JMIR Rehabilitation and Assistive Technologies

Development and Evaluation of Rehabilitation, Physiotherapy and Assistive Technologies, Robotics, Prosthetics and Implants, Mobility and Communication Tools, Home Automation and Telerehabilitation
Volume 13 (2026) ISSN 2369-2529 Editor in Chief: Sarah Munce, MS, PhD

Contents

Original Papers

Leveraging Large Language Models for Early Detection of Anomaly Work Injury Cases: Data-Driven Approach to Rehabilitation Efficiency

Peter Q Chen[1], MSc; Hayley Y W Gu[1], MSc; Heidi K Y Lo[2], MBChB, MRCPsych; Wing Chung Chang[2], MBChB, MD; Cameron J M Lai[3], MBA; Sun H S Lai[3], BSc, PhD; Andy S K Cheng[4], BSc, PhD; Peter H F Ng[1], MSc, PhD

[1]Department of Computing, Hong Kong Polytechnic University, PQ710, Mong Man Wai Building, Hong Kong, China (Hong Kong)

[2]Department of Psychiatry, University of Hong Kong, Hong Kong, China (Hong Kong)

[3]Total Rehabilitation Management (HK) Limited, Hong Kong, China (Hong Kong)

[4]Department of Health and Physical Education, The Education University of Hong Kong, Hong Kong, China (Hong Kong)

Corresponding Author:
Peter H F Ng, MSc, PhD
Department of Computing, Hong Kong Polytechnic University, PQ710, Mong Man Wai Building, Hong Kong, China (Hong Kong)

Abstract

Background: Large language models (LLMs) have demonstrated potential in automating the analysis of unstructured clinical data, yet their application in rehabilitation therapy for work injury cases remains underexplored.

Objective: We aimed to evaluate the performance of an LLM-assisted approach for the rapid identification of anomalous rehabilitation cases related to work injuries to enhance scalability and precision in case management.

Methods: We retrospectively analyzed 110,346 deidentified work injury cases between 2001 and 2024 from a leading rehabilitation coordination company in Hong Kong, representing approximately 20% of all work injury incidents in the region. LLMs were used to estimate the expected duration of recovery based on free-text injury descriptions. The cases in which the actual number of medically certified sick leave days exceeded the LLM-predicted maximum were classified as anomalies.

Results: The LLM-assisted method achieved high accuracy, with GPT-4o achieving over 73% accuracy in nonanomalous classification and 79% accuracy in all dataset detection, outperforming comparator models. The model maintained high accuracy across subgroups and demonstrated the reliable extraction of information from free-text notes.

Conclusions: The proposed method demonstrated robustness when evaluated on a large-scale dataset with a bimodal age distribution. This study highlights the potential of LLMs to transform rehabilitation workflows by automating anomaly detection at scale. The method also shows promise in tailoring rehabilitation strategies to age-specific needs and leveraging LLM tools for efficient case management. However, a key limitation is that the dataset includes only injury cases from a single geographic region, potentially limiting the generalizability of the findings to other populations or health care systems.

(JMIR Rehabil Assist Technol 2026;13:e80607) doi:10.2196/80607

KEYWORDS

anomaly detection; large language model; work injury; rehabilitation prediction; machine learning

Introduction

Efficient and targeted rehabilitation management is essential to ensure that individuals with severe conditions receive timely and appropriate care [1]. However, current work injury management often lacks the precision needed to allocate resources optimally, resulting in delays and inefficiencies in addressing high-priority cases [2]. A core challenge lies in the misallocation of attention and services, where relatively minor injuries with predictable recovery trajectories are sometimes treated with the same urgency as more complex cases. This misdirection not only burdens the health care system but also diverts valuable clinical time and expertise away from anomalies, cases involving severe injuries or irregular patterns of recovery that require specialized evaluation or intervention [3]. Without a robust mechanism to distinguish these cases early in the workflow, rehabilitation systems risk compromised patient outcomes, inefficient use of limited resources, and waste critical resources. In Hong Kong, the term “anomalies” refers to cases where the injured worker’s records indicate a severe injury requiring special or intensive care [4]. These cases may also suggest potential discrepancies, such as claims that have been overstated for additional compensation or legal benefits, indicating that the incident may not follow typical recovery

patterns according to industrial practice. In the Asia-Pacific context, including Singapore's employment practices guidance and Australian public sector leave management, cases that exceed expected or allowable sick leave provisions are treated as requiring attention, which aligns with our operational use of sick leave exceedance to flag potential anomalies [5-7]. Nonanomalous data refer to cases in which the injured worker experiences a light injury expected to heal in the usual course, with a standard recovery process leading to a timely return to work. These records represent the standard outcomes, without complications or indications of potential fraud. Some anomalies may also signal potential inconsistencies, such as exaggerated claims made for extended compensation or legal advantage. In industrial practice, comprehensive annotations about why a case is "special" are typically unavailable. The only consistent, objective post hoc indicator of atypical recovery trajectory is the realized count of medically certified sick leave days. Consequently, this study operationalizes cases requiring attention via a fast filter that flags cases whose realized sick leave exceeds a large language model (LLM)−estimated expected range. We emphasize that this is a pragmatic triage proxy, not a clinical determination of pathophysiology or fraud.

Addressing this challenge requires innovative methods for quickly and accurately identifying severe cases to optimize the distribution of resources. With the advancement of artificial intelligence (AI) techniques, clinical decision support systems have been increasingly employed across various domains to assist therapists in decision-making [8-11]. In the context of workplace injury, recent research has integrated machine learning methodologies, such as the variational autoencoder, for predicting sick leave outcomes and establishing a high alertness cliff [12]. Nevertheless, the prediction process still partially depends on the initial judgment of work injury case managers, who serve as the primary decision-makers in these cases. Senior work injury case managers consistently achieve higher accuracy compared to AI-based predictions [13]. Even with the assistance of neural networks, AI cannot rapidly achieve an acceptable level of performance without proper data preprocessing and customization. The research gap lies in the lack of efficient, data-driven methods to proactively identify anomalies in work injury management workflows. Current practices predominantly rely on random case assignment and retrospective corrections, resulting in wasted time and resources.

Recently, LLMs have demonstrated exceptional capabilities in processing language-related data, even passing the United States Medical Licensing Examination [14,15]. It has also been studied in several medical fields [16-18]. Numerous studies and surveys have investigated LLMs' ability to assume specific roles based on provided profiles, with results indicating that LLMs can effectively simulate profiled characters [19-21]. Simply prompting LLMs with a data description can generate responses in less than 1 minute without requiring additional model training. However, a critical research gap remains in determining how to constrain the outputs of LLMs and how to effectively design methods that leverage LLMs to detect anomalies in incoming injury cases efficiently.

Unlike traditional rehabilitation workflows, where senior work injury case managers must spend considerable time manually identifying anomalous cases, we developed an LLM-assisted method to streamline and accelerate this process. By leveraging prompt engineering techniques, we structured the input and constrained the output format to support accurate and efficient initial screening. The method is grounded in clinical reasoning: each injured worker is expected to follow a typical recovery trajectory, reaching a work-ready state within a medically anticipated time frame based on the nature and severity of the injury. LLMs, trained on extensive digital corpora that include medical and occupational content, are well positioned to infer such expectations and assist in detecting deviations from normative recovery patterns [22].

The scenario mirrors current practices in work injury management, where cases are often assigned randomly to junior or senior work injury case managers, only to discover later that certain cases would have benefited from senior-level expertise from the outset. This misallocation frequently results in delays and inefficient use of resources.

To this end, we proposed a novel approach leveraging LLMs to detect anomalies in occupational rehabilitation in the context of work injury management. Our method offered a potentially fast, scalable, and highly accurate solution for identifying severe cases based on data from work injury cases. Furthermore, this research collected over 110,000 work injury cases from a local company in Hong Kong, which handles nearly 20% of the total work injury cases yearly [23]. The data were used to validate our method, and several pilot studies were conducted for feasibility assessment, including model selection. Meanwhile, this research aims to uncover the key factors that characterize anomalies in this dataset, providing deeper insights into the decision-making process and facilitating a more informed allocation of resources. Our objectives are 2-fold: (1) developing a robust LLM-based method for anomaly detection in work injury management and (2) utilizing exploratory data analysis to uncover potential age and work-injury patterns in these anomaly cases.

Methods

LLM-Assisted Anomaly Fast Detection Method

This method uses a fast and reliable alertness cliff to classify cases more effectively. If the total number of medically certified sick leave days for an injured worker exceeds this cliff, the case is flagged as potentially anomalous and prioritized for review by senior-level personnel or a detailed evaluation. An LLM predicts the expected duration of sick leave for each case from injury and accident descriptions. By comparing realized sick leave with the model-predicted recovery days, cases exceeding the cliff are classified as anomalies, and those within the expected range are considered normal. To enhance precision, 3 aggregation rules are used to define the decision cliff, referred to as the cliff: the maximum, the average, and the median of 3 independent LLM−generated recovery estimates. A case whose realized sick leave exceeds this cliff is classified as an anomaly. To mitigate variability and improve reliability, the LLM is queried 3 times per case using the same prompt structure, and the final decision is derived from the aggregated predictions to produce a robust, data-driven anomaly detection process [24].

As illustrated in Figure 1, the workflow begins when a new work injury case is received. The "Query LLM 3 times and aggregate results" step rapidly determines the cliff using an LLM. The procedure begins by preparing case information, followed by preprocessing to retrieve demographics and extract key details, such as accident and diagnosis information. This content is embedded into a prompt based on the template shown in Figure 2. For each case, the LLM application programming interface (API) is queried 3 times to obtain predicted recovery periods, which serve as the cliff indicator for anomaly classification. A case is classified as nonanomalous if its sick leave has not yet exceeded the predicted cliff. For ongoing cases, sick leave days are incremented and reassessed against the cliff until the case is closed. Once the sick leave surpasses the cliff, the case is classified as an anomaly and referred to a senior work injury case manager for intervention.

Figure 1. Large language model (LLM)−assisted anomaly fast detection method. API: application programming interface.

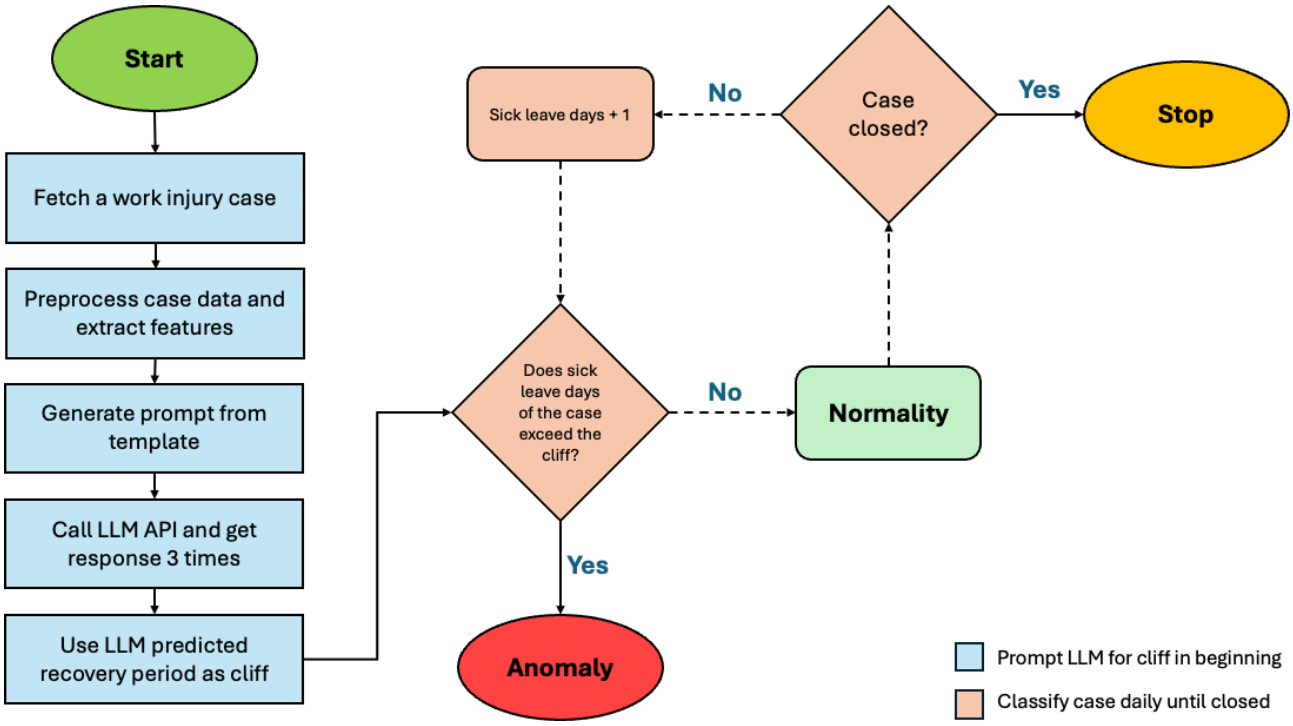
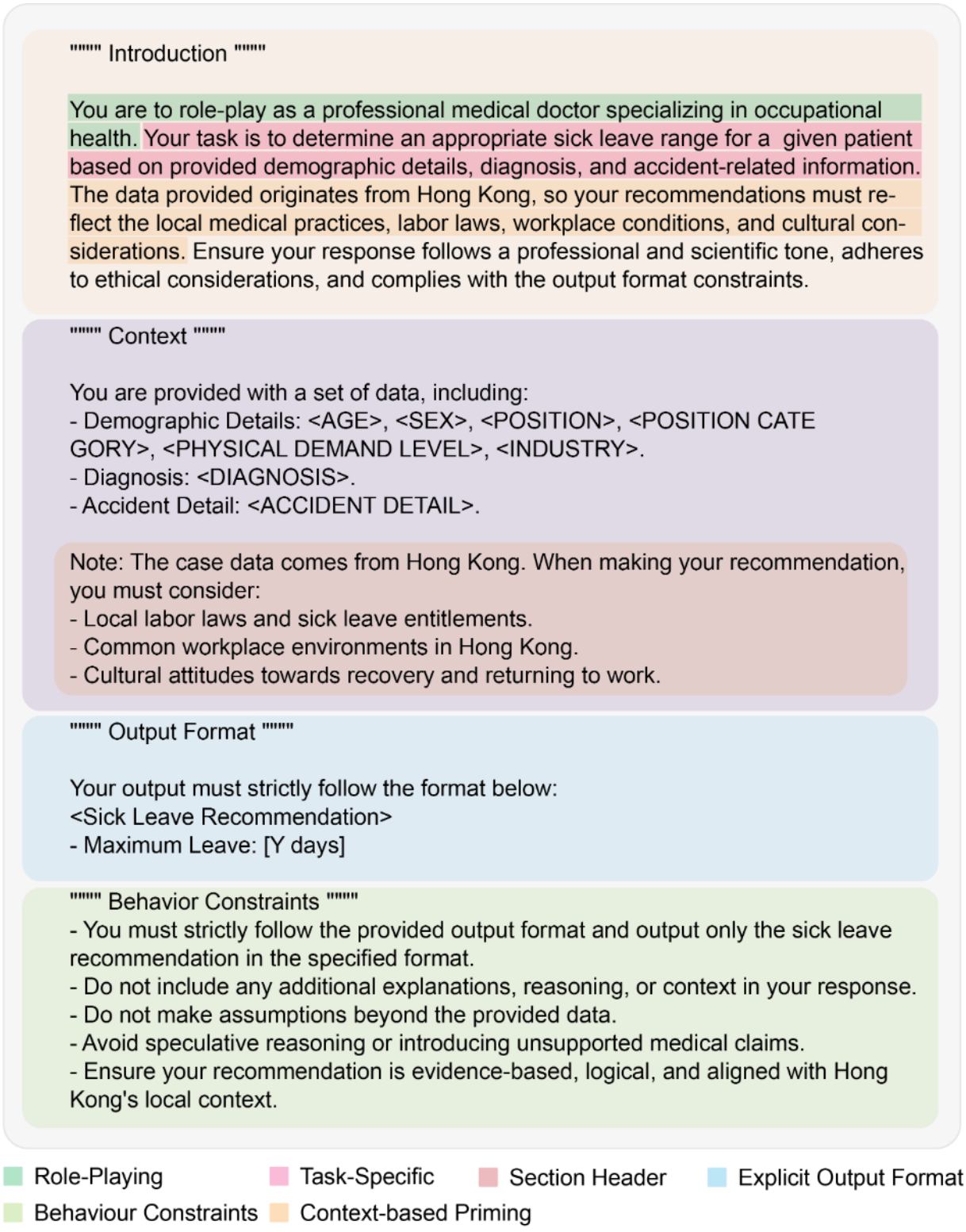


Figure 2. Concrete prompt used to acquire information from a large language model.



The primary outcome is the accuracy of anomaly and nonanomalous classification in work injury cases, assessed on a dataset from a leading local work injury management company with expert-verified labels. The systematic use of LLM predictions enhances detection sensitivity while reducing manual workload, enabling case managers to focus on genuinely complex cases that require expert attention.

In routine operations, rich clinical detail is often unavailable at intake, so triage relies on minimal text and basic demographics. To enable low-latency and low-cost prioritization, a case-specific cliff is defined as the LLM-estimated expected duration of sick leave, serving as a data-driven prior on typical recovery given available notes. As the case progresses, if the running tally of medically certified sick leave exceeds this prior, particularly early in the timeline, the case is automatically queued for senior review. This mechanism functions as a workload triage heuristic rather than a diagnostic judgment, triggering timely escalation in cases of information scarcity,

improving allocation precision, and deferring definitive clinical determinations to expert assessment.

The method was validated using a dataset of 110,346 real-world work injury cases provided by a leading work injury management company in Hong Kong.

Prompt Template

Figure 2 demonstrates the detailed prompt engineering template, which utilizes multiple prompting techniques to enhance LLM performance. For clarity, the section header technique, such as "Introduction," "Context," "Output Format," and "Behavior Constraints," is used in the prompt template [25]. In the first part of the prompt template, the role-playing technique enhances contextual understanding, adaptability, and response accuracy by simulating specific personas, perspectives, or expertise in a given scenario [26]. To improve response relevance and coherence, the template specifies that all cases occurring in Hong Kong should utilize the context-based priming technique [27]. In the "Context" section of the prompt template, the input data for the injury case, including demographic details (eg, age, occupation), diagnosis details (textual description), and accident details (textual description), were included. In the "Output Format" part, the Explicit Output Format avoids undesired reasoning steps (eg, Chain-of-Thought) or other deviations, ensuring the LLM generates responses strictly in the intended structure without adding irrelevant content [28]. The "Behavior Constraints" part also serves a similar purpose, ensuring that the model's responses remain factual, precise, and contextually appropriate. Prompts are explicitly contextualized to Hong Kong's legal and clinical environment to enhance ecological validity, with strict output schemas for parsability and consistency [29]. Explicit instructional constraints are embedded to reduce hallucination and enforce adherence to the analytical task.

Pilot Study

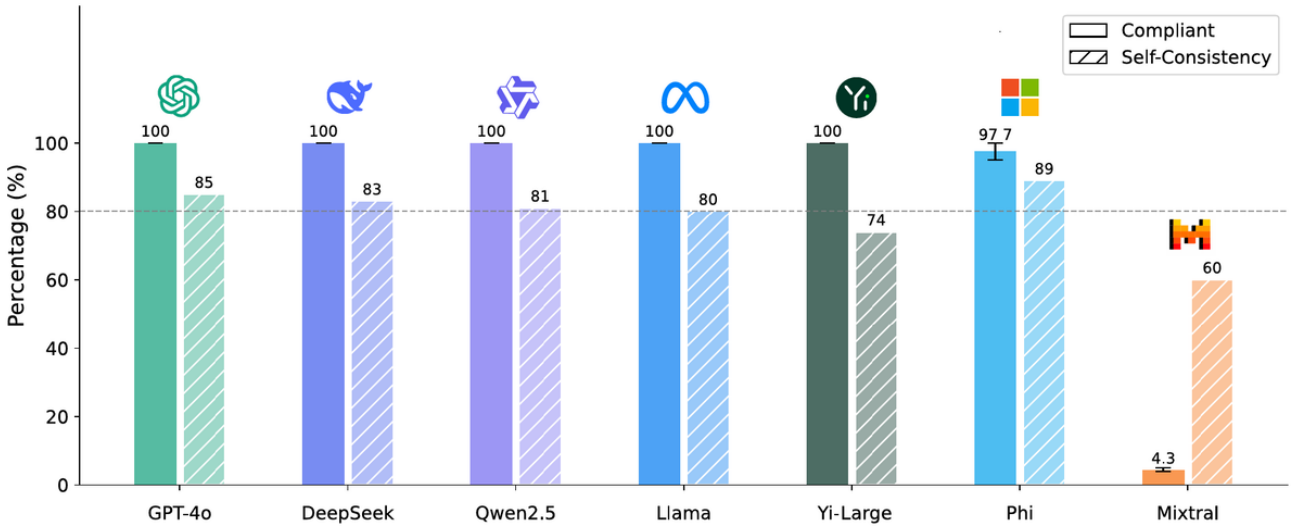
In our pilot study, we examined the varying strengths of different LLMs (eg, mathematical reasoning) and recognized that model size and architecture significantly influence performance [25,30-32]. To ensure a robust evaluation, we selected the largest and most widely recognized models from diverse LLM families, encompassing a broad range of architectures and capabilities.

Our primary objective was to determine whether these models could interpret a predefined prompt template and generate outputs that conformed to the required structure. Rather than examining their reasoning or predictive abilities, we focused on consistency, introducing 2 metrics: Compliance, which measures adherence to guidelines for producing the desired content, and Self-Consistency, which assesses whether the same response is generated across 3 repeated trials. We randomly sampled 100 cases from the dataset, prompted each model 3 times per case, and recorded the number of outputs that followed the required format. We focus on per-case anomaly triage using an upper bound on expected sick leave, so the prior work's center-focused reproducibility framework and large-run benchmarking do not fit our objective, evaluation needs, or deployable low-stochasticity protocol [32].

Figure 3 presents the results of this pilot study. Compliance represents the proportion of responses that satisfied our prescribed guidelines, while Self-Consistency quantifies each model's consistency across repeated prompts. Among 7 leading commercial LLMs, ChatGPT-4o, DeepSeek-V3, Qwen2.5-72B Instruct, Llama-3.1-405B Instruct, and Yi-Large achieved perfect Compliance; other models failed to avoid generating undesired content. We also examined internal consistency by comparing outputs across 3 prompts, categorizing them as identical across all trials, identical in at least 2, or unique each time. ChatGPT-4o exhibited the highest Self-Consistency, and several other models met our chosen 80% cliff for the subsequent experiments.

However, certain models with high Self-Consistency struggled with Compliance. This discrepancy is especially concerning, given our objective of providing a fast and reliable anomaly detection system for work injury management companies. Strict adherence to instructions is critical: any deviation can introduce erroneous or fabricated data, ultimately undermining the detection process. Ensuring compliance is thus essential to maintain the integrity of anomaly detection.

Figure 3. Compliance and self-consistency of different large language models.



Work Pipeline

Based on the results of the pilot study, we selected DeepSeek-V3, ChatGPT-4o, LLaMA-3.1-405B-Instruct, and Qwen2.5-72B to serve as LLM agents in our framework. All selected models achieved 100% compliance with the prompt template and demonstrated over 80% self-consistency across repeated responses.

A request-response framework was implemented using the FireWorks API in Python. Decoding parameters were configured for low-variance outputs using a temperature of 0.2 and a top-p of 1.0, which empirically reduced variability while maintaining robust adherence to the required output schema. A temperature of 0.0 was considered for full determinism; however, some models exhibited occasional output truncation or schema noncompliance at strictly deterministic settings during preliminary checks.

The demographic information, incident accident details, and clinical diagnoses were embedded into a standardized prompt template, as shown in Figure 2. Each prompt was submitted to the LLM via an API. A Python script extracted the responses to generate a predicted maximum duration of medically certified sick leave for each case. The data were then visualized to assess the accuracy of the LLM-based method.

Data Sources

Our dataset originates from a leading work injury management company in Hong Kong, which manages approximately 20% of work injury cases annually, covering records from 2001 to 2024. The dataset comprises 110,346 cases, with a gender distribution of 41.3% female and 53.6% male. The study population is predominantly Chinese, with individuals ranging in age from 18 to 83 years. This broad demographic coverage provides a robust basis for analyzing patterns across different age groups and genders within a relatively homogeneous ethnic context. Input leakage was not possible in this study. All records are confidential medical data, fully classified, and never publicly released or exposed to the models beyond the controlled evaluation pipeline, thereby preventing any external contamination that could bias LLM outputs or compromise validity.

Within this dataset, 15,575 cases were recorded as having zero sick leave days, which were treated as potential data entry errors. We exclude zero-day legitimate cases as outliers because they are immediately escalated to senior rehabilitation coordinators on day zero and handled outside our fast detection system, which targets anomalies only after predicted sick leave durations are exceeded. An additional 9230 cases had nonzero sick leave durations but contained missing values. For the primary predictive analysis, we used 85,541 cases that reported a nonzero number of medically certified sick leave days and had no missing data. These were inputs for the LLMs to predict the expected duration of normal recovery. Although excluded from the prediction task, the remaining data groups were also analyzed to extract relevant insights, given their substantial size.

Data Preprocessing

The dataset underwent staged preprocessing to ensure consistency and analytical suitability. Records outside the target time window were removed, implausible values were constrained within reasonable bounds, and entries with nulls in critical analytical fields were excluded. Noncritical descriptive fields with missing values were imputed using a neutral placeholder to preserve coverage while signaling incompleteness.

