully digitized and reproduced using ResearchKit and its available functions.

Table 1 presents a section of the IPAQ test, specifically Section Two, which asks about moderate physical activity, including the number of days and duration. This excerpt from the original test is easy to digitize since it consists solely of written questions.

Table 1. Example of section two of IPAQ

Think about all the moderate activities that you did in the last 7 days. Moderate activities refer to activities that take moderate physical effort and make you breathe somewhat harder than normal. Think only about those physical activities that you did for at least 10 minutes at a time.	
Question	Response
During the last 7 days, on how many days did you do moderate physical activities like heavy lifting, digging, aerobics, or fast bicycling?	() Days per week () No moderate physical activities (Skip to question 3)
How much time did you usually spend doing moderate physical activities on one of those days?	() Hours per day () Minutes per day () Don't know/Not sure

Figure 6 illustrates how this statement and its corresponding questions are digitized. Two types of question formats are used: a value picker answer format, similar to a caster wheel, and a numerical input answer format.

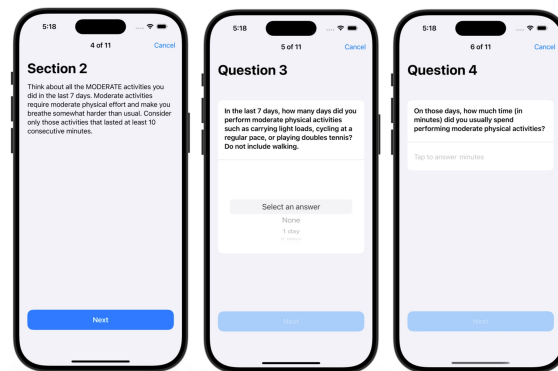


Figure 6. Example of the IPAQ, section two digitized.

As participants navigate through the questions in the two questionnaires, the application collects their responses in various formats, including boolean, numeric, text, and file inputs, among others. Each participant is assigned a Unique Identifier (UID), ensuring that their responses are securely stored under their respective identifier. Additionally, each question has a unique identifier, allowing for precise tracking and interpretation of responses.

As previously mentioned, the collected data can be stored on various platforms. Listing 1 provides an example of how questionnaire responses are stored in a JSON file, demonstrating the structured approach to data management.

This process can also be implemented using other types of databases; however, database selection is beyond the scope of this article. The focus here is on the digitization of questionnaires and using ResearchKit. Nevertheless, it is essential to understand the data format and how responses are stored along with their associated participant and question identifiers.

Listing 1: Example regarding two entries in the IPAQ questionnaire

```

1  {
2  "IPAQ": {
3    "FL0f9UxM0GUbt8HY3ictZi72Qv83": {
4      "1Vigorous1Question": "4 days",
5      "2Vigorous2Question": 4,
6      "3Moderate1Question": "No days",
7      "4Moderate2Question": 45,
8      "ModerateDays": "None",

```

```
9     "ModerateMinutes": 0,  
10     "SittingTime": 0,  
11     "VigorousDays": "None"  
12 },  
13     "illV0hcJMxgWUuhYJcq4qFgJcP02": {  
14         "1Vigorous1Question": "3 days",  
15         "2Vigorous2Question": 2,  
16         "3Moderate1Question": "5 days",  
17         "4Moderate2Question": 30,  
18         "ModerateDays": "Some",  
19         "ModerateMinutes": 15,  
20         "SittingTime": 120,  
21         "VigorousDays": "Few"  
22     }  
23 }  
24 }
```

The data download format varies depending on the developer's platform. Nevertheless, the data can be exported in multiple formats and analyzed with Machine Learning tools. Performing a comprehensive data analysis is essential as it may uncover valuable insights and relationships among participants.

4. Discussion

ResearchKit is a promising framework as it offers pre-built functionalities for key components of medical applications, including informed consent, surveys, specific tasks, and digital forms. These features facilitate the rapid collection of biomarkers (medical data) from a large number of users simultaneously, significantly enhancing efficiency in clinical research.