Categorical features were standardized through controlled vocabulary mapping, consolidation of multivalued entries into explicit multicategory indicators, and aggregation of low-frequency categories to mitigate sparsity. Text fields were sanitized by removing noninformative placeholders, and duplicates were eliminated based on content equivalence. A focused set of salient variables was retained for downstream analysis.

Data Analysis

In this study, we utilize a comprehensive set of metrics to rigorously evaluate the performance of the LLM-assisted anomaly detection method, encompassing classification accuracy, error magnitude, and model reliability. Classification accuracy, a core metric, is calculated based on 3 cliff-based methods: method 1 (maximum of 3 LLM predictions), method 2 (average of 3 predictions), and method 3 (median of 3 predictions). To further assess prediction deviations between realized sick leave days and LLM-predicted cliffs, we compute mean absolute error (MAE), mean squared error, root mean squared error (RMSE), mean absolute percentage error (MAPE), and mean percentage error (MPE). Additionally, Compliance (adherence to structured output formats) and Self-Consistency (reproducibility of outputs across repeated prompts) are quantified to ensure model reliability. We evaluate misclassification deviation, summarized through percentiles (eg, 50th and 75th percentiles), to analyze error distribution. Furthermore, exploratory data analysis is conducted to provide insights into the dataset, including descriptive statistics such as injury frequency, demographic distributions (eg, age, gender, occupation), anomaly prevalence, and misclassification patterns across key variables like body part and industry type. Together, these metrics and calculations form a robust framework for assessing the precision, reliability, and operational effectiveness of the proposed LLM-based anomaly detection system.

We evaluated triage performance using cliff-based classification derived from LLM-predicted “cliffs” of expected sick leave duration. For each case, the LLM was queried 3 times with the same prompt. The per-case decision cliff was then defined by 1 of the 3 aggregation rules: method 1, method 2, and method 3. A case was classified as an anomaly if it realized medically certified sick leave days exceeded the chosen cliff; otherwise, it was classified as nonanomalous. The primary performance metric was accuracy, computed as the proportion of correctly classified cases over the evaluation set, and reported overall and stratified by anomaly and nonanomalous subsets to characterize trade-offs across decision rules and models.

To assess reliability and error magnitude, we further quantified misclassification deviation, defined for errors as the absolute difference between realized sick leave days and the decision cliff, summarized via percentiles (eg, 50th and 75th).

To gain a deeper understanding of the phenomenon in the dataset, salary is treated as a composite proxy that captures differences across job types, seniority, contract structures, and work experience. Given the absence of granular role-level pay scales, salary should not be interpreted as a pure measure of experience but as an indicator shaped by occupational category and tenure.

Ethical Considerations

This study introduced an innovative anomaly detection method for work injury rehabilitation, validated using real-world cases from Hong Kong. The project has been approved by the PolyU Institutional Review Board (reference HSEARS20250406002). A pilot study was first conducted to evaluate the performance of several well-known commercial LLMs, which informed the selection of the most effective models for the subsequent experiments.

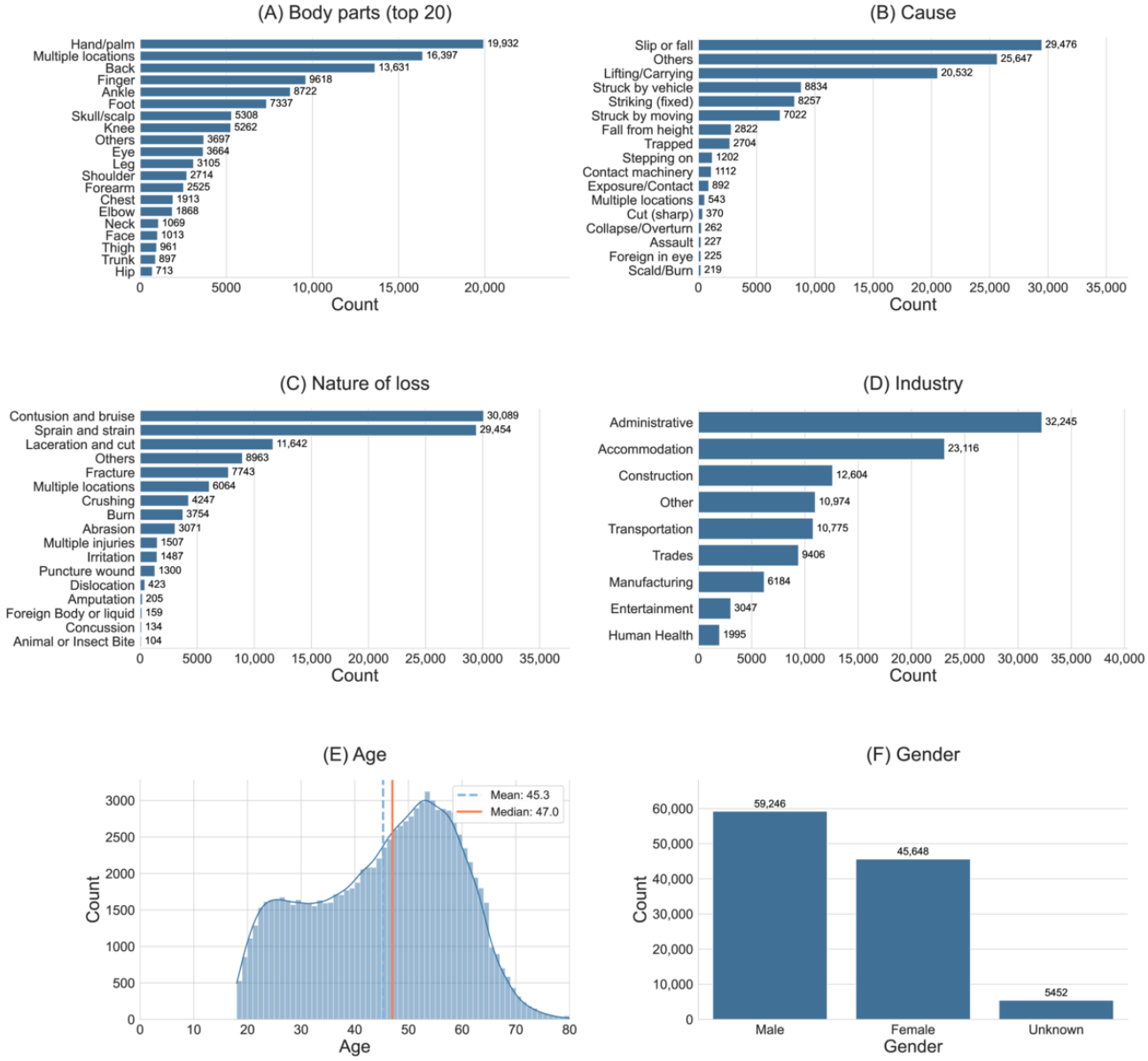
The dataset was provided by one of the largest work injury management companies in Hong Kong, which manages nearly 20% of the government-reported work injury rehabilitation cases. The data were shared exclusively for research purposes under a strict confidentiality agreement. Prior to transfer, all records were anonymized by the provider to ensure the protection of personal information.

Results

Work Injury Dataset

Figure 4 presents a comprehensive analysis of 110,346 deidentified occupational injury cases managed by a leading rehabilitation coordination company in Hong Kong between 2001 and 2024. Injuries predominantly involved peripheral anatomical regions, with fingers (n=16,397) and backs (n=13,631) collectively accounting for nearly one-quarter of all incidents, followed by hand or palm injuries (n=9618) and ankle injuries (n=8722). Consistent with these anatomical findings, the most common types of loss were contusions or bruises (n=30,089) and sprains or strains (n=29,454), whereas open wounds, such as lacerations and cuts (n=11,642), and fractures (n=7743), occurred less frequently. Industry-specific data indicate a substantial burden arising from labor-intensive service sectors, notably “Administration and support services” (n=32,245) and “Accommodation and food service activities” (n=23,116), collectively accounting for more than half of all cases and surpassing the construction sector (n=12,604) in this dataset. Precipitating events were predominantly same-level slips, trips, and falls (n=29,476) and manual lifting or carrying tasks (n=20,532). Demographically, male workers represented a modest majority (53% of claims); nevertheless, female workers comprised a substantial proportion (42%). The age distribution was right-skewed, with a mean age of 45.3 (SD 13.3) years and a median of 47.0 (IQR 21.0) years.

Figure 4. Demographic data statistics in the dataset. (A) The top 20 most common categories of injured body parts. (B) Most common categories in the industry. (C) Most common categories in the nature of loss. (D) Most common categories in the cause of injury. (E) Gender distribution in the dataset. (F) Age distribution in the dataset.



Performance Assessment of LLMs

Table 1 shows the LLM classification accuracy across the selected models for different data categories. All LLMs mentioned in Figure 3 have been tested. For the entire dataset, all selected LLMs achieved more than 70% accuracy in the maximax, expected maximum, and median cliff criteria methods. Among the selected LLMs, GPT-4o achieved the best performance in all cliff criteria, while DeepSeek-V3 had the lowest performance. In the separate anomaly dataset containing only anomalies, the prediction accuracy of all selected models exceeds 95%. The best performance for the nonanomalous dataset comes from GPT-4o as well, achieving more than 76% under maximax cliff criteria, over 73% accuracy under expected maximum cliff criteria, and a median cliff criterion.

Table . Case classification accuracy across models[a].

	Method 1 (%)	Method 2 (%)	Method 3 (%)
Qwen	78.71	78.26	78.19
DeepSeek	76.31	75.75	75.66
Llama	78.15	76.93	76.76
GPT-4o	81.77	79.95	79.51
Anomaly			
Qwen	97.72	97.79	97.79
DeepSeek	97.59	97.81	97.80
Llama	97.69	98.02	97.97
GPT-4o	96.32	97.16	97.21
Nonanomalous			
Qwen	71.71	71.07	70.98
DeepSeek	68.48	67.62	67.51
Llama	70.96	69.16	68.95
GPT-4o	76.40	73.61	72.99

[a]Method 1 uses the maximum value among the 3 large language models (LLMs) predictions as a cliff to classify anomalies and nonanomalous. Method 2 uses the average of the 3 LLM predictions as a cliff for classification. Method 3 uses the median of the 3 LLM predictions as a cliff for classification.

Table 1 further highlights the trade-offs in classification accuracy when different criteria are applied. When using the expected maximum as the classification criterion, the model achieves higher accuracy in anomaly detection but at the cost of reduced accuracy in nonanomalous classification. However, misclassifying nonanomalous cases is relatively less consequential, as such cases typically resolve quickly, with injured workers returning to work in a short period. In contrast, misclassifying anomalies can have significant financial and operational implications. Suppose an anomaly is incorrectly classified as a normal case. In that case, the company may need to allocate additional resources to reassign a senior work injury case manager later in the process, leading to prolonged recovery times and potentially missed rehabilitation windows. From an anomaly detection perspective, Llama demonstrates the highest detection rate. However, when considering overall performance across both anomaly and nonanomalous classification, GPT-4o outperforms other models, making it the most balanced and practical choice for real-world deployment. These findings highlight the importance of selecting an LLM that optimally balances accuracy, adherence to instruction, and overall classification performance.

Table 2 shows that across the 3 aggregation strategies, absolute and squared errors remain high: MAE is approximately 72 days, and RMSE is around 158 days for all methods, indicating substantial pointwise deviations and volatility in predicting the maximum sick leave duration. Relative errors are also large: MAPE ranges from 169.86% to 195.48%, and MPE exceeds 100% for all methods, evidencing pronounced systematic overestimation. Among the alternatives, the median-based cliff (method 3) yields the lowest relative error (MAPE=169.86%, MPE=106.77%) and slightly lower dispersion, whereas using the maximum prediction as the cliff (method 1) amplifies both bias and variance; the mean (method 2) lies in between. Despite these differences, the small gaps in MAE or RMSE across methods suggest that aggregation choice alone does not resolve the core error magnitude, and bias calibration or robustness enhancements are warranted. Directly using LLMs to predict sick leave duration from demographics yields large absolute and relative errors (≈72-day MAE, ≈158-day RMSE, MAPE >169% with systematic overestimation), revealing unstable and biased point forecasts, which motivates our shift to a fast exceedance−based detection method rather than relying on raw predictions.

Table . Standard metrics for 3 methods[a].

	MAE[b]	MSE[c]	RMSE[d]	MAPE[e] (%)	MPE[f] (%)
Method 1	72.37	24,741.36	157.29	195.48	136.39
Method 2	72.42	25,173.20	158.66	172.39	110.32
Method 3	72.54	25,266.41	158.95	169.86	106.77

[a]Method 1 uses the maximum value among the 3 large language models (LLMs) predictions as a cliff to classify anomalies and nonanomalous. Method 2 uses the average of the 3 LLM predictions as a cliff for classification. Method 3 uses the median of the 3 LLM predictions as a cliff for classification.

[b]MAE: mean absolute error.

[c]MSE: mean squared error.

[d]RMSE: root mean squared error.

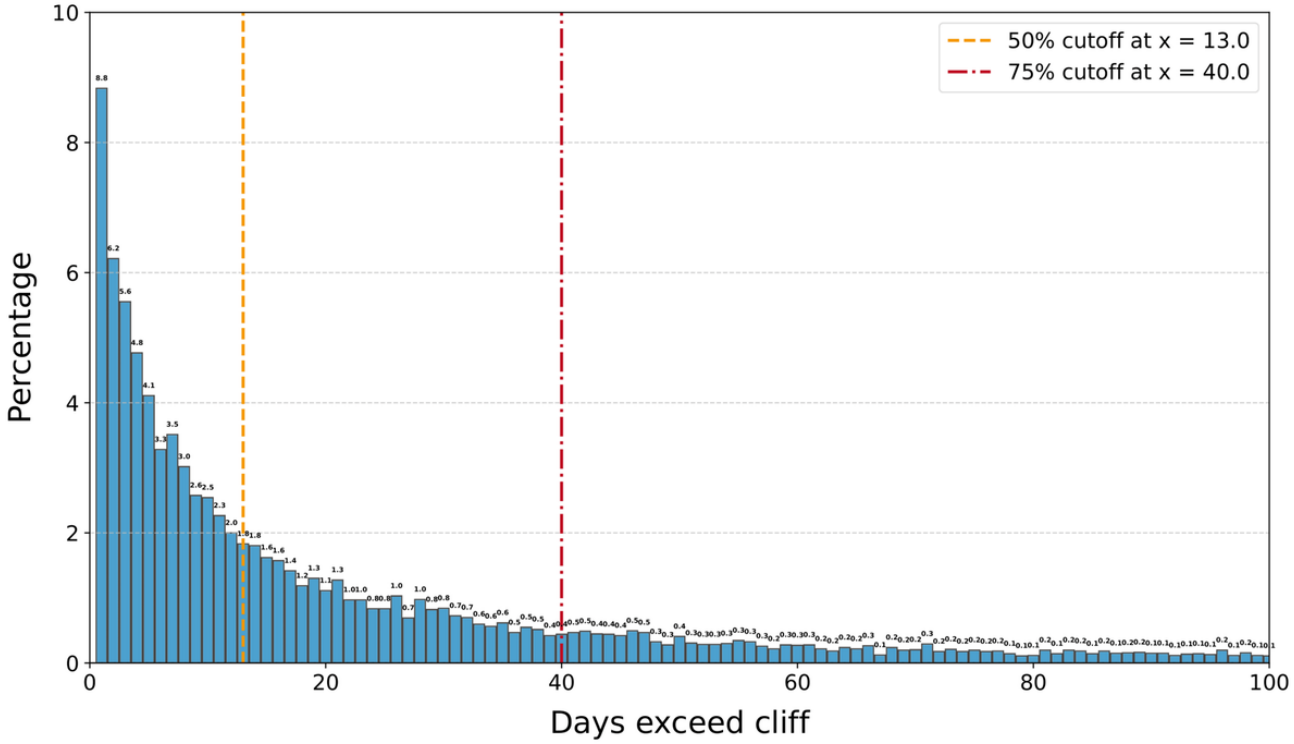
[e]MAPE: mean absolute percentage error.

[f]MPE: mean percentage error.

LLM Misclassifications Case Study

Since method 1 achieved the highest overall accuracy, this section primarily focuses on analyzing the predictions of the LLM based on method 1, as well as those generated by ChatGPT-4o, which demonstrated the highest average accuracy among all the selected models. Figure 5 illustrates the distribution of erroneous predictions that exceed the specified cliff when method 1 is applied. The yellow dashed line represents the 50th percentile of the cumulative distribution, corresponding to a deviation of 13 days. In contrast, the 75th percentile of the cumulative distribution corresponds to a deviation of 40 days. These results indicate that half of the erroneous predictions deviate from the ground truth by no more than 13 days, further underscoring the robustness of the proposed method. The previous work on similar datasets shows that the traditional variational autoencoder could only achieve an average error of 107.447 days, and they failed to predict a cliff for classifying the anomaly and nonanomalous cases [12,13]. Table 2 shows the absolute errors of direct LLM predictions, demonstrating substantial bias, while Figure 5 highlights erroneous cases flagged by our fast detection method, underscoring that our approach is intended to assist rather than replace rehabilitation coordinators.

Figure 5. Distribution of erroneous predictions beyond the cliff.



To gain a deeper insight into the LLM's prediction performance, we examined the distribution of GPT-4o outputs using method 1 (as described in the previous section). Specifically, 14,751 nonanomalous cases were misclassified as anomalies, and 847 anomalies were misclassified as normal. Figure 6 illustrates the distribution of the key variables within these misclassified nonanomalous cases. Finger has the highest proportion of misclassifications, at approximately 23%, whereas other body parts each account for around 10%. In the "Nature of Loss" variable, "Sprain & Strain" accounts for over 28% of the misclassifications, followed by "Contusion & Bruise" and "Laceration & Cut," both of which exceed 15%; the remaining

categories each fall below 10%. Regarding the Industry, the Administrative and Support Service sector and the Accommodation and Food Service Activities sector exhibit notably higher proportions of misclassifications (over 20%) relative to others. Finally, in the "Position Category" variable, most misclassifications occur under the "unknown" category.

Figure 6. Distribution of large language model misclassifications nonanomalous across the key variables (top 10).

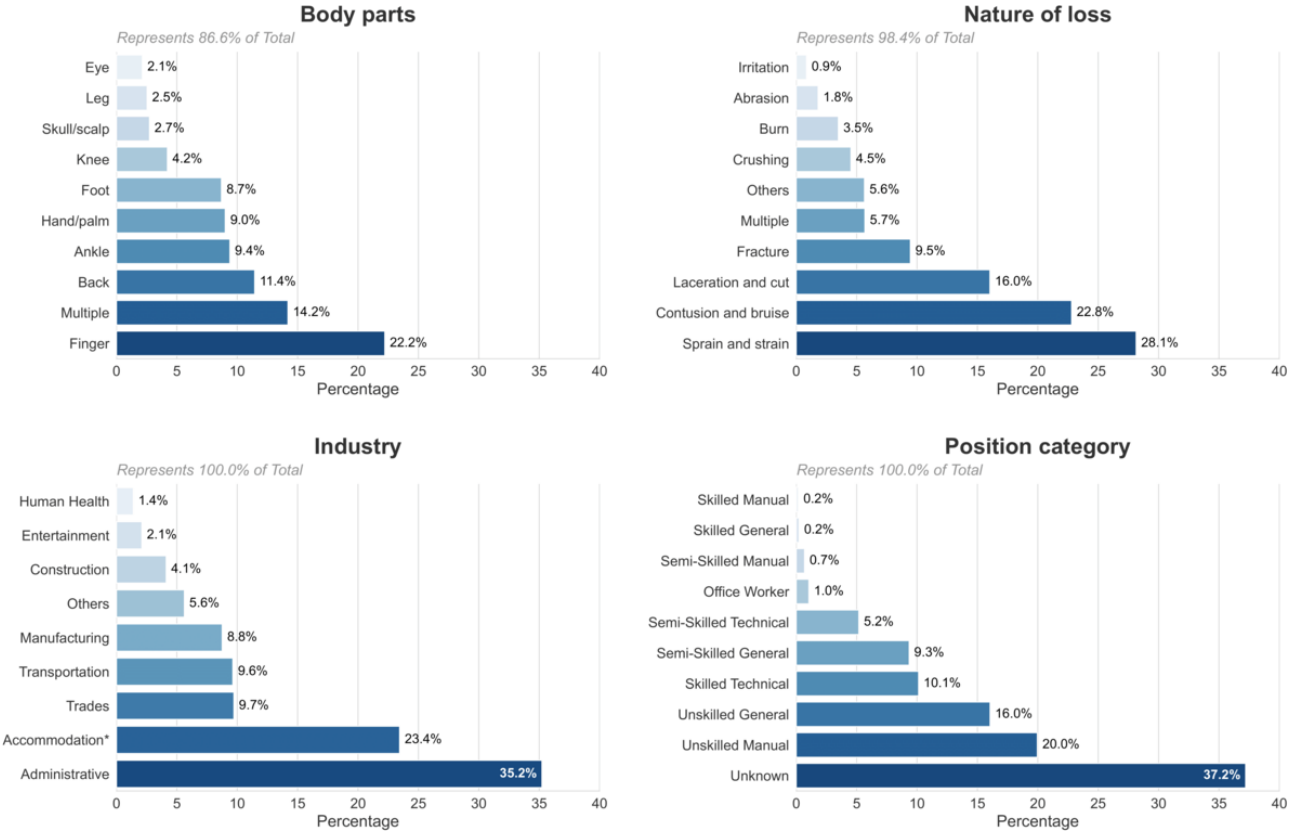
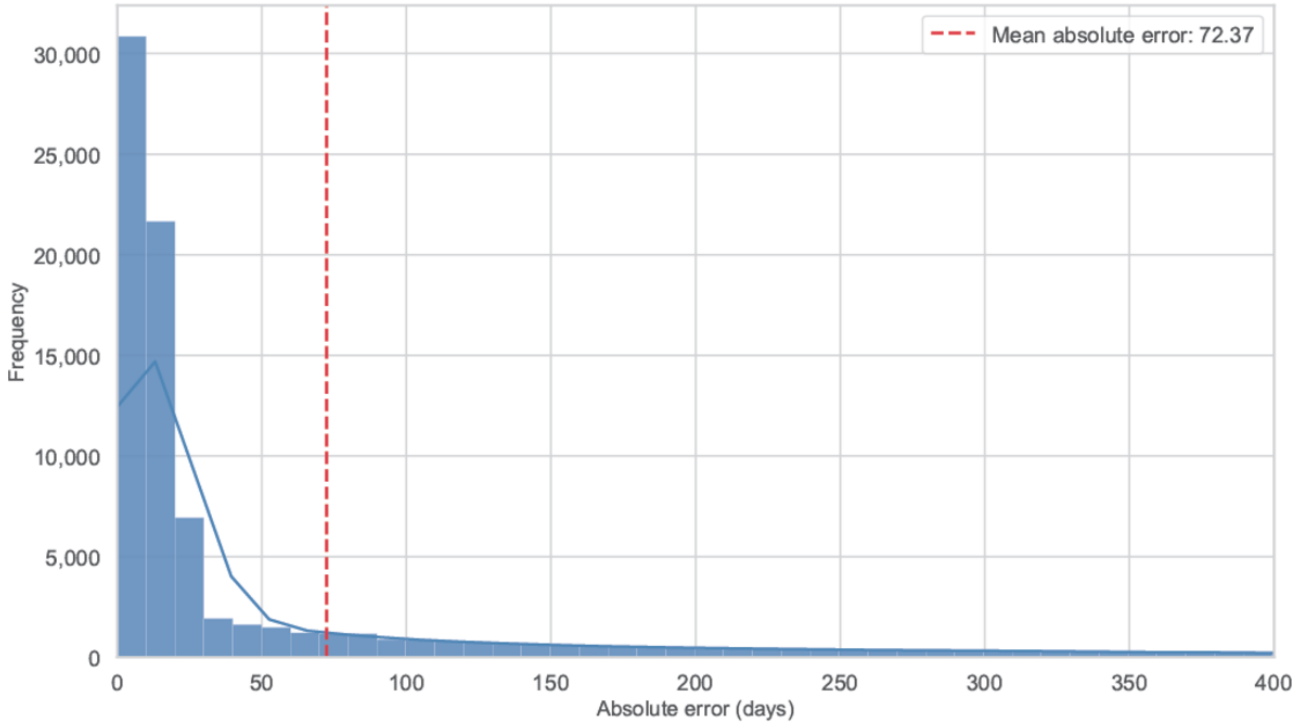


Figure 6 also suggests that certain misclassifications may arise from incomplete data, such as the absence of a position category, which impairs the LLM's ability to predict the cliff accurately. Without this critical information, the model must rely on its intrinsic knowledge, leading to increased variability and uncertainty. Additionally, for variables such as the nature of loss, injured body parts, and industry type, human judgment is also prone to significant bias, particularly in those variables with high percentages of misclassification [13].

LLM Direct Prediction Error Distribution

Figure 7 shows the distribution of absolute errors for raw LLM day predictions; the mean error is 72 days, which indicates that direct prediction yields deviations too large for practical use. Therefore, we use the LLM as a fast anomaly filter that flags cases whose realized sick leave exceeds an estimated upper bound, rather than predicting sick leave days directly.

Figure 7. Distribution of the large language model raw prediction absolute error.



Case Study of Our Approach

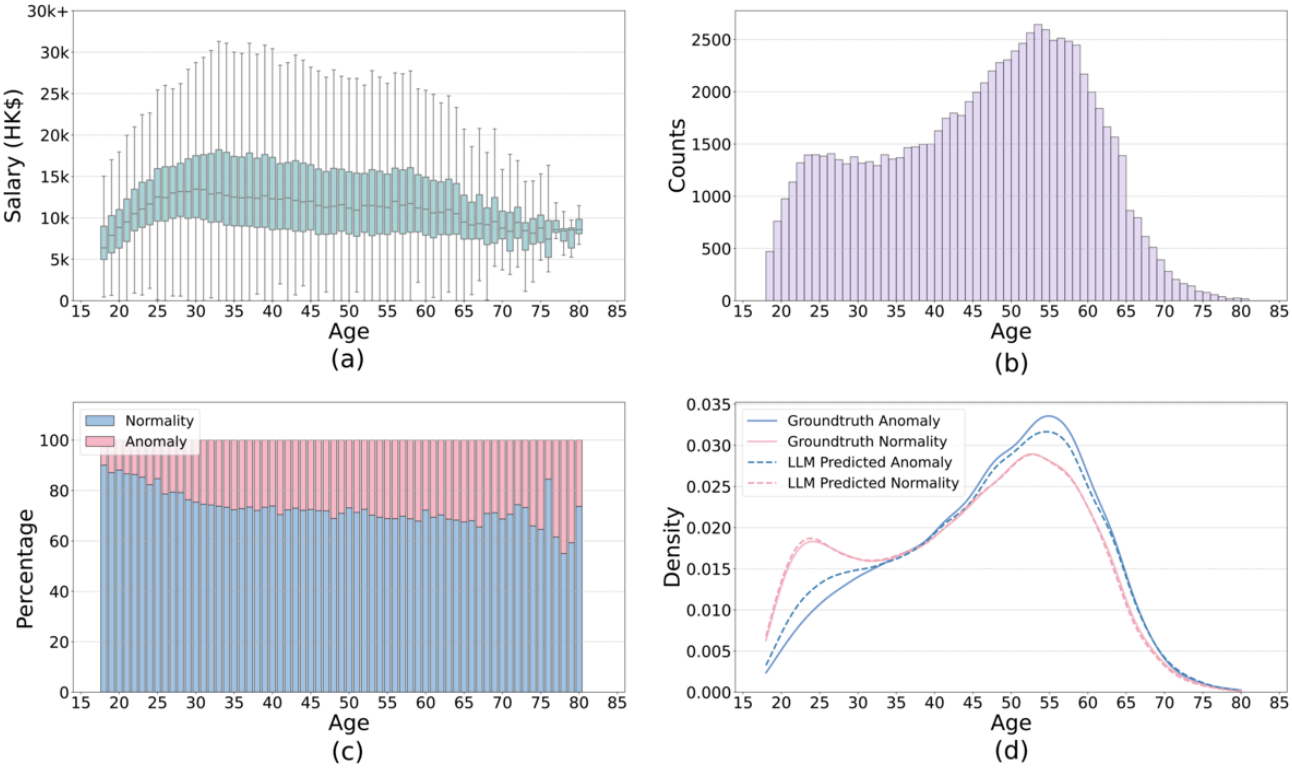
In this case study, we analyze a 51-year-old female worker who sustained a back injury while lifting a heavy basket of packaged bread. The injury was classified under Body Parts: Back, Nature of Loss: Sprain & Strain, and Cause: Injured whilst lifting or carrying, with the worker employed as a packer in the Accommodation and Food Service Activities industry. The actual sick leave duration was 6 days, while the LLM predictions for recovery duration were consistent across all methods: result 1=14 days, result 2=14 days, and result 3=14 days, leading to cliff values of max_max=14 days.

The difference between the actual sick leave and the maximum predicted cliff sick leave duration max_max was –8 days, indicating that the actual recovery period was well within the predicted range and did not exceed the anomaly cliff. This demonstrates the effectiveness of the LLM-assisted method in identifying cases that conform to expected recovery patterns, ensuring that resources are not misallocated to cases that align closely with typical recovery trajectories.

Exploratory Insights From the Dataset

To examine anomalies, we present a salary box plot against age in Figure 8. The top-left panel of the figure presents a box plot of salary against age, where the median salary for each age group is extracted to reveal a clearer trend. The median salary initially increases, peaking between ages 35 and 40, before gradually declining. The top-right panel illustrates the distribution of age counts from 18 to 80 years, showing an initial peak in the early 20s, followed by a decline until the 40s, after which the number of cases rises again, reaching its highest peak around the age of 55 years. The bottom-left panel displays the percentage of anomalies across different age groups, revealing a steady increase in anomaly prevalence with age. Finally, the bottom-right panel illustrates the normalized age distributions of both anomalies and nonanomalous, alongside the LLM's predicted anomalies and nonanomalous. The red line (representing nonanomalous) closely mirrors the overall age distribution, while the blue line (representing anomalies) steadily increases and peaks near the age of 60 years. Notably, the LLM's predicted distributions align closely with the ground-truth trends.

Figure 8. Overview of salary and age distributions with anomaly analysis. (A) Salary boxplot across ages. (B) Age distribution of all data. (C) Percentage of anomaly across ages. (D) Age distribution of anomaly or normality data.



Discussion

Principal Findings

The performance of the proposed LLM-assisted anomaly fast detection method demonstrates promising results, 110,346 deidentified work-injury cases between 2001 and 2024 from a leading work injury management company in Hong Kong, representing approximately 20% of all work injury incidents in the region. Compared to previous research, which primarily assessed AI prediction accuracy by comparing it to work injury case managers' judgments, our study provides a more comprehensive evaluation [12]. Prior studies demonstrated that AI predictions could surpass human work injury case managers in both nonanomalous and anomalous cases. However, while they also attempted to predict anomaly cliffs, their findings did not explicitly report the accuracy of such predictions. In contrast, our approach ensures that the accuracy of anomaly cliff prediction is systematically analyzed, contributing to a more reliable anomaly detection framework.

In Figure 7, we illustrate the trend of salary changes, considering salary as a key indicator of work experience. It is well known that physical ability peaks in the 20s, remains stable or slightly declines until around 30 - 35, and then drops significantly thereafter [33,34]. Comparing this physical ability trend with the salary trend, we observe that workers in their early 20s possess peak physical strength but lack experience. As a result, they may be more prone to injuries; however, their quick recovery often prevents these cases from becoming anomalies. This explains the initial peak in injury cases around the age of 20 years.

By the time workers reach their 30s, they have gained significant experience (as indicated by higher salaries), while their physical ability has not yet declined substantially. Consequently, the number of work injury cases decreases between the ages of 30 and 40 years. However, after 40, salaries may begin to decline slightly as workers are unable to maintain the same working hours as before, while their physical ability drops sharply. This results in a second peak in work-related injury cases. At this stage, injuries are more severe, and recovery is less likely, contributing to an increase in both the number and proportion of anomalies.

Limitations and Future Directions

This study tests LLMs for spotting unusual rehabilitation cases. The models read injury diagnoses and accident notes, estimate a normal recovery time, and flag cases that fall outside that range. We judge the approach by how well it separates nonanomalous from anomalies, not by exact prediction accuracy. However, the limitations are that no comparison is made with human rehab coordinators, only general-purpose LLMs without medical models, and data from a single region; therefore, the findings may not be generalizable. As LLMs improve, they could streamline the rehabilitation triangle and resource planning. This research can be readily transferred to other rehabilitation systems within Hong Kong, but adaptation to other regions would be domain specific due to significant differences in work injury frameworks, legal and policy requirements, and cultural practices.

Deterministic decoding with temperature 0.0 could further enhance reproducibility. As a robustness check, future or supplementary analyses can compare performance and schema compliance at 0.0 versus 0.2 to quantify any trade-off between

determinism and adherence to output constraints. If comparable, deployment would favor 0.0 to maximize reproducibility; if not, a small nonzero temperature remains justified to preserve formatting reliability under operational constraints.

Conclusions

We present an LLM-based approach that estimates expected recovery time from injury records and flags deviations as anomalies, streamlining rehabilitation triage. Tested on more than 110,000 Hong Kong work-injury cases, the method improved classification efficiency; GPT-4o delivered the most balanced accuracy, with DeepSeek-V3 and Qwen2·5-72B Instruct close behind. Demographic analysis reveals that injuries are more frequent yet milder in younger workers, whereas those aged 40 and above experience more anomalies, reflecting reduced resilience. The approach advances data-driven rehabilitation coordination and optimizes resource allocation.

Funding

We acknowledge the support of the Innovation and Technology Fund (ITP/039/24LP) of the Government of the Hong Kong Special Administrative Region for providing the support.

Data Availability

Medical notes and the full list of responses from large language models will not be shared to protect the privacy of the patients. The responses of the large language models or derived data are available from the corresponding author.

Authors' Contributions

PQC and PHFN conceived the study, oversaw its administration, and provided supervision. PQC, HYWG, CJML, and SHSL accessed, verified, and curated the data; performed the formal analysis and methodology; developed the software; validated the results; and produced the visualizations. PQC prepared the original draft, while HYWG, HKYL, WCC, ASKC, and PHFN reviewed and edited the manuscript. PHFN additionally secured funding and supplied resources. All authors had full access to all study data and approved the final version for publication.

Conflicts of Interest

None declared.

References

1. Szeto GPY, Cheng ASK, Lee EWC, Schonstein E, Gross DP. Implementing the work disability prevention paradigm among therapists in Hong Kong: facilitators and barriers. J Occup Rehabil 2011 Mar;21(1):76-83. [doi: 10.1007/s10926-010-9256-2] [Medline: 20652377]
2. Remus L, Grope M, Lemke S, Bethge M. An innovative case management intervention for people at high risk of permanent work disability to improve rehabilitation coverage and coordination of health services: a randomized controlled trial (AktiFAME, DRKS00024648). BMC Health Serv Res 2022 Mar 15;22(1):342. [doi: 10.1186/s12913-022-07482-9] [Medline: 35292005]
3. Brownlee S, Chalkidou K, Doust J, et al. Evidence for overuse of medical services around the world. Lancet 2017 Jul 8;390(10090):156-168. [doi: 10.1016/S0140-6736(16)32585-5] [Medline: 28077234]
4. Cooper DA, Slaughter JE, Gilliland SW. Reducing injuries, malingering, and workers' compensation costs by implementing overt integrity testing. J Bus Psychol 2021 Jun;36(3):495-512. [doi: 10.1007/s10869-020-09681-9]
5. Paid sick and carer's leave. Fair Work Ombudsman. URL: https://www.fairwork.gov.au/leave/sick-and-carers-leave/paid-sick-and-carers-leave [accessed 2025-12-11]
6. Management of staff leave in the Australian public service. : Australian National Audit Office (ANAO); 2022 URL: https://www.anao.gov.au/work/performance-audit/management-staff-leave-the-australian-public-service [accessed 2025-12-11]
7. Tripartite standards: frequently asked questions (FAQs). TAFEP. 2024. URL: https://www.tal.sg/tafep/-/media/tal/tafep/getting-started/files/tripartite-standards-faqs [accessed 2025-11-12]
8. Lutz W, Deisenhofer AK, Rubel J, et al. Prospective evaluation of a clinical decision support system in psychological therapy. J Consult Clin Psychol 2022 Jan;90(1):90-106. [doi: 10.1037/ccp0000642] [Medline: 34166000]
9. Antoniadi AM, Du Y, Guendouz Y, et al. Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review. Appl Sci 2021;11(11):5088. [doi: 10.3390/app11115088]
10. Musen MA, Middleton B, Greenes RA. Clinical decision-support systems. In: Biomedical Informatics: Computer Applications in Health Care and Biomedicine: Springer; 2021:795-840. [doi: 10.1007/978-3-030-58721-5_24]
11. Borges do Nascimento IJ, Abdulazeem HM, Vasanthan LT, et al. The global effect of digital health technologies on health workers' competencies and health workplace: an umbrella review of systematic reviews and lexical-based and sentence-based meta-analysis. Lancet Digit Health 2023 Aug;5(8):e534-e544. [doi: 10.1016/S2589-7500(23)00092-4] [Medline: 37507197]

12. Ng PHF, Chen PQ, Sin ZPT, Lai SHS, Cheng ASK. Smart Work Injury Management (SWIM) system: a machine learning approach for the prediction of sick leave and rehabilitation plan. Bioengineering (Basel) 2023 Jan 28;10(2):172. [doi: 10.3390/bioengineering10020172] [Medline: 36829666]
13. Yeung YYK, Chen PQ, Ng PHF, Cheng ASK. Evaluation of the accuracy of the Smart Work Injury Management (SWIM) system to assist case managers in predicting the work disability of injured workers. J Occup Rehabil 2025 Jun;35(2):320-332. [doi: 10.1007/s10926-024-10199-7] [Medline: 38874680]
14. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 8;9(1):e45312. [doi: 10.2196/45312] [Medline: 36753318]
15. Malgaroli M, Schultebraucks K, Myrick KJ, et al. Large language models for the mental health community: framework for translating code to care. Lancet Digit Health 2025 Apr;7(4):e282-e285. [doi: 10.1016/S2589-7500(24)00255-3] [Medline: 39779452]
16. Deiner MS, Deiner RY, Fathy C, et al. Use of large language models to classify epidemiological characteristics in synthetic and real-world social media posts about conjunctivitis outbreaks: infodemiology study. J Med Internet Res 2025 Jul 2;27:e65226. [doi: 10.2196/65226] [Medline: 40601927]
17. Huang W, Wei W, He X, et al. ChatGPT-assisted deep learning models for influenza-like illness prediction in mainland China: time series analysis. J Med Internet Res 2025 Jun 27;27:e74423. [doi: 10.2196/74423] [Medline: 40577658]
18. Huang J, Lai H, Zhao W, et al. Large language model–assisted risk-of-bias assessment in randomized controlled trials using the revised risk-of-bias tool: evaluation study. J Med Internet Res 2025 Jun 24;27:e70450. [doi: 10.2196/70450] [Medline: 40554779]
19. Li N, Gao C, Li M, Li Y, Liao Q. EconAgent: large language model-empowered agents for simulating macroeconomic activities. 2024 Presented at: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); Aug 11-16, 2024; Bangkok, Thailand p. 15523-15536 URL: https://aclanthology.org/2024.acl-long [accessed 2025-12-12]
20. Zhang Z, Zhang-Li D, Yu J, et al. Simulating classroom education with LLM-empowered agents. Presented at: Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics; Apr 29 to May 4, 2025; Albuquerque, NM p. 10364-10379 URL: https://aclanthology.org/2025.naacl-long [accessed 2025-12-12] [doi: 10.18653/v1/2025.naacl-long.520]
21. Wang L, Ma C, Feng X, et al. A survey on large language model based autonomous agents. Front Comput Sci 2024 Dec;18(6):186345. [doi: 10.1007/s11704-024-40231-1]
22. Longwell JB, Hirsch I, Binder F, et al. Performance of large language models on medical oncology examination questions. JAMA Netw Open 2024 Jun 3;7(6):e2417641. [doi: 10.1001/jamanetworkopen.2024.17641] [Medline: 38888919]
23. Occupational safety and health statistics. Labour Department. URL: https://www.labour.gov.hk/tc/osh/content10.htm [accessed 2025-12-11]
24. Wang X, Wei J, Schuurmans D, et al. Self-consistency improves chain of thought reasoning in language models. arXiv. Preprint posted online on Mar 21, 2022. [doi: 10.48550/arXiv.2203.11171]
25. Brown TB, Mann B, Ryder N, et al. Language models are few-shot learners. 2020 Presented at: 34th Conference on Neural Information Processing Systems (NeurIPS 2020); Dec 6-12, 2020; Vancouver, Canada p. 1877-1901 URL: https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf [accessed 2025-12-11]
26. Shanahan M, McDonell K, Reynolds L. Role play with large language models. Nature 2023 Nov;623(7987):493-498. [doi: 10.1038/s41586-023-06647-8] [Medline: 37938776]
27. Dong Q, Li L, Dai D, et al. A survey on in-context learning. arXiv. Preprint posted online on Dec 31, 2022. [doi: 10.48550/arXiv.2301.00234]
28. Kojima T, Gu SS, Reid M, Matsuo Y, Iwasawa Y. Large language models are zero-shot reasoners. 2022 Presented at: 36th Conference on Neural Information Processing Systems (NeurIPS 2022); Nov 28 to Dec 9, 2022; New Orleans, LA p. 22199-22213 URL: https://openreview.net/pdf?id=e2TBb5y0yFf [accessed 2025-12-12]
29. Yu F, Quartey L, Schilder F. Exploring the effectiveness of prompt engineering for legal reasoning tasks. Presented at: Findings of the Association for Computational Linguistics; Jul 9-14, 2023; Toronto, Canada p. 13582-13596 URL: https://pdfs.semanticscholar.org/9661/dda80241923420976 47c37423eb7fccee298.pdf [accessed 2025-12-12]
30. Achiam J, Adler S, Agarwal S, et al. GPT-4 technical report. arXiv. Preprint posted online on Mar 15, 2023. [doi: 10.48550/arXiv.2303.08774]
31. Hoffmann J, Borgeaud S, Mensch A, et al. Training compute-optimal large language models. arXiv. Preprint posted online on Mar 29, 2022. [doi: 10.48550/arXiv.2203.15556]
32. Wang JJ, Wang VX. Assessing consistency and reproducibility in the outputs of large language models: evidence across diverse finance and accounting tasks. SSRN. Preprint posted online on Mar 21, 2025. [doi: 10.2139/ssrn.5189069]
33. Burr DB. Muscle strength, bone mass, and age-related bone loss. J Bone Miner Res 1997 Oct;12(10):1547-1551. [doi: 10.1359/jbmr.1997.12.10.1547] [Medline: 9333114]
34. Westerterp KR. Changes in physical activity over the lifespan: impact on body composition and sarcopenic obesity. Obes Rev 2018 Dec;19 Suppl 1:8-13. [doi: 10.1111/obr.12781] [Medline: 30511504]

Abbreviations

AI: artificial intelligence
API: application programming interface
LLM: large language model
MAE: mean absolute error
MAPE: mean absolute percentage error
MPE: mean percentage error
RMSE: root mean squared error

Edited by S Munce; submitted 14.Jul.2025; peer-reviewed by AA Barr, E Fu, VX Wang; accepted 19.Nov.2025; published 30.Jan.2026.

Please cite as:
Chen PQ, Gu HYW, Lo HKY, Chang WC, Lai CJM, Lai SHS, Cheng ASK, Ng PHF
Leveraging Large Language Models for Early Detection of Anomaly Work Injury Cases: Data-Driven Approach to Rehabilitation Efficiency
JMIR Rehabil Assist Technol 2026;13:e80607
URL: https://rehab.jmir.org/2026/1/e80607
doi:10.2196/80607

Using a Co-Designed Digital Self-Management Program to Prepare Patients for Hip or Knee Replacement Surgery: Pragmatic Pilot Study

Elizabeth Horton[1], PhD; Hayley Wright[1], PhD; Andy Turner[1], PhD; Louise Moody[2], PhD; Lucy Aphramor[3], PhD; Anna Carlson[4], PhD; Hesam Ghiasvand[5], PhD; Shea Palmer[6], PhD

[1]Centre for Intelligent Healthcare, Coventry University, Priory Street, Coventry, United Kingdom

[2]Centre for Arts, Memory and Communities, Coventry University, Coventry, United Kingdom

[3]Centre for Agroecology, Water and Resilience, Coventry University, Coventry, United Kingdom

[4]Hope For The Community Community Interest Company, Enterprise Hub, Coventry, United Kingdom

[5]Centre for Healthcare and Communities, Coventry University, Coventry, United Kingdom

[6]School of Healthcare Sciences, Cardiff University, Cardiff, United Kingdom

Corresponding Author:
Elizabeth Horton, PhD
Centre for Intelligent Healthcare, Coventry University, Priory Street, Coventry, United Kingdom

Abstract

Background: The aging population has resulted in more people living longer with musculoskeletal conditions who require hip and knee replacement surgery. Lengthening waiting lists continue to be a challenge for patients and health care services.

Objective: This pragmatic study aimed to develop and test a digital self-management intervention (the HOPE [Help Overcome Problems Effectively] program) to better prepare patients waiting for hip and knee replacement surgery.

Methods: The study used a pragmatic, pre-post with follow-up, single-arm design. All intervention and data collection components were delivered online. Patients were recruited from those on the waiting list for hip or knee surgery. Following iterative co-development of the intervention, the content was refined and optimized into a final version for testing. The resulting program was an 8-week intervention delivered via the HOPE 4 The Community (H4C) digital platform (powered by H4C). Data were collected at baseline (pre-HOPE program), 8 weeks (post-HOPE program), and 6-month follow-up. Patient-reported outcome measures related to preparation for surgery, quality of life, physical function, pain, mental well-being, self-efficacy, and physical activity. Resource usage data were collected to calculate health and social care costs. System Usability Scale data were collected post-HOPE program.

Results: One hundred participants enrolled in the HOPE program. Of these, 57 (57%) consented to take part in the evaluation and returned the baseline questionnaire. Thirty-nine participants completed ≥5 of the 8 sessions and all surveys. Among the 25 participants who had surgery at 6 months, 23 (92%) felt better prepared due to the HOPE program. Median improvements in most outcomes were observed at 8 weeks, with several continuing to improve at 6 months. The Friedman test showed significant improvements over 6 months in self-efficacy (pain: P=.002; other symptoms: P<.01), pain (P=.04), health status (P=.02), and mental well-being (P=.01). No significant changes were noted in physical activity. While the early cost analysis did not reach statistical significance, it indicated potential cost savings from reduced patient interactions with health care professionals. Sixty-four percent (25/39) of participants had surgery, and this likely contributed in part to improvements in outcomes. System usability was rated above average (mean score 70.1, SD 15.9).

Conclusions: The results are promising in relation to participants attending the HOPE program feeling better prepared for surgery. A fully powered efficacy and cost-effectiveness trial is needed to determine the contribution of the HOPE program to outcomes, over and above the contribution of surgery.

(JMIR Rehabil Assist Technol 2026;13:e68286) doi:10.2196/68286

KEYWORDS

prehabilitation; self-management; peer support; digital platform; hip and knee arthroplasty; musculoskeletal

Introduction

In the United Kingdom, an estimated 20.3 million people are affected by musculoskeletal conditions. These conditions account for 21% of the years lived with illness and disability [1]. The global prevalence of osteoarthritis is increasing, and if the trend continues, osteoarthritis will become one of the most prevalent diseases in populations from high-income countries in the coming decades [2]. The aging UK population is living longer with complex musculoskeletal conditions and comorbidities, causing increased demand on National Health Service (NHS) health and social care services [1], accounting for up to 30% of general practice consultations in England [3].

Lengthening waiting lists are particularly problematic in musculoskeletal medicine. A 2019 report found that in England alone, 726,000 people had severe hip osteoarthritis and 1.4 million people had severe knee osteoarthritis [4]. For those whose condition is severe, joint replacement surgery is the only option to alleviate pain and improve mobility and the ability to self-manage. Under the NHS constitution, 92% of patients should be treated within 18 weeks as part of the referral-to-treatment scheme. However, in 2019, nearly 4000 patients had been waiting for over 2 years for surgery [4], and more than 690,000 were on waiting lists in 2021 [5]. The COVID-19 pandemic had an unprecedented impact on secondary care orthopedic services, with a significant increase in waiting times for the majority of patients [5]. While on the waiting list, patients are likely to experience worsening pain, reduced mobility, increased anxiety, and deteriorating health, leading to greater demand for health and care services. In recognition of wait times, Versus Arthritis [5] and Arthritis Action [6] offer resources for self-management on their websites. By 2060, it is projected that the demand for hip and knee joint replacements in the United Kingdom will rise by nearly 40% from current levels, which will have significant implications for the health care system [7].

New ways of working are needed to optimize support for patients, maximize capacity, and mitigate risk. It is also important to address inequities: the COVID-19 pandemic foregrounded deep-rooted equality, diversity, and inclusion issues in relation to morbidity and mortality that are entangled with access to health care services. Inequities in treatment waiting time [8] for musculoskeletal services and in treatment outcomes [9] reflect this general picture and highlight the need for action. There is a need for holistic support among those waiting for hip and knee surgery in England. The NHS personalized care team recommends that patients on the waiting list should receive self-management support to “wait well” by undertaking prehabilitation. This support should empower patients through information, health coaching, and digital resources [10].

Prehabilitation is an effective way of improving perioperative outcomes through support to increase physical and mental resilience for surgery. Systematic reviews and meta-analyses have found some, generally low-quality evidence that prehabilitation improves a range of postoperative outcomes for patients undergoing hip and knee surgery, including function, pain, strength, and quality of life [11-13]. A more recent systematic review and meta-analysis specifically focused on the effects of digital prehabilitation in a range of musculoskeletal conditions awaiting surgery, including knee and hip replacements [14]. They found evidence that advanced technologies supported greater improvements in function pre- and post-operatively than standard care for knee and hip replacements. Greater improvements were also seen in preoperative pain, preoperative risk of falling, and postoperative stiffness. There was no evidence for spinal surgery or other conditions. However, few orthopedic prehabilitation interventions are digitally delivered, nor do they provide peer or emotional support, which is highly valued by many patients living with long-term conditions [11-13]. Indeed, a recent survey conducted in the United Kingdom [15] found that, although the vast majority of hospitals (97%) offered preoperative education, only 59% and 48% offered prehabilitation for knee and hip arthroplasty, respectively. Education was mainly delivered as a single talk supported by a booklet, and prehabilitation mainly as strengthening exercise, advice, and written information. Reported barriers included lack of facilities, funding, and staff. There was also a reported lack of robust evidence to support practice [15]. Across various surgical specialties, multimodal prehabilitation includes nutrition and psychological support alongside exercise training. There is some evidence of psychological factors improving postsurgical outcomes [13]. A systematic review and meta-analysis found low-quality evidence that psychological interventions have a positive effect on postsurgical anxiety and on mental components of quality of life [16].