However, implementing these resources can be challenging, as each question type or activity requires a specific code structure. Additionally, prior knowledge of Swift language and the Xcode programming environment is essential. Xcode offers a suite of tools for the design of graphical interfaces and interactive elements, including buttons, images, sliders, and various other user interface components. This functionality enhances the versatility of application development; however, a comprehensive understanding is required to fully leverage these capabilities.

According to Samantha Meltzer-Brody, the developer of the Mom Genes Fight PPD App, "With the ResearchKit platform, we have been able to conduct the largest genetic study of postpartum depression (PPD) with a global population of participants" [20]. This highlights the potential of ResearchKit in facilitating large-scale medical studies.

Currently, many applications are being developed using ResearchKit due to its comprehensive set of tools that enable the creation of fully functional medical applications. These applications support the development of new diagnostic techniques, treatments, and medical advancements, all of which significantly impact healthcare.

The digitization of medical questionnaires effectively reduces the time burden on healthcare professionals [4]. Instead of requiring in-person interviews for every patient, digital questionnaires allow individuals to complete assessments remotely, increasing accessibility and efficiency. Figures 4 and 5 illustrate the user-friendly and intuitive interface designed to enhance the user experience.

The IPAQ and MMSE are excellent examples of digitized medical assessments, as they monitor two critical aspects of health: physical activity and cognitive function. If test results indicate potential issues, healthcare professionals can promptly contact the individual for further evaluation or intervention.

In conclusion, frameworks such as ResearchKit serve a critical function in the design and development of medical applications. They offer resources that facilitate development, minimize costs, and enhance both the quality and quantity of data collection. As digital health solutions continue to progress, utilizing such frameworks will be indispensable in furthering clinical research and personalized medicine.

5. Conclusion

Although developing applications for medical use can be complex, their implementation is crucial in advancing technological innovations in healthcare. Digital tools enhance disease monitoring, facilitate data collection, and support medical research, ultimately improving the ability of healthcare professionals to develop new treatment plans and preventive strategies.

By integrating technology into clinical practice, the medical field is evolving toward more personalized, data-driven, and efficient healthcare solutions. Frameworks like ResearchKit enable faster, large-scale data acquisition, which is essential for refining diagnostic tools, optimizing treatments, and advancing public health initiatives.

Furthermore, the integration of advanced computational technologies, such as artificial intelligence (AI) and Machine Learning (ML), enhances the analysis and interpretation of medical data. These technologies allow for the identification of pathological correlations, the detection of patient trends, the improvement of diagnostic accuracy, and even the prediction and prevention of diseases before they manifest.

Today, medicine and computing are increasingly interconnected, driven by groundbreaking discoveries facilitated by frameworks like ResearchKit. Understanding the development and implementation of such software is crucial for further advancing patient care, disease management, and medical research.