In a review [17] of over 30 prehabilitation surgery schools in the United Kingdom and Ireland (these schools inform patients about what to expect and guide them on how to prepare physically and mentally to reduce postoperative risks of surgery), only 40% contained content to manage emotional well-being, and only 13% used digital apps. Further, many interventions were not underpinned by behavior change theory and techniques.

In 2022, Coventry University and its university spin-out social enterprise, H4C (HOPE [Help Overcome Problems Effectively] for The Community) interest company, developed a proof-of-concept digital intervention, called the Help Overcome Problems Effectively (HOPE) program, to help patients prepare for hip and knee surgery. The HOPE program for hip and knee patients shares the same underlying theoretical framework as other HOPE programs for long-term conditions offered by H4C, which have been taxonomized using the taxonomy of self-management support [18] and are described in detail in published papers [19-21]. All 14 digital versions of the HOPE program have been approved by the Quality Institute for Self-Management Education and Training for the provision of self-management structured education (QIS2020 and QIS2023 [22]) and certified by the Organisation for the Review of Care and Health Apps (ORCHA [23]), scoring 88% for Android and iOS (Apple Inc), and 86% for WebApp, indicating compliance with best practice in data security, professional assurance, usability, and accessibility.

The HOPE program for hip and knee patients combines evidence-based self-management content with a validated exercise program, incorporating a home exercise component tailored to individual needs and abilities, drawing from the work of Ageberg et al [24].

In 2023 H4C was awarded funding through the UK Research and Innovation Healthy Ageing Challenge Scaling Social Ventures competition to co-design and evaluate the HOPE program for hip and knee surgery patients. The funding competition was to support social enterprises in scaling products and services to support healthy aging and deliver social value.

The pragmatic, multimethod study aimed to optimize and evaluate the HOPE program to determine whether patients were better prepared for surgery. The study objectives included optimizing the HOPE program through co-design with stakeholders, implementing and testing the program with patients waiting for a joint replacement, and assessing their preparedness for surgery.

Methods

Study Design

This study used a co-design phase followed by a pragmatic, pre-post, with follow-up, single-arm intervention study. All intervention components and data collection were delivered online. This study is reported according to the CONSORT (Consolidated Standards of Reporting Trials) 2016 statement: extension for nonrandomized pilot trials [25]. CHERRIES (Checklist for Reporting Results of Internet E-Surveys) was used to guide the survey report [26]. All intervention and data collection activities took place online. All study data were collected online via questionnaires administered through Qualtrics Survey Software (Qualtrics).

Co-Design Phase to Optimize the HOPE Program

Ten participants took part in the development activities, which included 3 online workshops. One workshop was undertaken with 6 patient participants waiting for a hip (n=3) or knee replacement surgery (n=3), who had completed an earlier proof-of-concept HOPE program (5 female participants, aged 60 - 80 years). The purpose of the workshop was to explore their experiences and generate feedback on the HOPE program.

Two health professional workshops involving 4 NHS staff from our partner organizations were held to discuss referral pathways and useful resources for patients awaiting surgery. The roles of the professionals were Elective Recovery Lead, Team Lead Physiotherapist in Elective Orthopedics, Project Manager of a Musculoskeletal Clinical Program, and Senior Primary and Community Care Lead. Workshops and interviews were conducted online via Zoom (Zoom Video Communications, Inc) and MS Teams (Microsoft Corp) to allow for geographically dispersed participation.

Development of the Exercise Program

The exercise program central to the intervention was based on the The Neuromuscular Exercise training program for patients with knee or hip osteoarthritis assigned for total joint replacement the neuromuscular exercise training program for patients with knee or hip osteoarthritis assigned for total joint replacement program [24,27], which was specifically developed for older patients with severe knee and hip osteoarthritis before having total joint replacement surgery. Only the exercises from the neuromuscular exercise training program for patients with knee or hip osteoarthritis assigned for total joint replacement program were adopted within the HOPE program. Those exercises have also been incorporated into the Good Life with osteoArthritis: Denmark program [28-30]. The exercises have been demonstrated to be safe, patients can successfully progress them, and they contribute to improvements in a range of outcomes, including symptoms, function, medication use, and sick leave. A range of video and visual resources had previously been developed to support the exercise components [31]. Following feedback from the co-design phase, new video resources were developed to illustrate how the exercises could be adapted within the home environment. Forty-three videos were filmed in a home setting (living room, bedroom, and kitchen), using home furniture (sofa, chair, and bed) and both exercise equipment and everyday household items as exercise props, with volunteers representing different ages and genders, and incorporating visual prompts and voiceover instructions. The exercises target major lower limb muscle groups and can be adapted to individual capabilities, with 3 difficulty levels and encouragement to alter repetitions and sets. Participants could build their own home-based exercise program by answering 6 questions about their ability (eg, if they can easily get on and off the floor) and equipment (eg, if they have a step they can use at home). An algorithm was then built to create their personalized exercise program from the 43 videos. Participants progressed up and down levels of difficulty at their own pace, monitored progress, and set exercise reminders. Participants could download their exercise record in PDF format to keep or share with a health care professional. Tips on creating a safe exercise space, as well as important information to mitigate any worries or injuries, were part of the program.

The HOPE program: Intervention Content and Structure

The resulting program comprised 8 modules and was hosted on H4C's digital platform, powered by H4C. The content comprised text, images, videos, downloadable documents, interactive activities (eg, quizzes, self-monitoring tools, and diaries), and discussion forums and messaging facilities. The digital content was released at set times over the 8 weeks but could be accessed at any time (asynchronous). Participants had the option to "fast-track" the content if they were due to have surgery during the 8 weeks.

Once accessed and viewed, the app content could be viewed offline, reducing the need for a data plan or high-quality internet connection. An analog print booklet was produced, containing the same content as the digital version of the HOPE program, for those who were digitally excluded and/or experienced low digital literacy.

Pre-Post With Follow-Up Study

Participants

Broad eligibility criteria were used to ensure the study was as inclusive as possible and to provide ample opportunity for participation. Individuals were eligible if they were adults aged 18 years or older, lived in the South West of England in the United Kingdom, were currently on a waiting list for hip or knee replacement surgery, had access to the internet and a suitable device to engage with the intervention, and were able to interact with all materials provided as part of the intervention.

Patients interested in attending the HOPE program were referred to the study sign-up webpage through several routes. NHS South West referral sources included secondary care, primary care, and musculoskeletal clinics. Eligible participants were referred directly to H4C to enroll in the HOPE program and given the option to take part in the research study. Patients who chose to take part in the study were directed to the participant information sheet and consent form in Qualtrics Survey Software. Patients were informed that participation was voluntary and that their decision would not affect their quality of care.

We collected the following sociodemographic information: name, email address, gender, age, postcode, occupation and employment, and some details about their emotional health and their illness diagnosis, level of physical activity, health care visits, time on the waiting list, and date of surgery. Postcode data were used to calculate the English index of multiple deprivation (IMD [32]). IMD is an official measure of deprivation ranked from 1 (most deprived) to 10 (least deprived).

The questionnaire was administered through Qualtrics, using responsive and mobile-ready question formats. Adaptive questioning was used to conditionally display questions based on previous responses to reduce the number and complexity of the questions. Most pages contained between 1 and 6 items. Excluding the introduction, participant information sheet, and consent form, the survey was distributed over 14 pages. The responses were made mandatory to avoid missing data. The survey was not set up to allow participants to change their responses. The procedure, as outlined in the participant information sheet and survey structure, involved collecting identifiable information at registration—specifically, name and email address (rather than via technical means such as cookies or IP addresses)—which was then used by the research team to ensure each individual only completed the survey once per time point. Pre-HOPE program (baseline) questionnaires were completed during the period of July 6-13, 2023, for the first HOPE program and July 20-31, 2023, for the second HOPE program. Participants received a £60 (approximately US $80) electronic gift voucher for completion of all pre- and postprogram questionnaires. Participants were informed in the Patient Information Sheet how their data would be processed in accordance with the Data Protection Act 2018. Participation in the study was optional for patients who accessed the HOPE program.

The HOPE Program: Accessing and Completing the Program

Following completion of the pre-HOPE program survey, participants were given access to the HOPE program (start dates: July 13 or 27, 2023) through a personalized log-in link.

Throughout the program, participants were supported by 2 facilitators who were trained in line with Quality Institute for Self-Management Education and Training standards. The program content was organized into themed sessions across the 8 weeks, with an integrated tailored exercise program (described in the “Development of the Exercise Program” section above; Table 1 lists session content; refer to Multimedia Appendix 1 for a brief description of each session and screenshots of the intervention).

Table . Session content of the HOPE[a] program.

Session	Session content
1	Instilling HOPE
2	Managing pain and fatigue
3	Stress and shifting your thinking
4	Communication
5	Sleep and mindfulness
6	Setbacks and hospital stay
7	Happiness and strengths
8	Moving on with HOPE

[a]HOPE: Help Overcome Problems Effectively.

Patient-Reported Outcome Measures

Surgery Preparation

At 6-month follow-up, participants were asked if they felt better prepared for surgery using the following question from the Patient Preparedness for Surgery questionnaire [33]: “Overall, I feel or felt (if I had surgery) prepared for my upcoming surgery.” There were 6 response options: strongly agree, agree, somewhat agree, somewhat disagree, disagree, and strongly disagree. Participants were also asked to provide reasons for their answers. Those who had surgery indicated whether they felt the HOPE program helped them prepare before surgery, after surgery, or both. Participants provided textual responses

to explain why they agreed or disagreed that the HOPE program helped them prepare for surgery.

The following validated patient-reported outcome measures (PROMs) were collected at baseline, post-HOPE program (8 weeks), and 6-month follow-up via Qualtrics.

Short Warwick-Edinburgh Mental Well-Being Scale

The Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS [34]) is a short version that assesses mental well-being within the adult population. The SWEMWBS uses 7 items from the full WEMWBS [35], which relate more to mental functioning than feelings. The 7 statements are positively worded, with 5 response categories ranging from "none of the time" to "all of the time." Total scores range from 7 to 35, with higher scores indicating higher mental well-being. A change of one point or more on the SWEMWBS total score represents a minimally important level of change.

The EQ-5D Index and EQ-Visual Analogue Scale

The EQ-5D index [36] and the EQ-Visual Analogue Scale (EQ-VAS) are widely used measures of health status and health-related quality of life, respectively. The EQ-5D index assesses patients' health state across 5 dimensions (self-care, mobility, anxiety and depression, usual activities, and pain and discomfort) that are weighted to provide a utility value based on a population tariff. Scores range from 0 (death) to 100 (perfect health). The EQ-VAS is a vertical rating scale for health, scored between 0 (worst imaginable health) and 100 (best imaginable health).

Western Ontario and McMaster Universities Arthritis Index

The Western Ontario and McMaster Universities Arthritis Index (WOMAC [37]) consists of 24 items divided into 3 subscales: Pain (5 items), Stiffness (2 items), and Physical Function (17 items). Items are scored on a scale of 0 - 4, which corresponds to: None (0), Mild (1), Moderate (2), Severe (3), and Extreme (4). The scores for each subscale are summed, with possible score ranges of 0 - 20 for Pain, 0 - 8 for Stiffness, and 0 - 68 for Physical Function. A sum of the scores for all 3 subscales gives a total WOMAC score (maximum 96). Higher scores indicate worse pain, stiffness, and functional limitations.

Arthritis Self-Efficacy Scale

The Arthritis Self-Efficacy Scale (ASES [38]) measures a person's confidence to self-manage their arthritis symptoms and consists of 2 subscales: Pain (5 items) and Other Symptoms (6 items). Items are scored from 1 (very uncertain) to 10 (very certain). The scores for each subscale are summed, with a possible score range of 10 - 50 for Pain and 10 - 60 for Other Symptoms. Higher scores indicate higher self-efficacy.

International Physical Activity Questionnaire–Short Form

The International Physical Activity Questionnaire–Short Form (IPAQ-SF [39]) assesses physical activity undertaken across a comprehensive set of domains including: (1) leisure-time physical activity, (2) domestic and gardening (yard) activities, (3) work-related physical activity, and (4) transport-related physical activity. The items are structured to provide separate scores on walking, moderate-intensity, and vigorous-intensity activity, as well as a combined total score to describe the overall level of activity. Computation of the total score requires summation of the duration (in minutes) and frequency (days) of walking, moderate-intensity, and vigorous-intensity activity. The IPAQ-SF scoring protocol assigns the following metabolic equivalent of task (MET) values to walking, moderate, and vigorous-intensity activity: 3.3 METs, 4.0 METs, and 8.0 METs, respectively. Participants are considered to have met Centers for Disease Control and Prevention and American College of Sports Medicine physical activity recommendations if they reported at least 150 minutes per week of walking, moderate, or vigorous intensity physical activity.

Numerical Pain Rating Scale

The Numerical Rating Scale (NPRS)-11 [40] is an 11-point scale for self-report of pain. It is the most commonly used unidimensional pain scale. The respondent selects a whole number (integers 0 - 10) that best reflects the intensity (or other quality, if requested) of their pain. The anchors are 0=no pain and 10=worst possible pain (there are various wordings of the upper anchor).

HOPE Program Usability and User Engagement

Usability

The usability of the system was assessed by the System Usability Scale (SUS [41]), which was embedded in the last session of the HOPE program. It was optional for participants to complete. The SUS uses a 5-point Likert scale ranging from 1=strongly disagree to 5=strongly agree across 10 items. Odd-numbered questions (1, 3, 5, 7, and 9) generate a positive response. Even-numbered questions (2, 4, 6, 8, and 10) generate a negative response, which must be inverted. All the points added up together could gain a maximum of 40, thus the multiplication by 2.5 to make the scale out of 100. A total score of ≥68 is considered above-average usability.

User Engagement

The intervention platform collected user engagement data. For this study, we report the number of sessions completed, the number of participants who used the personalized exercise program, and the most commonly bookmarked content or activities.

Sample Size

This pragmatic study enrolled participants from an opportunity sample (n=100) comprising eligible candidates. Potential participants received an email containing a link to the study website hosted by Qualtrics. Here, participants were required to review the digital Participant Information Sheet, provide digital consent, and complete the digital questionnaires.

Analytical Methods

Data relating to sociodemographic characteristics and outcome measures were collated and presented descriptively at the group level. Outcome data were mostly ordinal and nonnormally distributed, so descriptive data were limited to frequencies (and proportions) and medians (and IQRs). While the study was not powered to detect statistically significant changes in outcomes between time points, nonparametric Friedman and post hoc Wilcoxon signed-rank tests were used to explore changes over

time between baseline, post-HOPE program (8 weeks), and 6-month follow-up. All analyses were performed using IBM SPSS (version 28). The level of statistical significance was set at $P<.05$. Textual responses to the question about surgery preparedness at the 6-month follow-up survey were summarized to illustrate the quantitative findings.

Given this was a feasibility study with complete-case analysis as the prespecified approach, we focused on participants who engaged with ≥5 sessions and completed follow-ups. This decision was made because (1) the primary aim was assessing intervention feasibility and acceptability under optimal conditions, (2) minimal data were available from noncompleters (only 4/15 provided follow-ups), and (3) high follow-up rates among completers (98% at 8 weeks and 93% at 6 months) reduced concerns about attrition bias. Future efficacy trials will use intention-to-treat (ITT) analysis.

Resource Usage

An early cost-impact analysis evaluated the change in costs associated with patients' appointments and visits with NHS England to understand the potential cost impact of the program and assess whether it could be expanded into a broader study. These data were captured via the Qualtrics survey at baseline, post-HOPE program (8 weeks), and 6-month follow-up.

The economic analysis focused on changes in the number of interactions patients had with NHS health and social care staff, measured by appointments and visits. A decision model was developed using parameters from a before-and-after analysis, literature review, and incorporating assumptions. The mean values, associated SEs, and assumptions populated the model, detailed in Multimedia Appendix 2. The total cost impact was calculated from the NHS personal and social care perspective, both per patient and per patient per week.

Costs associated with interaction changes were evaluated at 8 weeks and 6 months compared to baseline using unit costs from the Unit Costs of Health and Social Care report by the Personal Social Services Research Unit at the University of Kent [42] and the NHS National Tariff [43]. A probabilistic sensitivity analysis explored uncertainty around the results.

Ethical Considerations

The user requirements research undertaken by Coventry University received ethical approval from the Coventry University Research Ethics Committee (P151751). The research and evaluation activity has also received approval from Coventry University (P106036) and, as an amendment to a preexisting HOPE evaluation, from the Health Research Authority and Health and Care Research Wales (Integrated Research Applications System, project ID 283172).

Results

Co-Design Phase Adaptations

Adaptations to the intervention, as an outcome of patient and health professionals' feedback, were as follows: Adaptations suggested by patients were (1) guidance on how to adjust the exercises to meet individual needs and capabilities; (2) a broader range of additional activities to try, for example, pool-based exercises; (3) reassurance for people who may struggle to keep up with the program; (4) information to challenge misinformation, controversies, and conflicting advice; and (5) clearer guidance on how to access some features, for example, messaging functions.

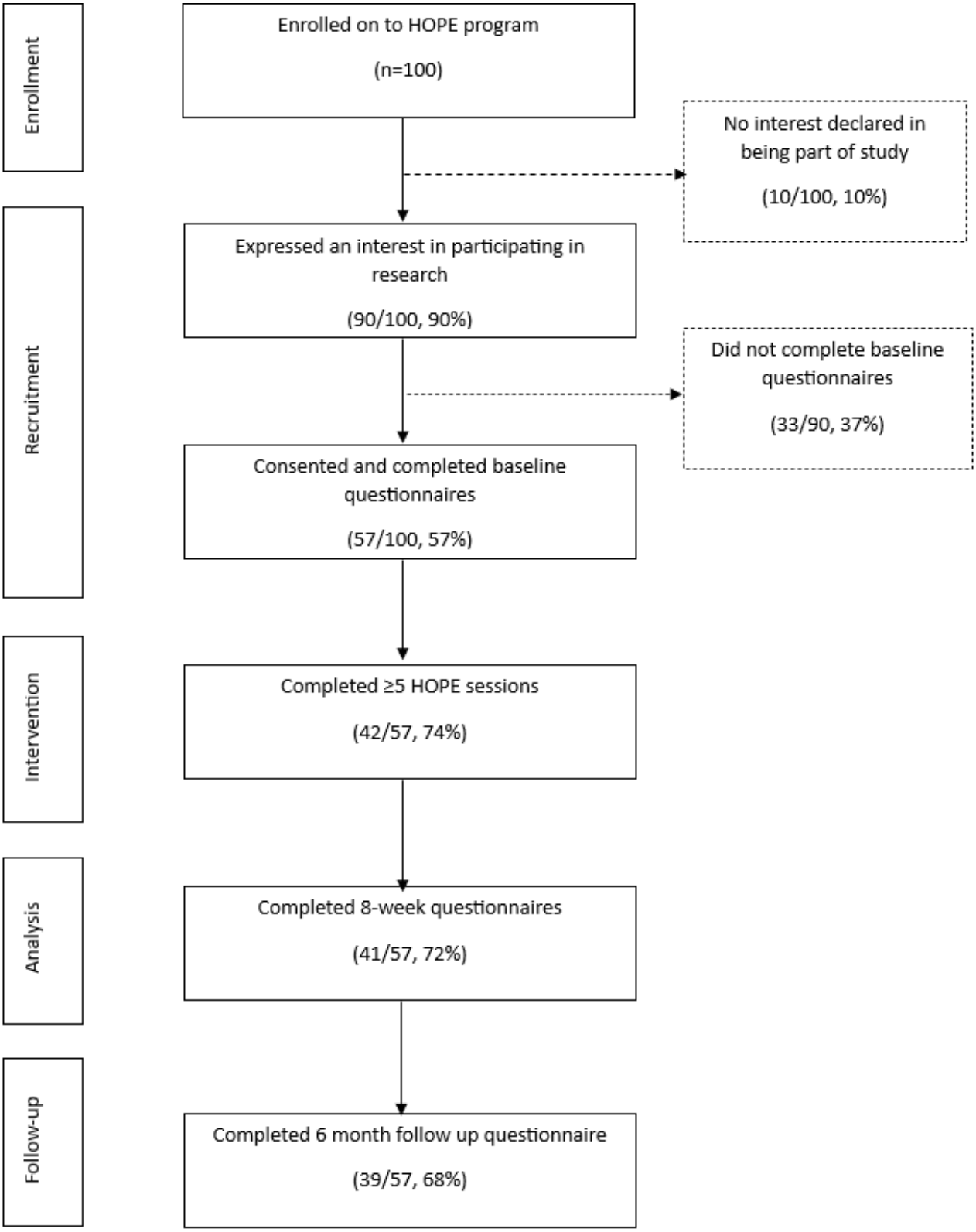
Adaptations suggested by health professionals were (1) reminders and nudges to make healthy changes and prevent deconditioning, (2) long-term access to information for use postoperatively, (3) adjustment of exercises to cater to different abilities and comorbidities, (4) program certification to demonstrate credibility, and (5) reference to the expert input that informed the program content.

Participants

One hundred participants enrolled in the 2 HOPE programs (HOPE 1: n=59 and HOPE 2: n=41). Of these, 57 (57%) consented to take part in the evaluation and returned the baseline questionnaire (n=39, HOPE program 1 and n=18, HOPE program 2). Forty-one participants returned follow-up questionnaires at 8 weeks (41/57, 71.9%), and 39 participants returned questionnaires at 6-month follow-up (39/57, 68.4%). Forty-two participants (42/57, 73.7%) accessed ≥5 of the 8 sessions and were considered program completers.

Almost all of the HOPE program completers (41/42, 98%) returned follow-up questionnaires at 8 weeks, and 39 (93%) returned questionnaires at 6-month follow-up (Figure 1). HOPE program completers who returned both questionnaires (39/42, 93%) were included in the primary analysis. There was no missing outcome data, as these fields were required during questionnaire completion.

Figure 1. Flow of participants through the study. HOPE: Help Overcome Problems Effectively.



Participant characteristics are presented in Table 2 and are similar in the total sample (n=57) and completers (n=39). All completer participants (n=39) identified as White-English, Welsh, Scottish, Northern Irish, or British ethnicity and described English as their first language (all 39/39, 100%). One-third of participants were male (13/39, 33%) and two-thirds were female (26/39, 67%). Age was only reported by 21 (54%) participants, with a median age of 66.0 (IQR 63.0-69.5) years. The majority of participants were retired (23/39, 59%). A third of participants (13/39, 33%) were listed for hip replacement surgery, and two-thirds (26/39, 67%) for knee replacement surgery. The median IMD was 7.00 (IQR 2.5-13) and the median time on the waiting list for surgery was 6.00 (IQR 2-12) months.

Table . Participant baseline characteristics of completers (n=39) and total sample (N=57).

Characteristic	Completers (n=39)	Total sample (N=57)
Gender, n (%)		
Male	13 (33)	20 (35)
Female	26 (67)	36 (63)
Not specified	0 (0)	1 (2)
Age (years), median (IQR)	66 (63-69.5)	66 (63-69.5)
Ethnicity, n (%)		
White-English, Welsh, Scottish, Northern Irish, or British	39 (100)	56 (98)
Black, African, or Caribbean	0 (0)	1 (2)
Disability, n (%)		
Mental health condition (long-term)	5 (13)	7 (12)
Blind or partially sighted	1 (3)	1 (2)
Hard of hearing or deaf	0 (0)	1 (2)
Long-term illness or health condition (lasting more than 12 months or terminal)	4 (10)	7 (12)
Mobility impairment	24 (62)	32 (56)
Employment, n (%)		
In paid work: full-time	4 (10)	8 (14)
In paid work: part-time	4 (10)	8 (14)
Retired	23 (59)	31 (54)
Not in paid work	8 (21)	10 (18)
Not in paid work due to hip or knee condition?	8 (21)	10 (18)
Index of multiple deprivation, median (IQR)	7 (2.25)	7 (2)
Joint replacement, n (%)		
Hip	13 (33)	22 (39)
Knee	26 (67)	35 (61)
Waiting time (months), median (IQR)	6 (2-12)	7 (2.5-13)

Twenty-four out of 39 (62%) participants considered themselves to have a disability, with 9 (23%) participants reporting that day-to-day activities were “limited a little,” and 15 (39%) reporting that activities were “limited a lot.” Seven (18%) participants reported more than one specific type of disability (refer to Table 2).

User Engagement

Just over half of all participants completed all 8 sessions (30/57, 53%), with 6 participants completing 7 sessions (6/57, 11%), 1 completing 6 sessions (1/57, 2%), and a further 5 participants completing 5 sessions (5/57, 9%). Forty-nine out of 57 (86%) participants used the personalized exercise program. The top 5 bookmarked content or activities were (1) exercise program, (2) relaxed breathing, (3) mindfulness meditation, (4) compassionate approach to pain, and (5) cognitive diffusion activity.

Patient-Reported Outcomes and Estimations

By the time of the 6-month follow-up, 25 out of 39 (64%) participants had already received their surgery. Of those who had their surgery, the majority (23/25; 92%) agreed with the statement: “As a result of attending the HOPE program, overall, I felt better prepared for my surgery.” Eight out of 23 (35%) participants selected “strongly agree,” 10 (44%) selected “agree,” and 5 (22%) selected “somewhat agree.” Of the 23 participants who agreed that they were better prepared, 16 (70%) felt better prepared in the presurgery period, 3 (13%) felt better prepared postsurgery, and 5 (17%) felt better prepared pre- and postsurgery.

Of those who had not yet had surgery, the majority (13/14, 93%) agreed with the statement: “As a result of attending the HOPE program, overall, I feel better prepared for my surgery.” Of these, 1 participant (1/14, 7%) selected “strongly agree,” 7 (n=7/14, 50%) selected “agree,” and 5 (5/23, 30%) selected “somewhat agree.”

All 39 participants who completed the 6-month follow-up questionnaire provided reasons why they agreed or disagreed that the HOPE program helped them prepare for surgery. The findings are presented under 4 headings: personalized exercise, physical and mental preparation, peer support, and nothing new. Participant ID numbers 1 - 14 are those that were still waiting for surgery at 6-month follow-up, and IDs 15-39 are participants who had undergone surgery. No harms or unintended consequences were reported during the study.

Personalized Exercise

The program offered exercises that helped patients improve their physical condition and overall preparedness for surgery.

> *Better exercised and with better muscle definition.* [ID19]
>
> *It gave me some exercises to prepare for surgery.* [ID29]
>
> *It encouraged me to do the preparation exercises and helped lift my mood when needed.* [ID20]

Physical and Mental Preparation

Patients found the program beneficial for preparing both physically and mentally for surgery. It provided information and insight about what to expect before and after surgery, helping to manage pain, reduce anxiety, increase hope, and plan for the future.

> *The program gave me an insight into what to expect during and after the procedure.* [ID18]
>
> *I found the information useful and the relaxation techniques particularly helpful.* [ID38]
>
> *The information given was clear about the future after the operation.* [ID7]
>
> *I feel I manage pain better even if it becomes more painful.* [ID8]
>
> *I knew so much about what to expect, and I learned techniques to calm any anxiety.* [ID33]

Peer Support

Connecting with others who have arthritis and are waiting for surgery made patients feel less isolated. The program offered a platform for discussing shared challenges, such as surgery delays and recovery expectations, fostering a sense of community among participants. Participants valued the emotional support they received through the program. Sharing experiences with others who were undergoing similar surgeries provided comfort, while insights into the surgical process helped ease fears.

> *Hearing what other arthritis sufferers are going through made you feel that you are not alone in dealing with the pain.* [ID1]
>
> *The HOPE program gave me the opportunity to share my thoughts/fears with others who had either had their joint surgery or were waiting to undergo it.* [ID17]

Nothing New

A few participants found that the program covered what they already knew or that they already had a positive mindset.

> *I haven't found out anything new.* [ID9]
>
> *I already had a very positive view of how to deal with the issues arising from my arthritis.* [ID10]

Usability

Only 16 participants completed the optional SUS. Participants reported a mean SUS score of 70.1 (SD 15.9; range 50 - 95). The 10-item frequency response data are provided in Multimedia Appendix 3. Compared with the 23 participants who did not complete the SUS, the 16 SUS completers were younger (median age of 64, IQR 8 vs 67, IQR 6 years; sample size n=8 and n=13, respectively), included a higher proportion of males (44% vs 26%), and more knee surgery patients (75% vs 60%) with a mobility impairment (69% vs 50%). Other patient characteristics were broadly similar. On average, the 16 completers had slightly lower disease severity: total WOMAC median 49 (IQR 23) versus 53 (IQR 17).

Table 3 summarizes the patient-reported outcomes at baseline, 8 weeks (post-HOPE program), and 6-month follow-up. Median values suggested potential improvements in many outcome measures at the end of the HOPE program (8 weeks). There were sustained improvements in median values for several outcomes at 6 months.

Table . Summary of baseline, post–Help Overcome Problems Effectively (HOPE) program (8 weeks), and 6-month follow-up outcomes (n=39).

Outcome variable	Baseline, median (IQR)	8 weeks, median (IQR)	*P* value, Wilcoxon test (baseline to 8 week)	6 months, median (IQR)	*P* value, Friedman test
Arthritis Self-Efficacy Scale (ASES)					
Confidence to manage pain (1 - 10, ↑=better)	3.8 (2.0-5.6)	4.6 (3.6-5.4)	.07	5.6 (3.8-8.2)	.002[a]
Confidence to manage other symptoms (1 - 10, ↑=better)	4.5 (2.5-5.5)	5.0 (4.2-6.5)	.001[a]	6.5 (4.2-8.3)	<.001[a]
Western Ontario and McMaster Osteoarthritis Index (WOMAC)					
Pain (0 - 20, ↑=worse)	10.0 (8.0-12.0)	10.0 (7.0-13.0)	.19	8.0 (5.0-12.0)	.0[a]4
Stiffness (0 - 8, ↑=worse)	4.0 (4.0-5.0)	4.0 (4.0-5.0)	.92	4.0 (2.0-5.0)	.11
Physical functioning (0 - 68, ↑=worse)	35.0 (26.0-41.0)	34.0 (24.0-42.0)	.44	29.0 (10.0-42.0)	.07
Total (0 - 96, ↑=worse)	49.0 (40.0-58.0)	48.0 (33.0-59.0)	.38	39.0 (19.0-60.0)	.09
Numerical Pain Rating Scale (NPRS)					
Pain (0 - 10, ↑=worse)	6.0 (5.0-7.0)	5.0 (4.0-7.0)	.19	5.0 (2.0-6.0)	.002[a]
EQ-5D					
Quality of Life (EQ-VAS[c]; 0 - 100, ↑=better)	58.0 (35.0-80.0)	60.0 (35.0-80.0)	.37	70.0 (45.0-85.0)	.05
Health status (EQ-Index; 0 - 1, ↑=better)	0.62 (0.30-0.74)	0.60 (0.23-0.78)	.54	0.75 (0.30-0.83)	.02[a]
Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS)					
Mental well-being (SWEMWBS; 5 - 35, ↑=better)	25.0 (21.0-28.0)	25.0 (22.0-28.0)	.86	27.0 (23.0-29.0)	.01[a]
International Physical Activity Questionnaire–Short Form (IPAQ-SF)					
Total (MET-min/wk, ↑=better)	2340 (393-7464)	2628 (480-6152)	.98	2079 (306-5988)	.55
Sitting time (min/d, ↑=worse)	360 (285-480)	360 (181-540)	.41	300 (240-480)	.15
Inactive (<600 MET[b]-min/wk), n (%)	12 (30.8)	10 (25.6)	—[d]	12 (30.8)	—[d]

[a]Statistically significant *P*<05. Data is for participants who participated in ≥5 sessions and completed both follow-up questionnaires (n=39).

[b]MET: metabolic equivalent.

[c]EQ-VAS: EQ-Visual Analogue Scale.

[d]Not applicable.

The Friedman test indicated several potential improvements across the 6 months. Median scores for ASES pain were 3.8 (IQR 2.0-5.6), 4.6 (IQR 3.6-5.4), and 5.6 (IQR 3.8-8.2) at baseline, 8 weeks, and 6-month follow-up, respectively (*P*=.002); ASES other symptoms: 4.5 (IQR 2.5-5.5), 5.0 (IQR 4.2-6.5), and 6.5 (IQR 4.2-8.3*; P*<.001); WOMAC pain: 10.0 (IQR 8.0-12.0), 10.0 (IQR 7.0-13.0), and 8.0 (IQR 5.0-12.0*; P*=.04); NRPS: 6.0 (IQR 5.0-7.0), 5.0 (IQR 4.0-7.0), and 5.0 (IQR 2.0-6.0*; P*=.002); EQ-index: 0.62 (IQR 0.3-0.74), 0.60 (IQR 0.23-0.78), and 0.75 (IQR 0.30-0.83*; P*=.02); and SWEMWBS: 25.0 (IQR 21.0-28.0), 25 (IQR 22.0-28.0), and 27.0 (IQR 23.0-29.0*; P*=.01). Separate Wilcoxon tests at 8 weeks (immediately following the end of the HOPE program) found that only ASES other symptoms was statistically significant (*P*=.001; refer to Table 3).

Ancillary Analyses

Assessment of Bias: Program Completers Versus Program Noncompleters at Baseline

A total of 15 participants were categorized as noncompleters of the Hope program (ie, completing <5 of 8 sessions). Only 4 (27%) of these participants returned both follow-up questionnaires, which was insufficient for meaningful analysis.

Therefore, bias assessment was conducted using baseline data only. Potential differences at baseline between noncompleters (<5 sessions; n=15) and completers (≥5 sessions; n=42) were explored using descriptive statistics. There were no obvious differences between noncompleters and completers in age (median 65.50, IQR 64-72.75 vs median 66.00, IQR 63.0-69.0 years, respectively) or IMD (median 7.00, IQR 5.0-8.0 vs median 7.00, IQR 6.0-8.25). There were slight differences between noncompleters and completers in gender (male: 50% vs 31%), ethnicity (White: 93% vs 100%), disability (yes: 53% vs 62%), employment (in paid work—full time and part time: 54% vs 20%), joint replacement (knee: 47% vs 67%), and waiting time (7.00, IQR 4.0-16.0 vs 6.00, IQR 2.0-13.0 months). On average, noncompleters had slightly greater disease severity. For example, noncompleters reported more pain (NPRS: median 7.00, IQR 6.0-8.0 vs median 6.00, IQR 5.0-7.0, respectively), had a higher total WOMAC score (median 60.00, IQR 40.0-71.0 vs median 49.00, IQR 38.75-58.25), and a lower EQ-5D index value (median 0.45, IQR 0.18-0.73 and median 0.636, IQR 0.29-0.74).

Impact of Surgery on Outcomes

To understand the potential impact of surgery on outcomes, an exploratory descriptive comparison between those who had and had not received surgery at 6 months was made (data presented in Multimedia Appendix 4). This comparison was only based on the ASES data, since improvements in this outcome were statistically significant at both 8 weeks and 6 months. The results show that those who had received surgery at 6 months had larger median improvements in self-efficacy (for both pain and other symptoms). Those who had not had surgery showed marginal improvements in self-efficacy for pain and for other symptoms at 8 weeks. These were maintained at 6 months for pain self-efficacy but not for other symptoms.

Health Care Resource Usage

The component of the study that focused on resource use for this early cost-impact analysis had a sample size of 39 patients, who completed the web-based questionnaire at all 3 time points: baseline, after 8 weeks, and after 6 months. Of these, 25 patients had their surgical intervention within the period covered (ie, within 6 months) and were therefore excluded from the analysis, leaving a total sample size of 14 analyzed. Results are provided in Table 4 (total cost-impact per patient) and Table 5 (cost-impact per patient per week). Cost-impact per patient per week evaluation revealed overall cost savings over 8 weeks as well as over 6 months, but this failed to reach statistical significance. Face-to-face general practitioner interactions at the 6-month interval showed a statistically significant change. Further details of the economic analysis are provided in Multimedia Appendix 2.

Table . Total cost impact per patient (£per week).

Cost category[a]	Cost change from baseline to 8 weeks, mean (95% CI)	Cost change from baseline to 6 months, mean (95% CI)
Face-to-face visit with a physiotherapist	12.85 (–3.80 to 36.02)	–8.64 (–76.34 to 27.29)
Remote visit with GP[b]	–7.72 (–19.55 to 0.36)	0.22 (–3.38 to 4.07)
Face-to-face visit with GP	9.66 (–12.57 to 34.00)	14.09 (–5.39 to 36.29)
Face-to-face hospital visit	–7.46 (–29.65 to 10.52)	–8.91 (–38.35 to 10.84)
Total	7.34 (46.49 to –27.61)	–3.25 (46.10 to –87.02)

[a]Positive values correspond to cost savings.

[b]GP: general practitioner.

Table . Cost impact per patient (£ per week).

Cost category[a]	Cost change from baseline to 8 weeks, mean (95% CI)	Cost change from baseline to 6 months, mean (95% CI)
Face-to-face visit with physiotherapist	1.61 (–0.48 to 4.5)	0.69 (–3.34 to 3.94)
Remote visit with [c]GP	–0.96 (–2.44 to 0.04)	0.20 (–0.1 to 0.63)
Face-to-face visit with GP	1.21 (–1.57 to 4.25)	2.45 (0.50 to 5.08)[b]
Face-to-face hospital visit	–0.93 (–3.71 to 1.31)	–0.03 (–1.85 to 1.82)
Total	0.92 (–3.45 to 5.81)	3.31 (–1.93 to 8.12)

[a]Positive values correspond to cost savings.

[b]Statistically significant (P<.05).

[c]GP: general pratitioner.

Sample Size Calculation for Future Trial

Data collected as part of the current evaluation were used to inform likely sample sizes for future studies in this area. This sample size calculation was based on ASES-8 data. Unfortunately, the minimum clinically important difference of the ASES-8 is unknown [44]. However, it is sensitive to change, with an effect size of 0.31 previously reported for the ASES-8

following interdisciplinary group therapy for fibromyalgia [45]. Moderate effect sizes of this magnitude are common for conservative interventions in musculoskeletal conditions. In this pilot study, the mean and SD values for ASES pain and ASES other symptoms at baseline were 3.98 (SD 1.93) and 4.29 (SD 2.08), respectively. Assuming a 1-tailed hypothesis, an effect size of 0.3, α=.05, 90% power, and a 1:1 allocation ratio, 191 participants would be required in each group (N=382) to detect a ≥0.58-point difference in ASES pain and ≥0.62-point difference in ASES other symptoms.

Discussion

Principal Findings

This study evaluated the HOPE program, a digital self-management intervention designed to support patients awaiting hip and knee replacement surgery. Results from 39 completers suggested potential improvements in self-efficacy, pain, health status, and mental well-being over 6 months. Most participants felt better prepared for surgery, and the program was rated above average for usability (mean SUS score 70.1).

Participant feedback revealed some key areas that underscore the program's potential usefulness. Some participants appreciated the targeted exercises that improved their physical and mental readiness for surgery. The program provided comprehensive information about the surgical process, helping patients manage pain, reduce anxiety, and plan for the future. Studies have shown that patients have difficulties remembering information immediately after deciding to undergo surgery [46]. Having access to digital information, which can be regularly and quickly updated with evidence-based information, is a useful resource for patients. By fostering a sense of community, the program helped some participants connect with others facing similar challenges. However, some participants noted that the program offered nothing new, as they already enjoyed a positive mindset or previous knowledge.

The demographic profile of completers (median age 66, IQR 63-69.5 years; 100% White; and 66.7% female) was almost identical to a recent UK study, which found that digital health coaching delivered to patients waiting for lower limb arthroplasty improved patient activation and reduced length of hospital stay [47]. It should be noted that noncompleters of the program were more likely to be male, in paid employment, and awaiting a hip replacement.

Engagement with the HOPE program was high, with 73.7% (42/57) of participants attending ≥5 of 8 sessions. Follow-up and engagement rates were lower when based on the 100 participants who enrolled: 39% (39/100) completed the 6-month follow-up questionnaires, and 42% (42/100) who completed ≥5 sessions. Among those who completed all study procedures, 93% (39/42) engaged with the program.

A recent national digital attitudes and behavior survey conducted in the United Kingdom by ORCHA in 2023 described the willingness of older respondents to use digital apps for self-monitoring, symptom tracking, and managing recovery [48].

At the 6-month follow-up, nearly two-thirds (25/39, 64%) of participants had undergone surgery. More than 90% (23/25) of these participants agreed that the program helped them prepare better for surgery. Statistically significant median improvements in most PROMs were evident at the end of the HOPE program, and several scores continued to improve at 6-month follow-up, including self-efficacy, pain, health status, and mental well-being. The exercise program was the most bookmarked page, and despite the majority of participants (49/57, 86%) starting the personalized exercise program, there were no improvements in time spent sitting or in the proportion of participants classified as inactive. The exercise program may require greater input from facilitators to encourage optimum engagement. Research shows that exercise supervision involving trained physical therapists improves compliance with exercises, especially in older adults [13,49]. Alternatively, it may be that the IPAQ-SF lacks sensitivity to adequately assess physical activity [39]. More objective measures of physical activity, such as accelerometry, could be considered in future research.

The high number of participants undergoing surgery makes it challenging to attribute potential improvements in PROMs to either intervention. In their systematic review and meta-analysis, Punnoose et al [13] showed that variability in surgical procedures can influence postoperative recovery; therefore, postsurgical improvements cannot be attributable solely to prehabilitation. Owing to the often degenerative nature of musculoskeletal conditions, potential improvements in PROMs in this study were not anticipated a priori. Rather, it was hypothesized that attending the HOPE program would slow the rate of decline through the acquisition of effective self-management and coping strategies. Thus, the observed trend for median improvements across the majority of PROMs is encouraging.

Resource Usage

This early cost analysis suggests that the HOPE program may lead to a reduction in patient interactions with care professionals at both 8 weeks and 6 months. However, the small sample size results in wide CIs, which limits the reliability of these findings and affirms the need for further studies to assess the cost-effectiveness of the program. Despite this limitation, the initial results highlight the potential for the HOPE program to offer cost-saving benefits at a societal level.

Strengths and Limitations

A strength of this real-world study was the inclusion of participants with lived experience at all stages of the project, providing input into the HOPE program intervention development process and follow-up feedback to optimize it for further studies. The majority of participants started the exercise program, which is a cornerstone of prehabilitation. Other strengths include the use of validated PROMs, high levels of engagement with the intervention, and good survey completion rates at 6 months. This version of the HOPE program was rapidly developed and deployed by adding new musculoskeletal content to an existing taxonomized evidence-based intervention. Some of the health professionals involved in the co-design workshops suggested that patients needing only conservative management and not requiring surgery would also benefit from

the program. Our co-creation and intervention development process could develop and test a program for these patients and for other groups of nonorthopedic presurgery patients. The powered by H4C platform currently hosts more than 15 digital self-management and health interventions. Using a single platform to deliver multiple interventions and modules offers several advantages for funders, researchers, health care providers, and patients. Many patients live with comorbid conditions requiring diverse information and self-management techniques. Platform delivery can incorporate and streamline self-management support. Torous and Vaidyam [50] asserted that "instead of a plethora of apps, there is a need for a few that meet the needs of many." Drawing on successful examples from the automobile, space, and clean energy sector, Ansar and Flyvbjerg [51] outline the benefits of platforms over one-off designs, such as repeatability, extendibility, absorptive and adaptive capacity, resulting in "faster, better, cheaper" services and products. They concluded that sectors such as health "are ripe for a platform rethink." Another strength of this application is the partnership between a social enterprise company and an academic institution. A recent Wellcome report [52] recommended that companies, including nonprofits, can be better at developing and scaling digital health solutions than university research groups.

The limitations of the study include the small number (16/39) of participants who completed the SUS. It is possible that these participants had a more positive user experience compared to those who did not complete the scale. Among the 39 study completers, most participants (>90%) agreed that the program helped them prepare better for surgery, and the textual responses supporting this question provided limited feedback. A broader set of feedback questions and/or postprogram qualitative interviews or focus groups analyzed using rigorous and transparent methods—with participants who did not complete the program—could elicit more critical or negative experiences.

The self-selecting nature of recruitment may have resulted in participants who were inherently more inclined to seek assistance or engage in self-help efforts.

Without a control group comparator, it is not possible to directly attribute any change in the PROMs to the HOPE program. It is important to note that many improvements were not statistically significant, and the statistical analyses performed were likely to be underpowered. Furthermore, a recent systematic review of hip arthroplasty prehabilitation interventions suggested that measures such as the WOMAC may not be the most appropriate measure to detect differences and suggest alternative objective measures such as the chair rise test, gait speed, or stair climbing [53]. That review also found that more than 8 weeks of prehabilitation was associated with improved outcomes, suggesting that future trials of the HOPE program should consider extending the length of the intervention. While our completer analysis provides valuable proof-of-concept data, it limits generalizability to real-world implementation, where attrition is typically higher. The baseline differences between completers and noncompleters suggest our effect estimates may be optimistic. Future randomized controlled trials (RCTs) should combine ITT and per-protocol analyses to distinguish efficacy from effectiveness.

Separating the effects of the intervention from the effects of surgery is problematic. The ancillary analysis of the ASES data (refer to Table 4) suggests that surgery was probably a major contributor to improvements in self-efficacy at 6 months. This is not surprising, given that the excellent outcomes of hip replacement surgery have led to the procedure being described by The Lancet as the "operation of the century" [54]. Approximately 96.2% and 90.8% of patients have previously reported satisfaction with their hip and knee replacement surgery, respectively [55].

A future definitive RCT should be appropriately powered to directly compare an intervention group (ie, the HOPE program) against an appropriate control group (ie, treatment as usual). Subgroup analysis should compare PROMs in those who have had, or are still awaiting, surgery at 6-month follow-up. Such a design would help to distinguish the effects of the intervention from the effects of surgery.

The baseline data show that program noncompleters (ie, those who completed <5 sessions) had slightly greater disease severity at baseline than program completers. Owing to limited follow-up data, it is not known whether these participants could not complete the program due to factors relating to their musculoskeletal condition, their experience of the program, or random intervening factors. Nonresponders were also more likely to be male, in paid employment, and awaiting a hip replacement. Such findings raise questions about how to engage people with greater disease severity and these sociodemographic characteristics in future support programs. Further research is needed to understand individual needs and how they change as disease and pain progress, and to determine how best to support individuals through targeted interventions.

In line with the wealth of other UK health care research studies, the participant sample in this study lacked diversity in terms of ethnicity and socioeconomic characteristics. The study sample reflects the demographics of NHS waiting lists and can be understood as a manifestation of structural inequalities. People living in the most deprived areas of the United Kingdom are more likely to require replacement surgery but less likely to receive it [56] and less likely to have good outcomes [57,58] compared with those living in the least deprived areas. This recurring finding underscores the need for research into the impact of structural barriers to self-management, which may, in turn, suggest the need for more options or a new paradigm approach. Health care interventions that disproportionately meet the needs of nonmarginalized groups embed injustice by widening health inequity. The earlier statement that no harm was reported during the study holds when "harm" is understood within the parameters of evidence-based medicine and its associated framework of biomedical ethics. However, when a framing such as distributive justice is applied, the intervention may be associated with unintended adverse consequences that emerge from and perpetuate ideologies such as structural racism and classism. Lack of attention to unintended harm linked to the lack of diversity in self-management research highlights the need for an expanded ethical framework informed by disability justice scholarship [59]. Recommendations from a recent report into musculoskeletal health inequalities in the United Kingdom included prioritizing surgery and self-management support for

patients living in the most deprived areas [60]. More effort is required to understand the needs of and actively recruit these groups of participants in future self-management trials. A national digital attitudes and behavior survey conducted in the United Kingdom by ORCHA in 2021 [61] found that advocacy for digital health apps was highest among people of Black African heritage (89%), followed by Asian (80%), and then White (64%) respondents. Studies from the United States highlight the importance of recruiting low-income and ethnic minority participants, showing that these groups are more willing to attend [62] and engage more [63] with health interventions compared with White participants in higher-income groups. However, data from this study show that deprivation levels were similar between HOPE program completers and noncompleters.

Conclusion

The results are promising in relation to the acceptability of a peer-supported self-management program for people awaiting hip or knee surgery. Overall, participants felt better prepared for surgery. Textual feedback was generally positive, and participants attributed improvements in their mental and physical well-being to techniques they learned in the HOPE program. However, comparing self-efficacy in those who had and had not received surgery suggests that surgery might have been a more important agent of change than the HOPE program. Overall, the study has demonstrated potential benefit and no evidence of harm or unintended consequences. A randomized controlled efficacy and cost-effectiveness trial design, involving a socioeconomically and ethnically representative sample, is required to delineate the effects attributable to the HOPE program, as opposed to effects of having surgery or natural variation in PROMs. While these preliminary results are promising, they require confirmation in a fully powered RCT using ITT analysis to account for real-world attrition patterns.

Acknowledgments

The authors would like to thank the participants involved in the co-creation workshops, stakeholders, partners, and course participants. The HOPE program is delivered by the Coventry University spin-out social enterprise, Hope 4 The Community (H4C) CIC. We would also like to thank Health Tech Enterprise for contributing to the health economic analysis. The project, Hope and Prepare Effectively for Surgery (HoPES; 10053263), was funded by UK Research and Innovation as part of the Healthy Ageing Challenge Scaling Social Ventures competition, delivered by Innovate UK. Funding was awarded to H4C through the Small Business Research Initiative Social Ventures competition, run by UK Research and Innovation as part of the Healthy Ageing Challenge, delivered by Innovate UK. This pilot study was not registered with a clinical trials registry. There was no published protocol for this pilot study.

Funding

No external financial support or grants were received for this work.

Conflicts of Interest

AT is a co-founder and director of H4C and the co-inventor of the original HOPE program. AC is the head of Partnerships and Projects at H4C.