References

1. Singh, A.P.; Shahapur, P.R.; Vadakedath, S.; Bharadwaj, V.G.; Kumar, D.P.; Pinnelli, V.B.; Godishala, V.; Kandi, V.; Singh, D.A.P.; Shahapur, P.R.; et al. Research Question, Objectives, and Endpoints in Clinical and Oncological Research: A Comprehensive Review. *Cureus* **2022**, *14*. Publisher: Cureus, <https://doi.org/10.7759/cureus.29575>.
2. Novitzke, J.M. The significance of clinical trials. *Journal of Vascular and Interventional Neurology* **2008**, *1*, 31.
3. Regmi, P.R.; Waithaka, E.; Paudyal, A.; Simkhada, P.; van Teijlingen, E. Guide to the design and application of online questionnaire surveys. *Nepal Journal of Epidemiology* **2016**, *6*, 640–644. <https://doi.org/10.3126/nje.v6i4.17258>.
4. Martin-Sanchez, F.; Verspoor, K. Big Data in Medicine is Driving Big Changes. *Yearbook of Medical Informatics* **2014**, *9*, 14–20. <https://doi.org/10.15265/IY-2014-0020>.
5. Jardine, J.; Fisher, J.; Carrick, B. Apple's ResearchKit: smart data collection for the smartphone era? *Journal of the Royal Society of Medicine* **2015**, *108*, 294–296. <https://doi.org/10.1177/0141076815600673>.
6. Lucena, P. ¿Qué es el framework? | 2025, 2023.
7. ResearchKit.
8. What is Machine Learning? Guide, Definition and Examples.
9. L'Heureux, A.; Grolinger, K.; El Yamany, H.; Capretz, M. Machine Learning With Big Data: Challenges and Approaches. *IEEE Access* **2017**, *PP*, 1–1. <https://doi.org/10.1109/ACCESS.2017.2696365>.
10. Craig, C.L.; Marshall, A.L.; Sjöström, M.; Bauman, A.E.; Booth, M.L.; Ainsworth, B.E.; Pratt, M.; Ekelund, U.; Yngve, A.; Sallis, J.F.; et al. International physical activity questionnaire: 12-country reliability and validity. *Medicine & science in sports & exercise* **2003**, *35*, 1381–1395.
11. Truong, Q.C.; Cervin, M.; Choo, C.C.; Numbers, K.; Bentvelzen, A.C.; Kochan, N.A.; Brodaty, H.; Sachdev, P.S.; Medvedev, O.N. Examining the validity of the Mini-Mental State Examination (MMSE) and its domains using network analysis. *Psychogeriatrics* **2024**, *24*, 259–271.
12. Zhang-Xu, A.; Vivanco, M.; Zapata, F.; Málaga, G.; Loza, C. Actividad física global de pacientes con factores de riesgo cardiovascular aplicando el "International Physical Activity Questionnaire (IPAQ)". *Revista Médica Herediana* **2011**, *22*. <https://doi.org/10.20453/rmh.v22i3.1085>.
13. Arevalo-Rodriguez, I.; Smailagic, N.; Roqué i Figuls, M.; Ciapponi, A.; Sanchez-Perez, E.; Giannakou, A.; Pedraza, O.L.; Bonfill Cosp, X.; Cullum, S. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). *The Cochrane Database of Systematic Reviews* **2015**, *2015*, CD010783. <https://doi.org/10.1002/14651858.CD010783.pub2>.
14. CreatingSurveys Document.

15. Craig, C.L.; Marshall, A.L.; Sjöström, M.; Bauman, A.E.; Booth, M.L.; Ainsworth, B.E.; Pratt, M.; Ekelund, U.; Yngve, A.; Sallis, J.F.; et al. International Physical Activity Questionnaire: 12-Country Reliability and Validity. *Medicine & Science in Sports & Exercise* **2003**, *35*, 1381. <https://doi.org/10.1249/01.MSS.0000078924.61453.FB>.
16. Rodríguez, D.G.P. DR. HÉCTOR DAVID MARTÍNEZ CHAPA, 2017.
17. Barrera, R. Cuestionario Internacional de actividad física (IPAQ). *Revista Enfermería del Trabajo* **2017**, *7*, 49–54. Publisher: Asociación de Especialistas en Enfermería del Trabajo Section: Revista Enfermería del Trabajo.
18. Dr. Jesús Avilio Martínez Beltrán.; Dr. Manuel Gerónimo Lomas López.; Dr. Juan Humberto Medina Chávez.; Dr. Salvador Amadeo Fuentes Alexandro.; Dr. Raúl Agustín Sobrino Martínez de Arredondo. Diagnóstico y tratamiento de DEMENCIA VASCULAR En el adulto En los tres niveles de atención, 2017.
19. Irina Planelas.; Caterina Calderon. MINI-MENTAL EXAMEN COGNOSCITIVO MINI-MENTAL.
20. ResearchKit - ResearchKit & CareKit.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.