Multimedia Appendix 1
Intervention screenshot and session content.
[DOCX File, 560 KB - rehab_v13i1e68286_app1.docx]

Multimedia Appendix 2
Health economic evaluation analysis.
[DOCX File, 40 KB - rehab_v13i1e68286_app2.docx]

Multimedia Appendix 3
System Usability Scale scores.
[DOCX File, 22 KB - rehab_v13i1e68286_app3.docx]

Multimedia Appendix 4
Changes in Arthritis Self-Efficacy Scale.
[DOCX File, 29 KB - rehab_v13i1e68286_app4.docx]

Checklist 1
CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist.
[DOCX File, 28 KB - rehab_v13i1e68286_app5.docx]

Checklist 2
CONSORT-EHEALTH checklist.
[PDF File, 683 KB - rehab_v13i1e68286_app6.pdf]

References

1. The state of musculoskeletal health 2021: arthritis and other musculoskeletal conditions in numbers. Arthritis UK. URL: https://www.versusarthritis.org/media/24238/state-of-msk-health-2021.pdf [accessed 2025-12-12]
2. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. The Lancet 2019 Apr;393(10182):1745-1759. [doi: 10.1016/S0140-6736(19)30417-9]
3. Musculoskeletal health. NHS England. 2020. URL: https://www.england.nhs.uk/elective-care-transformation/best-practice-solutions/musculoskeletal/ [accessed 2024-08-07]
4. Versus arthritis position statement on the clinically-led review of NHS access standards. Versus Arthritis. 2019. URL: https://www.versusarthritis.org/media/21674/clinical-standards-review-nov19.pdf [accessed 2024-08-07]
5. What we're doing to tackle waiting times for people with arthritis. Versus Arthritis. 2023 Dec. URL: https://www.versusarthritis.org/news/2023/december/what-we-re-doing-to-tackle-waiting-times-for-people-with-arthritis/ [accessed 2024-08-07]
6. Treatment waiting times. Arthritis Action. 2024. URL: http://www.arthritisaction.org.uk/living-with-arthritis/resource-centre/waiting-times/ [accessed 2024-08-07]
7. Matharu G, Culliford D, Blom A, Judge A. Projections for primary hip and knee replacement surgery up to the year 2060: an analysis based on data from The National Joint Registry for England, Wales, Northern Ireland and the Isle of Man. Ann R Coll Physicians Surg Can 2022 Jun;104(6):443-448. [doi: 10.1308/rcsann.2021.0206]
8. People living in the poorest areas waiting longer for hospital treatment: The King's Fund and Healthwatch England share new analysis. The King's Fund. 2021 Sep. URL: https://www.kingsfund.org.uk/insight-and-analysis/press-releases/poorest-areas-waiting-longer-hospital-treatment#:~:text=New%20analysis%20from%20The%20King's,in%20the%20least%2Ddeprived%20areas [accessed 2024-08-07]
9. Bergstra SA. Health inequalities across patients with early inflammatory arthritis of different ethnicities: what could be the driving factors? Rheumatology (Oxford) 2022 Dec 23;62(1):7-8. [doi: 10.1093/rheumatology/keac383] [Medline: 35786709]
10. Elective recovery planning supporting guidance. NHS England. 2022 Apr. URL: https://www.england.nhs.uk/wp-content/uploads/2021/12/B1269-elective-recovery-planning-supporting-guidance.pdf [accessed 2024-08-07]
11. Vasta S, Papalia R, Torre G, et al. The influence of preoperative physical activity on postoperative outcomes of knee and hip arthroplasty surgery in the elderly: a systematic review. J Clin Med 2020 Mar 31;9(4):969. [doi: 10.3390/jcm9040969] [Medline: 32244426]
12. Wang L, Lee M, Zhang Z, Moodie J, Cheng D, Martin J. Does preoperative rehabilitation for patients planning to undergo joint replacement surgery improve outcomes? A systematic review and meta-analysis of randomised controlled trials. BMJ Open 2016 Feb;6(2):e009857. [doi: 10.1136/bmjopen-2015-009857]
13. Punnoose A, Claydon-Mueller LS, Weiss O, Zhang J, Rushton A, Khanduja V. Prehabilitation for patients undergoing orthopedic surgery: a systematic review and meta-analysis. JAMA Netw Open 2023 Apr 3;6(4):e238050. [doi: 10.1001/jamanetworkopen.2023.8050] [Medline: 37052919]
14. Guida S, Vitale JA, Swinnen E, et al. Effects of prehabilitation with advanced technologies in patients with musculoskeletal diseases waiting for surgery: systematic review and meta-analysis. J Med Internet Res 2024 Dec 12;26:e52943. [doi: 10.2196/52943] [Medline: 39666971]
15. Omar I, Wylde V, Fogg J, Whitehouse M, Bertram W. Pre-operative education and prehabilitation provision for patients undergoing hip and knee replacement: a national survey of current NHS practice. BMC Musculoskelet Disord 2025 Apr 29;26(1):421. [doi: 10.1186/s12891-025-08637-5] [Medline: 40301783]
16. Perry R, Herbert G, Atkinson C, et al. Pre-admission interventions (prehabilitation) to improve outcome after major elective surgery: a systematic review and meta-analysis. BMJ Open 2021 Sep 30;11(9):e050806. [doi: 10.1136/bmjopen-2021-050806] [Medline: 34593498]
17. Fecher-Jones I, Grimmett C, Carter FJ, Conway DH, Levett DZH, Moore JA. Surgery school—who, what, when, and how: results of a national survey of multidisciplinary teams delivering group preoperative education. Perioper Med 2021 Dec;10(1):20. [doi: 10.1186/s13741-021-00188-2]
18. Pearce G, Parke HL, Pinnock H, et al. The PRISMS taxonomy of self-management support: derivation of a novel taxonomy and initial testing of its utility. J Health Serv Res Policy 2016 Apr;21(2):73-82. [doi: 10.1177/1355819615602725] [Medline: 26377727]
19. Martin F, Wright H, Moody L, et al. Help to overcome problems effectively for cancer survivors: development and evaluation of a digital self-management program. J Med Internet Res 2020 May 19;22(5):e17824. [doi: 10.2196/17824] [Medline: 32209529]
20. Wright H, Martin F, Clyne W, et al. A digital self-management program (help to overcome problems effectively) for people living with cancer: feasibility randomized controlled trial. J Med Internet Res 2021 Nov 5;23(11):e28322. [doi: 10.2196/28322] [Medline: 34738912]

21. Wright H, Turner A, Ennis S, et al. Digital peer-supported self-management intervention codesigned by people with long COVID: mixed methods proof-of-concept study. JMIR Form Res 2022 Oct 14;6(10):e41410. [doi: 10.2196/41410] [Medline: 36166651]
22. Accreditation register. Quality Institute for Self Management Education and Training (QISMET). 2024. URL: https://www.qismet.org.uk/accreditation/accreditation-register/ [accessed 2024-08-07]
23. Delivering safe digital health. The Organisation for the Review of Care and Health Apps (ORCHA). 2024. URL: https://orchahealth.com [accessed 2024-08-07]
24. Ageberg E, Link A, Roos EM. Feasibility of neuromuscular training in patients with severe hip or knee OA: the individualized goal-based NEMEX-TJR training program. BMC Musculoskelet Disord 2010 Jun 17;11(1):126. [doi: 10.1186/1471-2474-11-126] [Medline: 20565735]
25. Lancaster GA, Thabane L. Guidelines for reporting non-randomised pilot and feasibility studies. Pilot Feasibility Stud 2019;5(1):114. [doi: 10.1186/s40814-019-0499-1] [Medline: 31608150]
26. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34. [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
27. Ageberg E, Nilsdotter A, Kosek E, Roos EM. Effects of neuromuscular training (NEMEX-TJR) on patient-reported outcomes and physical function in severe primary hip or knee osteoarthritis: a controlled before-and-after study. BMC Musculoskelet Disord 2013 Dec;14(1):232. [doi: 10.1186/1471-2474-14-232]
28. Skou ST, Roos EM. Good Life with osteoArthritis in Denmark (GLA:D™): evidence-based education and supervised neuromuscular exercise delivered by certified physiotherapists nationwide. BMC Musculoskelet Disord 2017 Feb 7;18(1):72. [doi: 10.1186/s12891-017-1439-y] [Medline: 28173795]
29. Roos EM, Barton CJ, Davis AM, et al. GLA:D to have a high-value option for patients with knee and hip arthritis across four continents: Good Life with osteoArthritis from Denmark. Br J Sports Med 2018 Dec;52(24):1544-1545. [doi: 10.1136/bjsports-2017-098904]
30. Barton CJ, Kemp JL, Roos EM, et al. Program evaluation of GLA:D® Australia: physiotherapist training outcomes and effectiveness of implementation for people with knee osteoarthritis. Osteoarthr Cartil Open 2021 Sep;3(3):100175. [doi: 10.1016/j.ocarto.2021.100175] [Medline: 36474815]
31. Goff AJ, De Oliveira Silva D, Ezzat AM, et al. Co-design of the web-based “My Knee” education and self-management toolkit for people with knee osteoarthritis. Digit HEALTH 2023;9:20552076231163810. [doi: 10.1177/20552076231163810] [Medline: 37009308]
32. English indices of deprivation 2019. GOV.UK. 2019 Sep. URL: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019 [accessed 2025-12-12]
33. Kenton K, Pham T, Mueller E, Brubaker L. Patient preparedness: an important predictor of surgical outcome. Am J Obstet Gynecol 2007 Dec;197(6):654. [doi: 10.1016/j.ajog.2007.08.059] [Medline: 18060968]
34. Ng Fat L, Scholes S, Boniface S, Mindell J, Stewart-Brown S. Evaluating and establishing national norms for mental wellbeing using the short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS): findings from the Health Survey for England. Qual Life Res 2017 May;26(5):1129-1144. [doi: 10.1007/s11136-016-1454-8] [Medline: 27853963]
35. Stewart-Brown S, Tennant A, Tennant R, Platt S, Parkinson J, Weich S. Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey. Health Qual Life Outcomes 2009 Feb 19;7(1):15. [doi: 10.1186/1477-7525-7-15] [Medline: 19228398]
36. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001 Jul;33(5):337-343. [doi: 10.3109/07853890109002087] [Medline: 11491192]
37. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988 Dec;15(12):1833-1840. [Medline: 3068365]
38. Lorig K, Chastain RL, Ung E, Shoor S, Holman HR. Development and evaluation of a scale to measure perceived self‐efficacy in people with arthritis. Arthritis & Rheumatism 1989 Jan;32(1):37-44. [doi: 10.1002/anr.1780320107]
39. Lee PH, Macfarlane DJ, Lam TH, Stewart SM. Validity of the international physical activity questionnaire short form (IPAQ-SF): a systematic review. Int J Behav Nutr Phys Act 2011 Oct 21;8(1):115. [doi: 10.1186/1479-5868-8-115] [Medline: 22018588]
40. Stratford PW, Spadoni G. The reliability, consistency, and clinical application of a numeric pain rating scale. Physiotherapy Canada 2001;53:88-91.
41. Lewis JR. The System Usability Scale: past, present, and future. Int J Hum Comput Interact 2018 Jul 3;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
42. Unit costs 2022-2027. University of Kent. URL: https://www.pssru.ac.uk/unitcostsreport/ [accessed 2024-08-07]
43. National tariff payment system documents, annexes and supporting documents. NHS England. 2023. URL: https://www.england.nhs.uk/publication/national-tariff-payment-system-documents-annexes-and-supporting-documents/ [accessed 2023-08-07]
44. Brady TJ. Measures of self-efficacy: Arthritis Self-Efficacy Scale (ASES), Arthritis Self-Efficacy Scale-8 Item (ASES-8), Children’s Arthritis Self-Efficacy Scale (CASE), Chronic Disease Self-Efficacy Scale (CDSES), Parent’s Arthritis

Self-Efficacy Scale (PASE), and Rheumatoid Arthritis Self-Efficacy Scale (RASE). Arthritis Care Res (Hoboken) 2011 Nov;63 Suppl 11:S473-S485. [doi: 10.1002/acr.20567] [Medline: 22588769]
45. Mueller A, Hartmann M, Mueller K, Eich W. Validation of the arthritis self-efficacy short-form scale in German fibromyalgia patients. Eur J Pain 2003;7(2):163-171. [doi: 10.1016/S1090-3801(02)00097-6] [Medline: 12600798]
46. Specht K, Kjaersgaard-Andersen P, Pedersen BD. Patient experience in fast-track hip and knee arthroplasty--a qualitative study. J Clin Nurs 2016 Mar;25(5-6):836-845. [doi: 10.1111/jocn.13121] [Medline: 26708610]
47. Powley N, Tew GA, Durrand J, et al. Digital health coaching to improve patient preparedness for elective lower limb arthroplasty: a quality improvement project. BMJ Open Qual 2023 Dec 7;12(4):e002244. [doi: 10.1136/bmjoq-2022-002244] [Medline: 38061840]
48. Digital health in the UK today, annual national attitudes and behaviour research. ORCA. 2023. URL: https://info.orchahealth.com/digital-health-attitudes-behaviour-2023-report [accessed 2025-12-12]
49. Shubert TE. Evidence-based exercise prescription for balance and falls prevention: a current review of the literature. J Geriatr Phys Ther 2011;34(3):100-108. [doi: 10.1519/JPT.0b013e31822938ac] [Medline: 22267151]
50. Torous J, Vaidyam A. Multiple uses of app instead of using multiple apps - a case for rethinking the digital health technology toolbox. Epidemiol Psychiatr Sci 2020 Jan 31;29:e100. [doi: 10.1017/S2045796020000013] [Medline: 32000876]
51. Ansar A, Flyvbjerg B. How to solve big problems: bespoke versus platform strategies. Oxford Review of Economic Policy 2022 May 18;38(2):338-368. [doi: 10.1093/oxrep/grac009]
52. The Wellcome Trust. Catalytic environments to drive progress in digital mental health research. Wellcome Open Res 2024;9:301. [doi: 10.21955/wellcomeopenres.1115389.1]
53. Widmer P, Oesch P, Bachmann S. Effect of prehabilitation in form of exercise and/or education in patients undergoing total hip arthroplasty on postoperative outcomes-a systematic review. Medicina (Kaunas) 2022 May 30;58(6):742. [doi: 10.3390/medicina58060742] [Medline: 35744005]
54. Learmonth ID, Young C, Rorabeck C. The operation of the century: total hip replacement. The Lancet 2007 Oct;370(9597):1508-1519. [doi: 10.1016/S0140-6736(07)60457-7]
55. Appiah KOB, Khunti K, Kelly BM, et al. Patient‐rated satisfaction and improvement following hip and knee replacements: development of prediction models. Evaluation Clinical Practice 2023 Mar;29(2):300-311. [doi: 10.1111/jep.13767]
56. Cook MJ, Lunt M, Ashcroft DM, Board T, O'Neill TW. The impact of frailty and deprivation on the likelihood of receiving primary total hip and knee arthroplasty among people with hip and knee osteoarthritis. J Frailty Aging 2023;12(4):298-304. [doi: 10.14283/jfa.2023.36] [Medline: 38008980]
57. Clement ND, Muzammil A, MacDonald D, Howie CR, Biant LC. Socioeconomic status affects the early outcome of total hip replacement. J Bone Joint Surg Br 2011 Apr;93-B(4):464-469. [doi: 10.1302/0301-620X.93B4.25717]
58. Clement ND, Jenkins PJ, MacDonald D, et al. Socioeconomic status affects the Oxford knee score and short-form 12 score following total knee replacement. Bone Joint J 2013 Jan;95-B(1):52-58. [doi: 10.1302/0301-620X.95B1.29749] [Medline: 23307673]
59. Piepzna-Samarasinha LL. Care Work: Dreaming Disability Justice: Arsenal Pulp Press; 2018:263.
60. Act now: musculoskeletal health inequalities and deprivation. : Arthritis and Musculoskeletal Alliance (ARMA); 2024 URL: https://arma.uk.net/wp-content/uploads/2024/03/Musculoskeletal-Health-Inequalities-and-Deprivation-report_v07.pdf [accessed 2024-08-07]
61. Digital health in the UK: national attitudes and behaviours. ORCHA. 2021. URL: https://us.orchahealth.com/wp-content/uploads/2021/07/2107_ICS_Research_Report_2021_National_final.pdf [accessed 2025-12-12]
62. Lorenzo-Luaces L, Wasil A, Kacmarek CN, DeRubeis R. Race and socioeconomic status as predictors of willingness to use digital mental health interventions or one-on-one psychotherapy: national survey study. JMIR Form Res 2024 Apr 11;8:e49780. [doi: 10.2196/49780] [Medline: 38602769]
63. Horrell LN, Kneipp SM, Ahn S, et al. Chronic disease self-management education courses: utilization by low-income, middle-aged participants. Int J Equity Health 2017 Jun 27;16(1):114. [doi: 10.1186/s12939-017-0604-0] [Medline: 28655319]

Abbreviations

ASES: Arthritis Self-Efficacy Scale
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
CONSORT: Consolidated Standards of Reporting Trials
EQ-VAS: EQ-Visual Analogue Scale
H4C: HOPE 4 The Community Interest Company
HOPE: Help Overcome Problems Effectively
IMD: index of multiple deprivation
IPAQ-SF: International Physical Activity Questionnaire–Short Form
ITT: intention-to-treat
MET: metabolic equivalent of task
NHS: National Health Service

NPRS: Numerical Pain Rating Scale
ORCHA: Organisation for the Review of Care and Health Apps
PROM: patient-reported outcome measure
RCT: randomized controlled trial
SUS: System Usability Scale
SWEMWBS: Short Warwick-Edinburgh Mental Well-Being Scale
WOMAC: Western Ontario and McMaster Universities Arthritis Index

Edited by S Munce; submitted 01.Nov.2024; peer-reviewed by G Martin-Alemany, J Bailey, P Marier-Deschenes; revised version received 23.Oct.2025; accepted 04.Nov.2025; published 07.Jan.2026.

Please cite as:
Horton E, Wright H, Turner A, Moody L, Aphramor L, Carlson A, Ghiasvand H, Palmer S
Using a Co-Designed Digital Self-Management Program to Prepare Patients for Hip or Knee Replacement Surgery: Pragmatic Pilot Study
JMIR Rehabil Assist Technol 2026;13:e68286
URL: https://rehab.jmir.org/2026/1/e68286
doi:10.2196/68286

Blockchain-Based Mobile App for Digital Identification of Older Adults in Rural Peru: Design and Usability Evaluation Study

Wilver Arana-Ramos*, BSc; Aldo Francisco Pastrana-Leon*, BSc; Juan Carlos Morales-Arevalo*, MBA, MSc

Department of Software Engineering, Faculty of Engineering, Peruvian University of Applied Sciences, Av. Alonso de Molina Nro. 1611, Lima Polo And Hunt Club (Av. Primavera 2390), Lima, Peru

*all authors contributed equally

Corresponding Author:
Juan Carlos Morales-Arevalo, MBA, MSc
Department of Software Engineering, Faculty of Engineering, Peruvian University of Applied Sciences, Av. Alonso de Molina Nro. 1611, Lima Polo And Hunt Club (Av. Primavera 2390), Lima, Peru

Abstract

Background: Older adults in rural areas of Peru encounter many challenges in accessing critical public services, such as health care, education, and social assistance, due to low levels of digital literacy, limited access to technology, and the lack of formalized, secure ID. This inhibits entry into digital health, education, and social assistance systems and increases their risk of vulnerability and social exclusion.

Objective: This study aimed to design a blockchain technology–based mobile app architecture that helps facilitate secure and inclusive digital ID for older adults in rural areas of Peru, enabling access to vital services through a decentralized, privacy-preserving solution.

Methods: This study followed the design thinking process, which consists of five phases: empathize, define, ideate, prototype, and evaluate. A total of 16 adults (aged 61-85 years) were interviewed to determine the usability barriers and security and privacy concerns with mobile technology, which was used to define functional and nonfunctional requirements. These requirements were developed based on the interviews. The primary features the target population valued included blockchain authentication, assisted registration, multilingual functionality, and a user-friendly interface. The features were prioritized and prototyped using the Figma web-based app. The architecture of the app was developed using the C4 model and accounted for sequential development while ensuring scalability, modularity, and decentralization. Usability was assessed quantitatively by administering the System Usability Scale to the same 16 participants after they had interacted with the prototype.

Results: The mean System Usability Scale score was 60.78 (SD 13.68), indicating acceptable usability. The main issues identified were a lack of skills to navigate digital interfaces, concerns regarding data security, and accessibility challenges for people with disabilities. Participants provided high ratings for the assisted registration system and notifications. The modular, blockchain-based system architecture showed substantial potential for scalability and broader inclusion. The prioritization matrix identified that, for adoption, features must incorporate good design, be multilingual, and require secure authentication.

Conclusions: The proposed blockchain-based mobile app offers a viable technical and socially inclusive model for secure digital ID of older adults in underserved contexts. Usability testing suggested that the solution was perceived as secure, usable, and appropriate for the target population. Although not fully deployed, our prototypes and system architecture provide a good starting point for future implementation. The findings in this study can contribute to efforts to facilitate digital inclusion, access to services, and respect for people's autonomy in identity management systems for vulnerable people.

(JMIR Rehabil Assist Technol 2026;13:e79553) doi:10.2196/79553

KEYWORDS

blockchain technology; data privacy; digital divide; essential public service; portable digital identity; older people; user-centric design

Introduction

Digital technologies have become increasingly important in identity management across all sectors of the economy. The proliferation of mobile apps designed to improve access to services has further deepened this relevance [1]. Older adults living in rural areas of Peru face unique challenges, such as higher levels of digital illiteracy and limited public service coverage. These challenges hinder access to essential services such as health care, education, and social support and mirror broader barriers reported for digital health innovations in low- and middle-income health care systems in South and Southeast Asia [2-4]. Advances in digital identity models, such as self-sovereign identities (SSIs), have opened up new possibilities

for improving security, privacy, and access to services [5]. The emergence of identity verification systems that leverage blockchain infrastructure and protocols exemplifies this change, as they ensure there are no gaps in identity verification and authentication processes and allow users to exercise holistic control over their personal information [6].

Nonetheless, traditional identity management systems that typically use centralized data formats and 2-factor authentication often fail older adults in rural settings [7]. Centralized databases are limited, as they rely on dependable connectivity, technical skills, and trust in institutions to function as intended, even though there isn't adequate infrastructure in many rural contexts. In addition, centralized database models present vulnerabilities by creating a single point of failure that makes any district-based database susceptible to unauthorized access or misuse of information. Furthermore, the use of passwords, tokens, and SMS codes presents usability challenges, as older adults may be less versed in the use of mobile devices and may have difficulty remembering and entering credentials [8]. Conversely, blockchain protocols decentralize validation across identifiable participants and enable user control over personal data by creating a single entry point to the data. Collectively, these properties make blockchain protocols safer, more transparent, and more empowering for this population.

These advances in technology go well beyond data protection. They are fundamental to improving efficiency in the delivery of public and private services. For example, health credential verification and the interoperability of electronic health records are relevant in the health care context. The same is true in education and public participation [9], where the use of mobile apps that integrate SSIs could help reduce the digital divide and provide access to vital resources for all [2,3]. However, their implementation remains limited due to the fragmentation of existing platforms and the lack of a widely accepted interoperable standard. In addition, challenges remain related to centralized data sets, transparency in verification, adherence to strong data privacy policies, and the centralized control of information [10].

In this context, the focus of this study is on designing a blockchain-based mobile app architecture to support digital identity verification and access to necessary services. The goal is to create a system that can be implemented quickly and efficiently while respecting appropriate privacy, security, and decentralization practices in relation to structural data [11].

This study follows a user-centered design thinking (DT) approach to ensure that the proposed solutions are accessible, secure, and appropriate for the target population, rural older adults [12]. The research will include a comprehensive review of current solutions, such as blockchain and smart contracts applied to identity systems, in order to identify the most effective methods for improving user interaction and building trust in digital service platforms [13]. This research is qualitative and will combine a detailed literature review with the design and creation of system architecture models that will eventually be applied in the real world. The goal will be to establish a robust framework for mobile services that use advanced identity verification technologies, highlighting the potential of blockchain to ensure data permanence and integrity. The results of this study provide useful approaches for building mobile systems that prioritize accessibility, security, and transparency, with a special focus on improving essential services [13]. This strategy is even more significant for rural communities, as it allows older adults in these communities to access digital ID solutions that respect their privacy and give them control of their personal information.

Therefore, the objective of this study was to design and evaluate the usability of a blockchain-based mobile app architecture that supports secure digital ID and access to essential services for older adults living in rural areas of Peru.

Methods

Ethical Considerations

This research involved 16 older adults (aged 61-85 years) living in rural areas of Peru who voluntarily took part in interviews and usability tests. All participants signed an informed consent form. No sensitive data were collected, and personal information was anonymized and managed in accordance with the principles of confidentiality and data protection.

Formal ethics board (IRB) approval was not required according to institutional guidelines [14], as the study focused on the design and usability evaluation of a digital prototype. The research was non-interventional in nature, involved voluntary adult participants, and did not include clinical procedures or the collection of sensitive personal or health data. All participants provided informed consent prior to participation. The study was conducted in accordance with the principles of the Declaration of Helsinki.

Research Methodology

This research adopted the DT methodology to develop a mobile app. The user-centered design approach is highly effective in digital product development because it supports users' needs at every stage while also ensuring a product is functional and accessible to the target audience [15-18]. Among the several established design frameworks, such as user-centered design, participatory design, and agile user experience, DT was selected because it provides a structured yet flexible process that begins with empathy as its core principle. This characteristic aligns with the study's focus on understanding the lived experiences of older adults in rural Peru, whose limited digital literacy and accessibility challenges require solutions grounded in human context rather than purely technical efficiency. By explicitly situating this work within a lineage of research that views empathy as the foundation for inquiry and intervention, the DT framework allows the research team to translate qualitative insights into practical, socially responsive design decisions [15-17].

The design process was broken into five phases: empathize, define, ideate, prototype, and test, and each phase was important for our design of a blockchain-based mobile product for digital ID and safe access to services. A complete overview of these activities is provided below.

Empathize Phase

A total of 16 older adults aged 61 to 85 years were interviewed to identify some of the key barriers when trying to access digital services or engage with new technologies. The interviews revealed a number of key issues directly related to participants' comfort levels and hesitancy with digital technology, and their anxiety surrounding cybersecurity, which often impaired their use of multiple digital platforms. These findings illustrate that solutions must be technically feasible, but also usable, secure, and easy to navigate [15].

Define Phase

The research team identified important findings from the interviews that influenced the project [19]. Based on the interview findings, the research team identified key user priorities, specifically the importance of a user-friendly interface, transparent privacy control, and user management (input). The study also highlighted the need for ID systems that are achievable and meaningful to users (at all digital literacy levels) [17].

Ideate Phase

At this stage of the research and development process, the team specified some core features of the app, such as authentication methods incorporating blockchain technologies, a security alerts feature, and personalization options for service access that the user sets according to their needs and preferences. The goal was to create a well-integrated solution in which users could access and manage strategic services relevant to them in a secure and efficient manner [18].

Prototype Phase

Prototypes and low- and medium-fidelity mockups were developed on the Figma platform. The prototypes helped to visualize the interface design of the app and observe users' interactions with the app's functions [19]. The technical system architecture was developed in accordance with the C4 model (this study included only "context," "containers," and "components," excluding "code," as we did not model the code level) to allow the system to be scalable and adaptable within a blockchain structure [20-22].

Testing Phase

To quantitatively evaluate prototype app usability, the study applied the System Usability Scale (SUS) developed by Brooke [23] in 1996. The SUS is a well-respected instrument used in usability research, due in part to its simplicity and its provision of a straightforward numerical metric of users' perceptions of system usability [23].

A survey was created using the standard 10-item SUS, in which respondents indicated whether they agreed or disagreed with statements using a 5-point Likert scale from 1 ("strongly disagree") to 5 ("strongly agree"). The full set of questions used is in Textbox 1.

The launch survey was conducted on 16 older adults that were representative of the targeted users, all of whom had previous interactions with the prototype. The sample size matched the guidelines recommended for initial usability testing using the SUS. Scoring was based on subtracting 1 from the participants' response for the odd-numbered items and subtracting the even-numbered item responses from 5. The scores were summed and multiplied by 2.5 to create a final score with a range of 0 to 100, with higher scores representing better usability. User feedback was an important part of the process for improving interface accessibility and ease of use [24]. This cyclical, user-centered approach ensures that the app meets technical specifications while being accessible, safe, and easy to use for older adults with various levels of technological capabilities [16,17].

Textbox 1. Questions on the System Usability Scale.

1. I would like to use this application frequently.
2. I found the application unnecessarily complex.
3. I found the application easy to use.
4. I would need the help of a person with technical knowledge to be able to use this application.
5. The various functions of the application are well integrated.
6. I found too much inconsistency in the application.
7. I imagine most people could learn to use this application quickly.
8. I found the application cumbersome to use.
9. I felt very confident using this application.
10. I learned to use the application quickly.

Results

This section elaborates on the results from the various design phases and the process of developing a mobile app designed for digital identity and secure access to vital services based on blockchain technology, which was achieved using the DT approach. This allowed the team to better understand the needs of the targeted end users, older adults, and ensured that the prototype genuinely reflects the expectations and needs of this group. The major findings and responses presented in each phase are described below.

Empathize Phase: Interview Findings

During the empathize phase, interviews were conducted with 16 older adults, ranging in age from 61 to 85 years, to identify

the barriers and difficulties they face when interacting with mobile apps. The most common problems encountered are presented in Textbox 2.

Based on the findings obtained, two empathy maps were developed to reflect the needs and emotions of users in relation to technology [25,26]. These are presented in Figures 1 and 2, where the first corresponds to users and the second to the local authorities.

Textbox 2. Identified problems.

- Difficulty navigating apps: users indicated that the app interfaces were complex and difficult to understand.
- Data security concerns: older adults expressed distrust about the handling of their personal data on digital platforms.
- Lack of interface customization: some users reported that the apps were not tailored to their individual needs, making adoption difficult.
- Lack of digital literacy: many older adults have difficulty understanding apps due to their unfamiliarity with digital technologies.
- Limited accessibility: the lack of accessibility options for people with visual or motor disabilities made interaction complicated for a segment of the population.

Figure 1. Empathy map: user perspective.

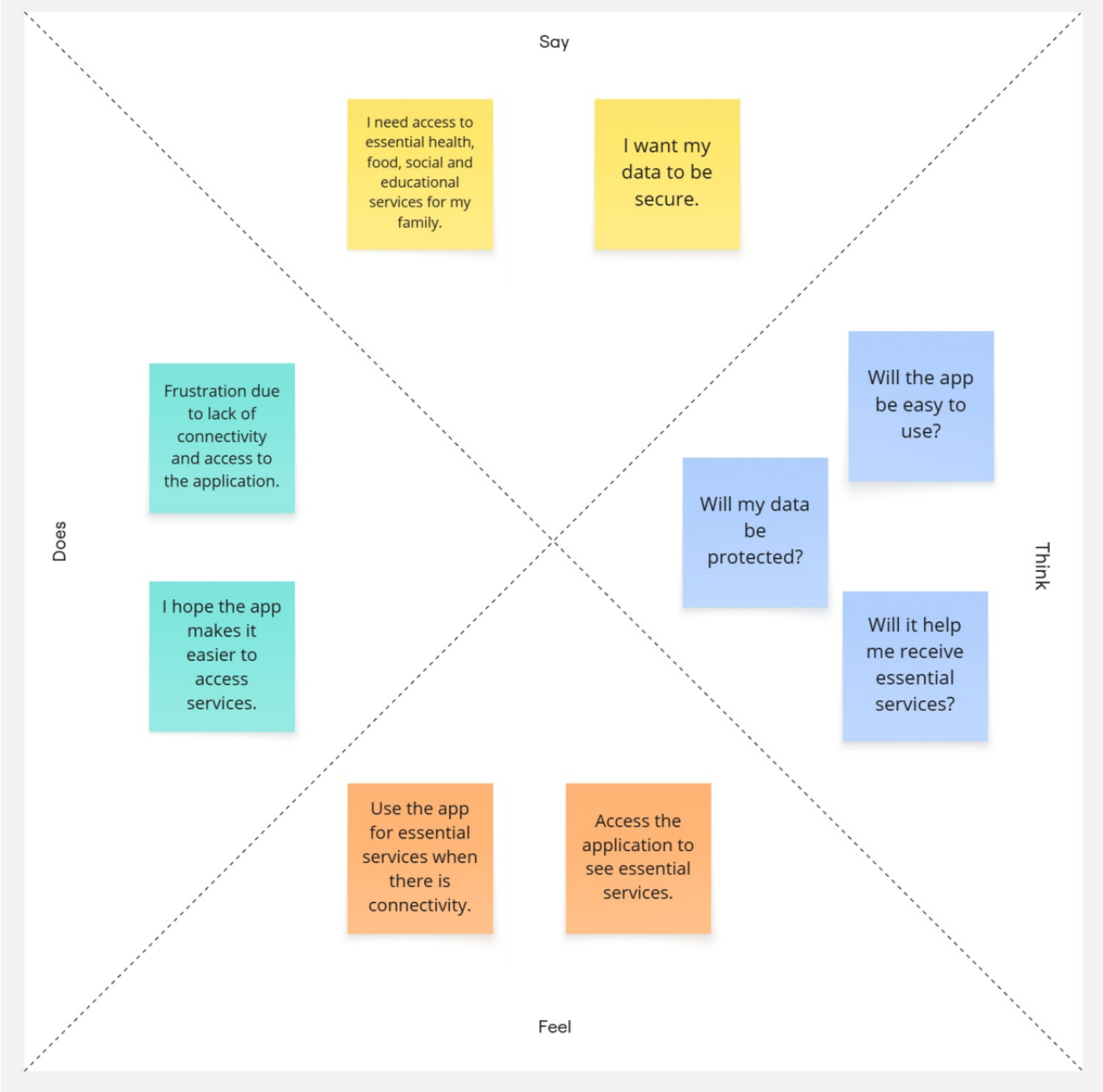
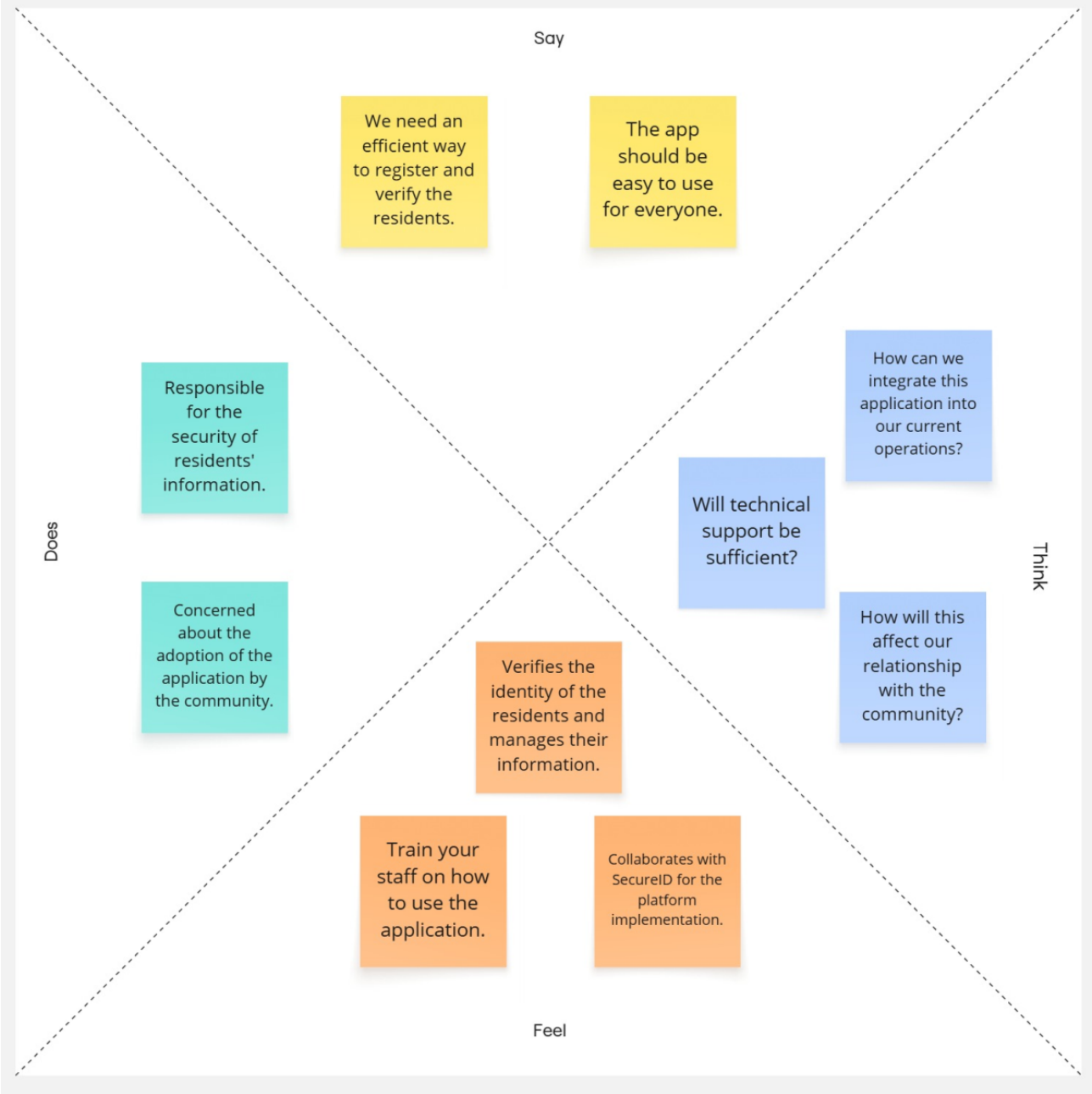


Figure 2. Empathy map: local authorities' perspective.



Define Phase: Functional and Nonfunctional Requirements

Based on the findings obtained in the empathize phase, the functional and nonfunctional requirements for the mobile app were defined. These were established based on the specific needs of the users. The most relevant requirements are presented in Table 1 below.

The following are the most relevant user statements, selected from those collected for their significant contribution to understanding users' expectations regarding interaction with the app and ensuring that their needs are adequately addressed. These user stories are presented in Table 2.

Table . Functional and nonfunctional requirements.

Requirement	Description	Type
Registration in the app	Allows the user to register in the mobile app with the support of a caregiver, enabling access to their digital identification and essential services.	Functional
Digital identification display	Allows the user to securely access their digital identification based on blockchain technology, guaranteeing the authenticity and protection of their personal data.	Functional
Selection of essential health service	Provides the user with the ability to view and select health services available in their locality, facilitating timely access to vital information.	Functional
Information security	The app must ensure the integrity, confidentiality, and authenticity of user data through the use of blockchain technology and robust security protocols.	Nonfunctional
Multilingual support	The app must support multiple languages to facilitate access to services for users who do not speak the platform's main language.	Nonfunctional
Notification system	The system must have a notification controller that manages reminders, alerts, and updates, sending them to users and local authorities.	Functional

Table . Relevant user stories.

User ID	User statement
US-01	As an older adult user, I would like to view my digital identification in order to access my personal data securely.
US-02	As a local authority, I want to filter the list of registered seniors by name or digital identification to quickly locate a specific user and facilitate records management.
US-03	As an older adult user, I would like to register in the application in order to access the available services.
US-04	As an older adult user, I wish to attach and upload images such as a signature, face, or identification to be registered in my account.
US-05	As a local authority, I wish to modify the distribution of essential services.
US-06	As an older adult user, I want to visualize the services provided by the application to easily access the available options.
US-07	As a local authority, I want to view the list of registered seniors to properly track their enrollment and access to available services.
US-08	As a local authority, I want to visualize the services provided by the application to easily access the available options and also have the possibility to add new broadcasts to each selected service.

Ideate Phase: Key Functionalities and Prioritization Matrix

In this phase, the key functionalities needed for the mobile app were identified. The functionalities were prioritized according to their impact on user experience and the effort required for their implementation. These results are presented in Table 3.

Table 4 shows the prioritization matrix, which ranks the key functionalities according to their impact and implementation effort.

Table . Selected key functionalities.

Key functionality	Description
Blockchain authentication	A decentralized authentication system to ensure the security and privacy of user data.
Intuitive interface	Simple and accessible design, with customization options to enhance the experience for older adults.
Assisted registration system	Provides personalized assistance during the registration process to ensure the inclusion of older adults who are not fully technologically autonomous.
Multilingual support	The app must support multiple languages to facilitate access to services for users who do not speak the platform's main language.
Notification system	The system should send notifications to remind users about important events or updates related to services, such as health care, education, and social assistance.

Table . Functionality prioritization matrix.

Functionality	Impact	Effort	Priority
Blockchain authentication	High	Medium	High
Intuitive Interface	High	High	Medium
Assisted registration system	High	Medium	High
Multilingual support	High	Medium	High
Notification system	High	Medium	High

Prototyping Phase: C4 Model, User Flow, and Mockups

During the prototyping phase, a comprehensive representation of the system was developed, covering both the technical architecture and the user experience. For this purpose, the C4 model was applied, which allows for structuring the system architecture at different levels of abstraction. Figure 3 shows the context diagram, which illustrates the general interaction between the mobile app, the users (older adults and local authorities), and the external identity verification services. Figure 4 presents the container diagram, which decomposes the system into its main modules: mobile interface, cloud backend, and decentralized database, along with the technologies used. Multimedia Appendix 1 presents the component diagram, which describes the internal elements of the back end, such as the authentication manager, the service controller, and the notification module.

Figure 3. C4 model context diagram.

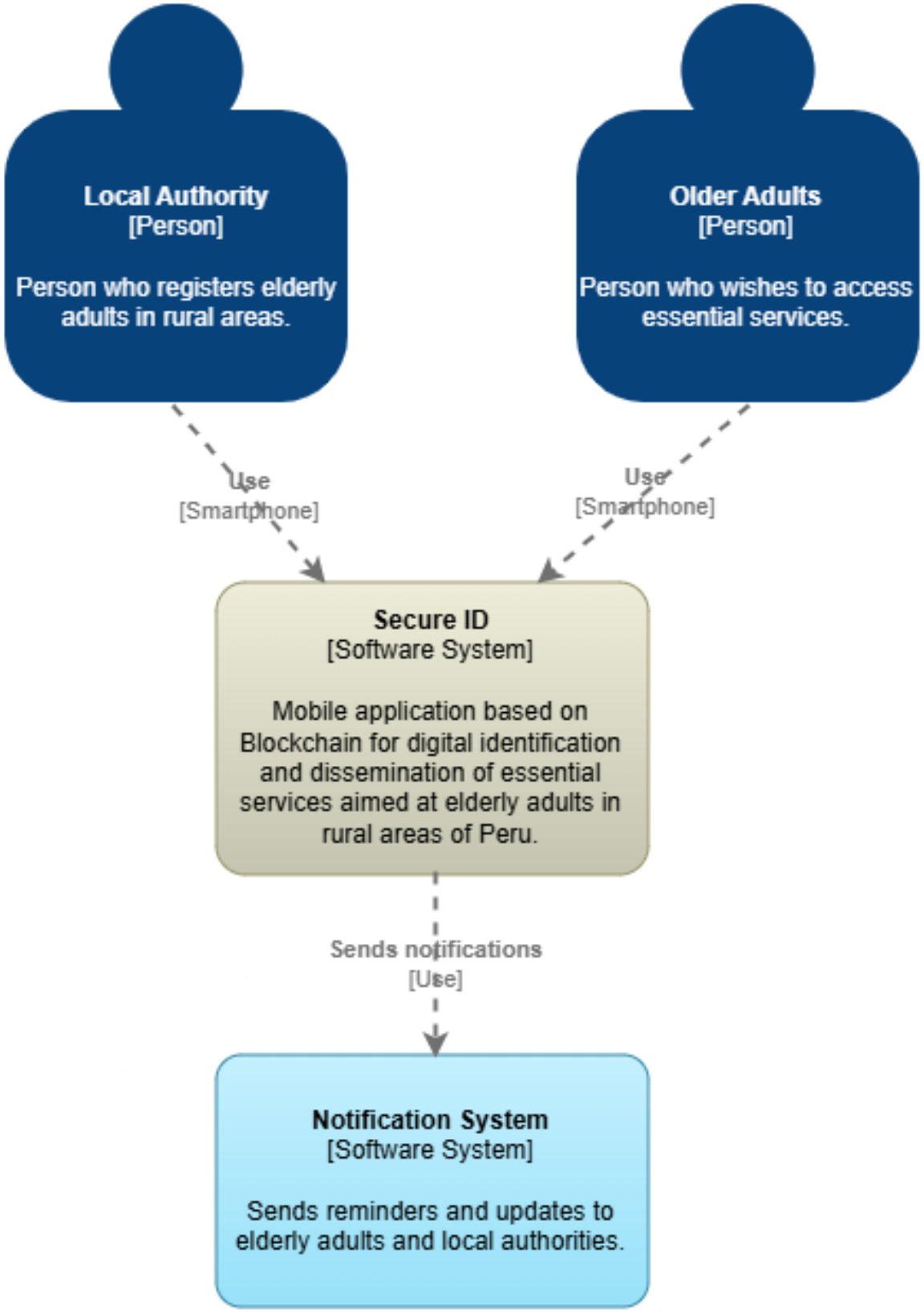
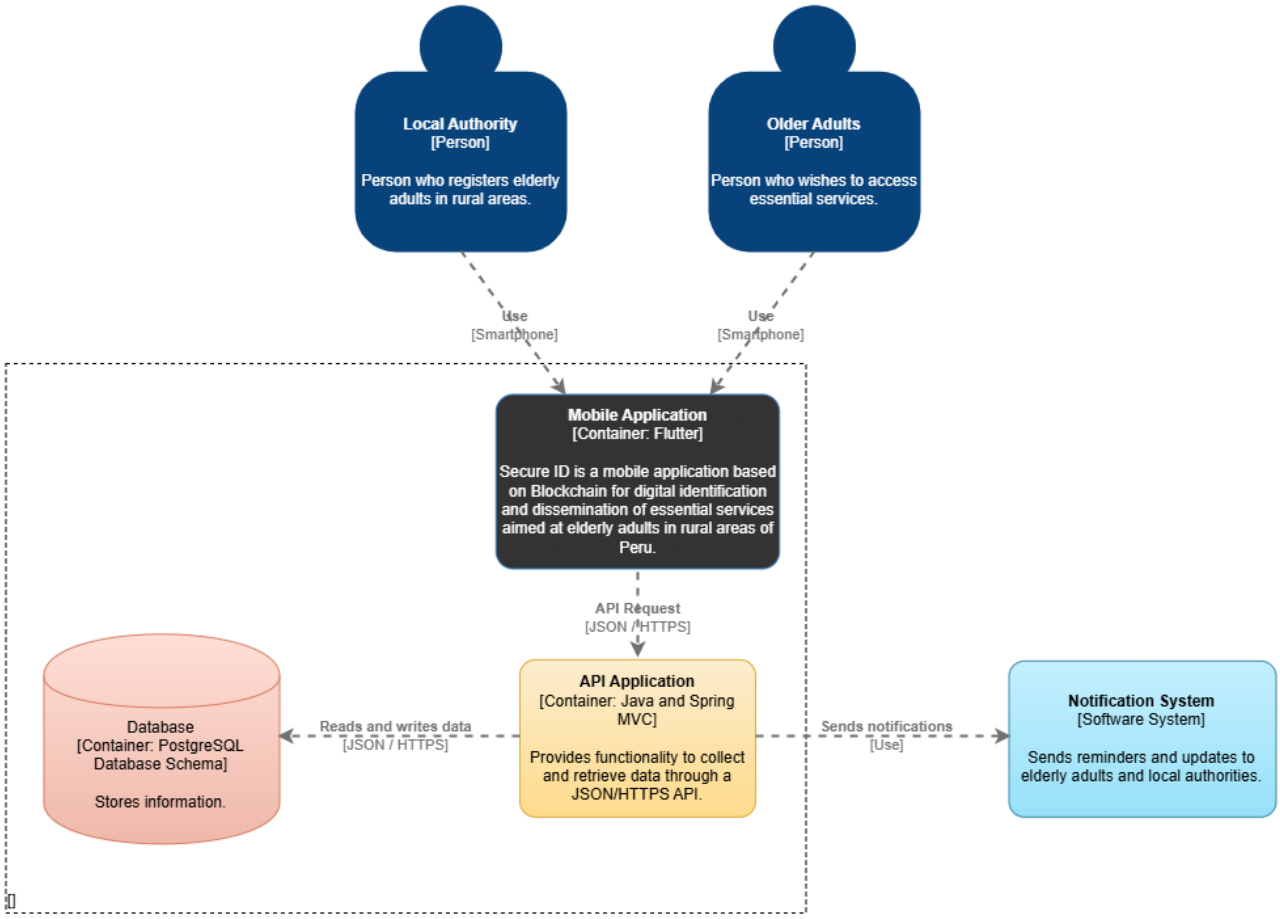


Figure 4. C4 model container diagram.



To validate the user experience, navigation flows were developed that represent common user steps from login to service access. Figure 5 shows one of these flows, focusing on digital ID.

Finally, low and medium fidelity mockups were designed in Figma. Figure 6 illustrates the most representative screens of the prototype: user registration, ID display, and service dissemination. These prototypes were essential to validate the suitability of the design to the capabilities and preferences of the target audience.

Figure 5 illustrates the proposed navigation flow for the digital ID process of older adults. This sequential design optimizes the user experience through a clear interface and guided steps from the initial registration to the display of the digital ID, ensuring accessibility, usability, and efficiency in identity management.

Figure 6 presents a series of mockups that illustrate key system functionalities from the administrator's perspective. These include user registration, display of ID information, and management of essential services such as health care, social support, education, and food. The design promotes efficient administration and clear dissemination of campaigns targeting vulnerable populations.

Figure 5. Ideal navigation flow for identification.

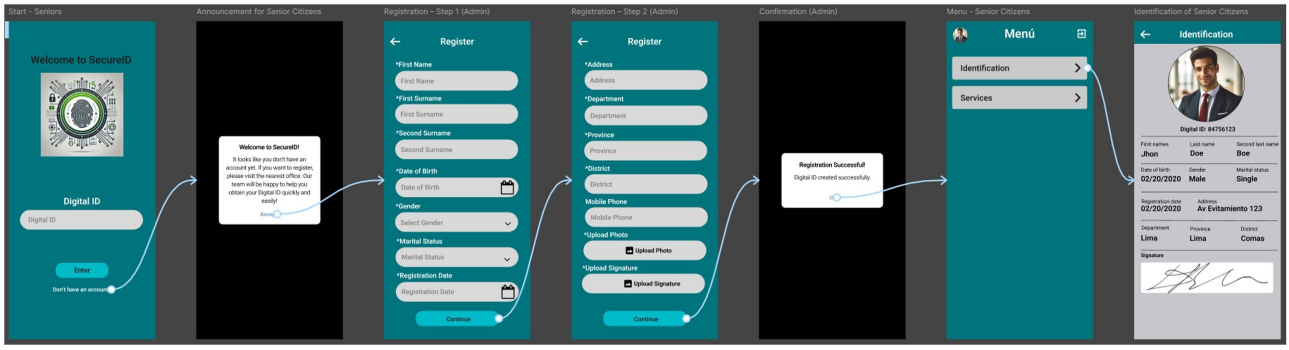
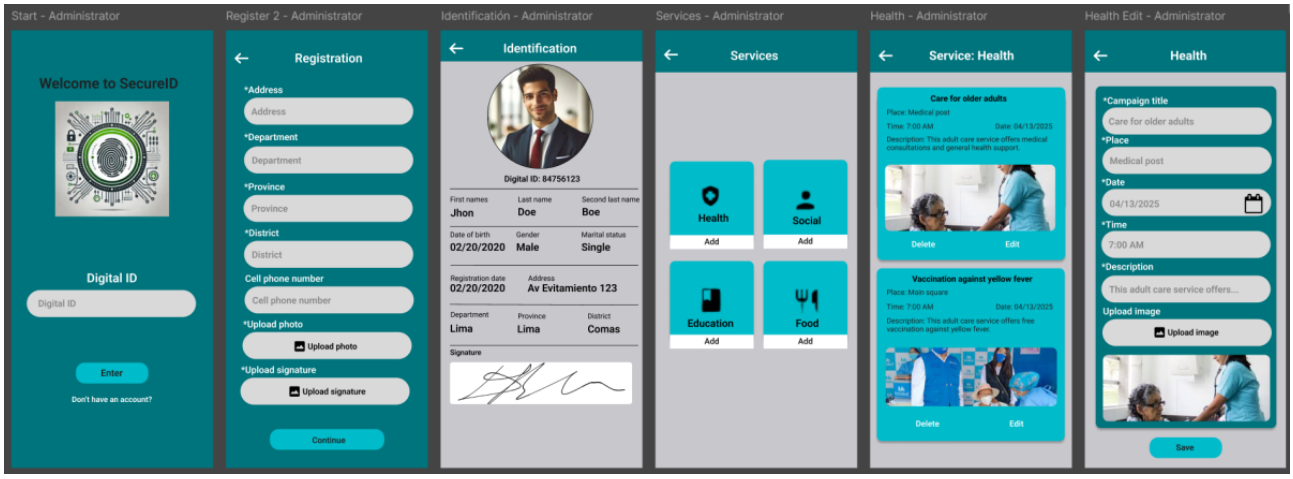


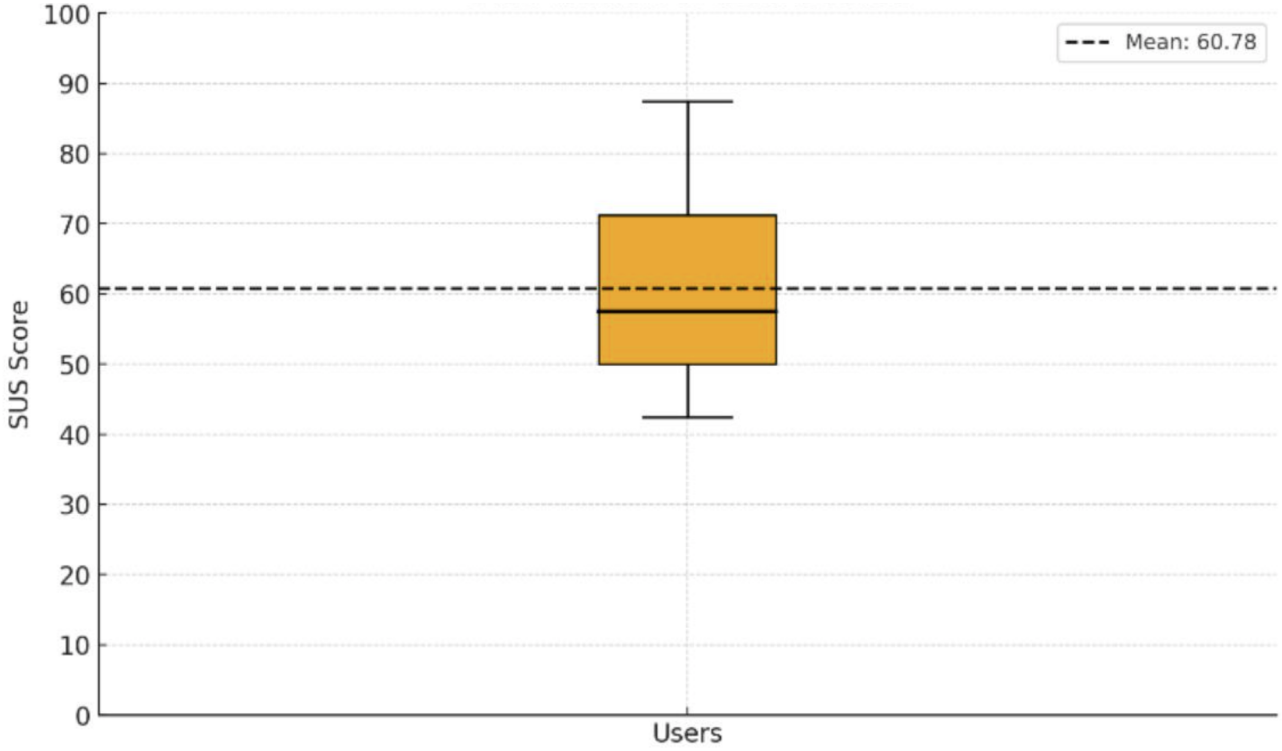
Figure 6. Mockups: registration, identification, and dissemination of essential services.



Testing Phase: Usability Evaluation

In this phase, the results obtained by applying the SUS to a sample of 16 older adults, representative users of the target population who interacted with the mobile app prototype, were analyzed. The mean SUS score was 60.78 (SD 13.68) on a scale from 0 to 100, where higher values indicate better perceived usability, as can be seen in Figure 7. This score is interpreted as an acceptable usability according to the standards established in the literature, which indicates that the designed prototype is perceived as easy to use, safe, and suitable for the target population.

Figure 7. Distribution of System Usability Scale (SUS) scores using a boxplot.



Discussion

Outline

In this study, we designed and evaluated the usability of a blockchain-based mobile app architecture intended to support secure digital ID and access to essential services for older adults in rural Peru. Usability testing with 16 older adults showed that the mean SUS score was 60.78 (SD 13.68), indicating acceptable usability. Older adults viewed the app as secure and mostly easy to use and specifically highlighted the intuitive interface, assisted registration process, and personalized notifications as positive features [25,27], although there were ongoing issues related to digital literacy, trust in data security, and accessibility for users with disabilities.

In terms of accessibility, the design targeted low digital literacy and disability-related barriers through large high-contrast fonts, simplified menus, intuitive iconography, and an assisted registration flow [19]. In our sample, most participants self-reported no visual or motor limitations; those who did generally required more time and occasional assistance during

formative sessions, which directly informed these adaptations [8]. The confirmatory evaluation will quantify effectiveness using prespecified metrics, including completion without assistance across core tasks, median task time, error rate, and SUS, stratified by limitation status.

The prototypes and C4 models developed demonstrate a clear and scalable architecture, since they are structured in modular layers that allow incorporating new functionalities without affecting the stability of the system [22]. This scalability is reflected in the use of the C4 model, which facilitates the expansion of the system both in terms of components and cloud services, adapting to different regions or user groups [28]. The solution presented is accessible and reliable, as evidenced by the results obtained in the usability tests with older adults [29].

Unlike the conceptual framework proposed by Tan et al [30], which focuses on a governance taxonomy for blockchain-based systems in the public sector at macro, meso, and micro levels, our proposal is based on a practical and applied approach, which directly implements such principles in a functional architecture with tested prototypes. This framework includes concrete technical decisions on decentralized authentication, user privacy, and accessibility. The designed app facilitates secure and efficient access to essential services, promoting digital inclusion and trust in the handling of personal information [11]. As a limitation, while the study initially focused on design and prototyping, the app is currently being evaluated in real-world settings with older adults from the target population to validate its effectiveness, usability, and adoption in practical environments.

Related Work

In recent years, academic studies have begun to examine how blockchain technology can reinforce governance structures, support the decentralization of public service delivery, enhance security for mobile apps, and improve the user experience in distributed digital environments. Below are four thematic categories that frame the most relevant work.

Application of Blockchain in Public Sector Governance Models

Tan et al [30] address the implementation of blockchain's potential to transform public services by improving transparency, efficiency, and security, key aspects for the design of blockchain-based mobile apps. Through a conceptual framework, the authors explore key governance decisions at three levels: micro, meso, and macro. These governance elements, illustrated in Figure 8, affect the design and implementation of blockchain-based systems in the public sector [31]. However, they identify limitations such as the lack of interoperable infrastructure and the need for effective governance models. This study contributes to these limitations by proposing a solution based on a more flexible approach to digital identity management and access to essential services, especially through mobile apps that enable better interaction with public services [30].

Figure 8. Key elements in the economic design of governance.

Specifically, the proposed architecture directly operationalizes Tan et al's [30] microlevel governance principle through user-centered control of digital credentials. For instance, the assisted registration system and blockchain-based authentication module enable users to manage their identity data without relying on centralized authorities. Likewise, the notification and verification features provide transparency and consent management, ensuring that users are aware of and approve any data exchanges. At the mesolevel, interoperability is supported by modular components that facilitate integration with local authority systems, while the macrolevel implications relate to potential scalability within national digital identity strategies. In this way, our design translates Tan et al's [30] theoretical governance taxonomy into practical app features that promote individual autonomy and trust in digital identity management.

For his part, Ibrahim [32] explores the impact of decentralization on improving public services, highlighting the need to optimize government efficiency and accountability by delegating authority to local governments. This approach, which highlights the importance of tailoring services to local needs, is essential when considering blockchain-based mobile apps for public services. However, Ibrahim notes that disparities in local resources and capabilities limit the effectiveness of decentralization. Our research complements this analysis by integrating blockchain-based technologies, providing a solution that improves security, accessibility, and efficiency in digital ID and access to essential services, overcoming local resource barriers through a decentralized and accessible infrastructure [32].

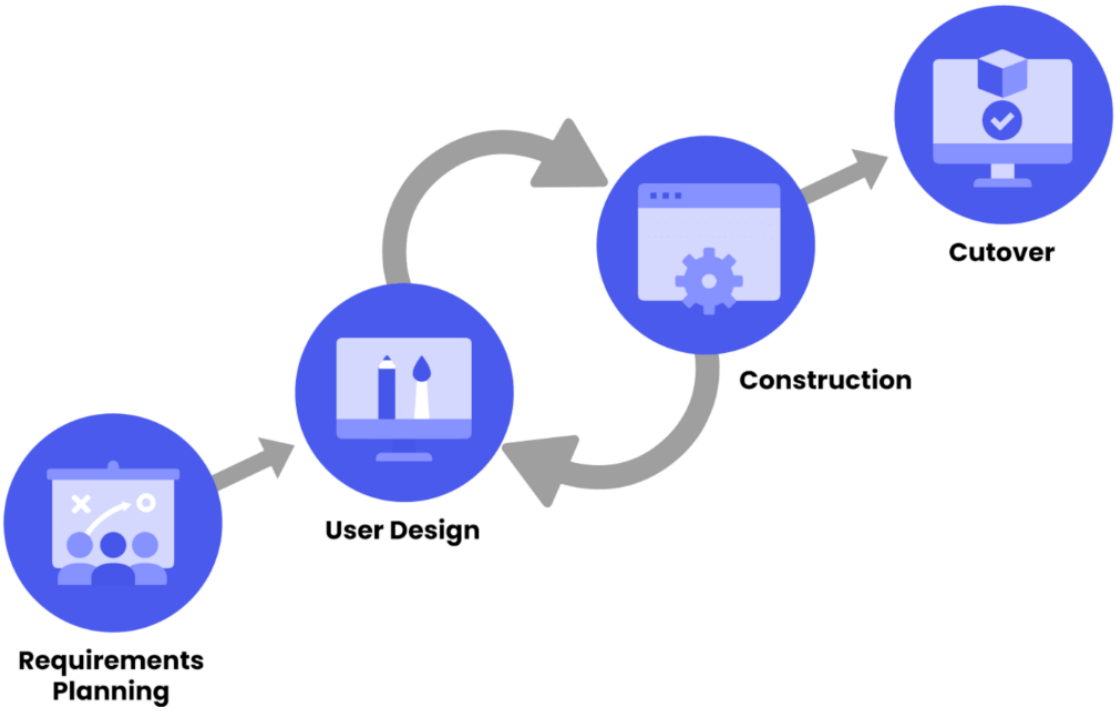
Innovations in Blockchain-Based Mobile Apps for Security and Education

Rizky et al [33] propose a blockchain-based mobile app for decentralization of information management in the field of e-journals, addressing security and reliability issues in

centralized systems. Although their focus is on data security in academic platforms, their findings on decentralization and security are highly relevant to the design of mobile apps that manage digital ID in essential public services. Using a SWOT (strengths, weaknesses, opportunities, and threats) analysis and a waterfall development approach, the study implements blockchain to ensure a network resilient to external interference. However, scalability in larger environments remains a challenge. Our research advances this by proposing technical solutions that improve the efficiency and accessibility of essential services, overcoming these limitations by integrating blockchain for secure and decentralized digital ID.

Similarly, Asmawi et al [34] developed BlockScholar, a blockchain-based mobile educational app to facilitate the understanding of blockchain. Their research addresses the gap in interactive and accessible educational resources, highlighting the need for platforms that offer immersive and accessible learning on blockchain. They used the rapid application development model to create interactive and gamified content, as illustrated in Figure 9. Although the study made progress in accessibility and comprehension, it still faces challenges in adaptability to diverse educational contexts. Therefore, our research proposes solutions that enhance learning personalization and expand the coverage of blockchain-related topics.

Figure 9. Rapid application development model phases.



Security, Identity, and Digital Inclusion with Blockchain

Gumilar et al [35] explore the integration of financial technology in digital inclusion, with an emphasis on digital financial literacy as a catalyst for reducing economic and social disparities [35]. Through a systematic literature review, they used Scopus databases (2020-2024) to examine how the adoption of digital financial services optimizes inclusion. The results highlighted the importance of improving digital financial literacy but noted limitations in consumer use and protection. In line with these findings, our research extends the analysis by proposing innovative technological solutions and a more accessible design to improve adoption in underserved populations.

On the other hand, Musa et al [36] propose a blockchain-based approach to improve the security of data storage in Android mobile apps. They address the problem of vulnerability and unauthorized access to sensitive data by using blockchain to provide decentralized and secure storage. Their methodology includes the implementation of a 6-layer framework called BSADS (blockchain-based secure android data storage), which optimizes efficiency and security, as detailed in their proposed framework. Despite advances, scalability and costs remain limitations. This research contributes to overcoming these challenges by proposing lightweight node solutions and cost optimization techniques.

Identity and User Experience Management on Blockchain

Alanzi and Alkhatib's [37] study presents blockchain-based identity management solutions aimed at improving privacy and security in traditional centralized systems. Using blockchain technologies such as Ethereum and smart contracts, they address issues of third-party control, single point of failure, and vulnerability to data manipulation. However, limitations in scalability and weak authentication of some proposed systems are identified. In this context, our research extends these approaches, proposing improved decentralized infrastructure and optimizing the design of smart contracts to increase security and efficiency in digital identity management systems.

On the other hand, Jang and Han [38] developed a user experience framework for blockchain-based services, specifically focused on improving user interaction in contexts

such as finance and health care [38]. Through an analysis of active services, they identified both general and blockchain-specific functions, highlighting improved efficiency and security. However, their study faces limitations in general applicability due to a lack of standardization. Consequently, our research extends this framework, addressing the implementation of blockchain in digital identity and essential services, improving accessibility and user experience through advanced technological solutions [38]. Table 5 summarizes the key approaches and inputs underlying the development of a blockchain-based mobile app for digital ID and access to essential services.

Table . Fundamental approaches for blockchain-based mobile apps.

Approach	Contribution	Referenced works
Taxonomy of governance	• Rationale for the governance approach	• Tan et al [30]
Decentralization	• Decentralization analysis	• Ibrahim [32]
Patterns of mobile apps	• Mobile app design patterns	• Rizky et al [33] • Asmawi et al [34]
Inclusion, security, identity, and user experience	• Design and accessibility considerations • Blockchain security and authentication • Digital identity management • User experience and accessible design	• Gumilar et al [35] • Musa et al [36] • Alanzi and Alkhatib [37] • Jang and Han [38]

Conclusions

This study designed a functional architectural proposal for a blockchain-based mobile app that facilitates digital ID and access to essential services, with a focus on the inclusion and data security of older adults. The solution addresses usability, privacy, and decentralization issues, overcoming the limitations of traditional centralized and vulnerable systems. In addition to its technical and usability benefits, the proposed system entails ethical challenges related to data governance. If local authorities or institutions act as blockchain nodes, they could potentially exert undue control over users' identity data. To prevent this, future implementations should ensure transparent node management, community oversight, and independent auditing. These measures are key to promoting genuine digital inclusion while safeguarding autonomy and trust in rural contexts. Nevertheless, usability tests applied using the SUS revealed high acceptance and perceived ease of use by older adults, demonstrating that the proposal represents a solid basis for the comprehensive development of the app, which has been implemented in a functional prototype currently undergoing real-world testing with older adult participants in rural areas.

Acknowledgments

The authors would like to thank the Dirección de Investigación de la Universidad Peruana de Ciencias Aplicadas (UPC) for the support provided for the publication of this article.

Funding

This work received financial support from the Dirección de Investigación de la Universidad Peruana de Ciencias Aplicadas (UPC) exclusively for the article processing fee (APF). No additional funding was provided for the research activities.

Data Availability

The data collected in this study consist of usability evaluation responses (eg, survey data such as System Usability Scale scores and qualitative feedback) obtained from voluntary participants during prototype testing. No clinical, laboratory, or medical record data were collected. The dataset is not publicly available due to privacy and ethical considerations, but can be provided by the corresponding author upon reasonable request.

Authors' Contributions

WA-R and AFP-L contributed equally to this work. They led the study execution, including participant recruitment, data collection, usability evaluation, analysis of results, and drafting of the manuscript. JCM-A provided research supervision and leadership, contributed to the study planning and methodological guidance, ensured alignment with journal requirements, coordinated the submission and revision process, and performed critical review and editing of the manuscript.
All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
C4 model component diagram.
[PNG File, 134 KB - rehab_v13i1e79553_app1.png]

References

1. Cai S, Du J, Chen X, Li E, Chen Y. The relationship between e-health literacy and educational participation motivation among elderly individuals: the chained mediating effects of self-identity and social capital. Br J Hosp Med (Lond) 2024 Sep 30;85(9):1-13. [doi: 10.12968/hmed.2024.0261] [Medline: 39347685]
2. Farao A, Paparis G, Panda S, Panaousis E, Zarras A, Xenakis C. INCHAIN: a cyber insurance architecture with smart contracts and self-sovereign identity on top of blockchain. Int J Inf Secur 2024 Feb;23(1):347-371. [doi: 10.1007/s10207-023-00741-8]
3. Morte-Nadal T, Esteban-Navarro M. Recommendations for digital inclusion in the use of European digital public services. Humanit Soc Sci Commun 2025;12(1):273. [doi: 10.1057/s41599-025-04576-7]
4. Yi S, Yam ELY, Cheruvettolil K, et al. Perspectives of digital health innovations in low- and middle-income health care systems from South and Southeast Asia. J Med Internet Res 2024 Nov 25;26:e57612. [doi: 10.2196/57612] [Medline: 39586089]
5. Song W, Nokhbeh Zaeem R, Liau D, et al. Self-sovereign identity and user control for privacy-preserving contact tracing. Presented at: WI-IAT '21; Dec 14-17, 2021; Melbourne, Australia p. 438-445 URL: https://dl.acm.org/doi/proceedings/10.1145/3486622 [doi: 10.1145/3486622.3493914]
6. George M, Chacko AM. Health passport: a blockchain-based PHR-integrated self-sovereign identity system. Front Blockchain 2023;6:1075083. [doi: 10.3389/fbloc.2023.1075083]
7. Ji H, Dong J, Pan W, Yu Y. Associations between digital literacy, health literacy, and digital health behaviors among rural residents: evidence from Zhejiang, China. Int J Equity Health 2024 Apr 9;23(1):68. [doi: 10.1186/s12939-024-02150-2] [Medline: 38594723]
8. Lu SY, Yoon S, Yee WQ, Heng Wen Ngiam N, Ng KYY, Low LL. Experiences of a community-based digital intervention among older people living in a low-income neighborhood: qualitative study. JMIR Aging 2024 Apr 25;7:e52292. [doi: 10.2196/52292] [Medline: 38662423]
9. Prieto-Egido I, Sanchez-Chaparro T, Urquijo-Reguera J. Impacts of information and communication technologies on the SDGs: the case of Mayu Telecomunicaciones in rural areas of Peru. Inf Technol Dev 2023 Jan 2;29(1):103-127. [doi: 10.1080/02681102.2022.2073581]
10. Li J, Liang N, Wu Y, Ma X, Liao Z. Effect of internet usage on the life satisfaction of the elderly spousal caregivers of disabled elderly. Sci Rep 2024;14(1):22294. [doi: 10.1038/s41598-024-73298-8]
11. Konopik J, Blunck D. Development of an evidence-based conceptual model of the health care sector under digital transformation: integrative review. J Med Internet Res 2023;25:e41512. [doi: 10.2196/41512]
12. Husain L, Greenhalgh T. Examining intersectionality and barriers to the uptake of video consultations among older adults from disadvantaged backgrounds with limited English proficiency: qualitative narrative interview study. J Med Internet Res 2025 Jan 6;27:e65690. [doi: 10.2196/65690] [Medline: 39761566]
13. Zuñe Chero L, Capuñay-Uceda OE, Idrogo Burga E, Capuñay-Uceda CE. Inequalities in learning performance caused by the digital divide during the health emergency. Presented at: 22nd LACCEI International Multi-Conference for Engineering, Education and Technology (LACCEI 2024; Jul 17-19, 2024 URL: https://laccei.org/LACCEI2024-CostaRica [doi: 10.18687/LACCEI2024.1.1.1042]
14. UPC Code of Ethics in Scientific Research of the Peruvian University of Applied Sciences (UPC). URL: https://investigacion.upc.edu.pe/assets/files/C%C3%B3digo-de-%C3%A9tica-en-la-investigacion-cientifica-de-la-UPC.pdf [accessed 2026-02-02]
15. Nimmanterdwong Z, Boonviriya S, Tangkijvanich P. Human-centered design of mobile health apps for older adults: systematic review and narrative synthesis. JMIR Mhealth Uhealth 2022 Jan 14;10(1):e29512. [doi: 10.2196/29512] [Medline: 35029535]
16. Sobrinho ADS, Gomes GDO, Bueno Júnior CR. Developing a multiprofessional mobile app to enhance health habits in older adults: user-centered approach. JMIR Form Res 2024 Apr 15;8(8):e54214. [doi: 10.2196/54214] [Medline: 38619865]
17. Essop H, Kekana R, Smuts H. Co-designing of a prototype mobile application for fetal radiation dose monitoring among pregnant radiographers using a design thinking approach. Health Informatics J 2024;30(3):14604582241284960. [doi: 10.1177/14604582241284960] [Medline: 39348214]
18. Chen E, Bishop J, Guge Cozon L, et al. Integrating human-centered design methods Into a health promotion project: supplemental nutrition assistance program education case study for intervention design. JMIR Form Res 2023 Apr 21;7:e37515. [doi: 10.2196/37515] [Medline: 37083485]
19. Gomez-Hernandez M, Ferre X, Moral C, Villalba-Mora E. Design guidelines of mobile apps for older adults: systematic review and thematic analysis. JMIR Mhealth Uhealth 2023 Sep 21;11:e43186. [doi: 10.2196/43186] [Medline: 37733401]
20. Rashidi H, Rashidi Z. Software architecture tools: a classification and survey with recommendation for an organization. J Comput Secur 2023(10):61-81. [doi: 10.22108/jcs.2023.137862.1131]

21. Voorheis P, Zhao A, Kuluski K, et al. Integrating behavioral science and design thinking to develop mobile health interventions: systematic scoping review. JMIR Mhealth Uhealth 2022 Mar 16;10(3):e35799. [doi: 10.2196/35799] [Medline: 35293871]
22. Liang X, Zhao J, Chen Y, Bandara E, Shetty S. Architectural design of a blockchain-enabled, federated learning platform for algorithmic fairness in predictive health care: design science study. J Med Internet Res 2023 Oct 30;25:e46547. [doi: 10.2196/46547] [Medline: 37902833]
23. Brooke J. SUS: a “quick and dirty” usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry: Taylor & Francis; 1996:189-194. [doi: 10.1201/9781498710411]
24. Hill JR, Brown JC, Campbell NL, Holden RJ. Usability-in-place-remote usability testing methods for homebound older adults: rapid literature review. JMIR Form Res 2021 Nov 2;5(11):e26181. [doi: 10.2196/26181] [Medline: 34726604]
25. Sweeney M, Barton W, Nebeker C. Evaluating mobile apps targeting older adults: descriptive study. JMIR Form Res 2023 Apr 27;7:e37329. [doi: 10.2196/37329] [Medline: 37103995]
26. Schroeder T, Dodds L, Georgiou A, Gewald H, Siette J. Older adults and new technology: mapping review of the factors associated with older adults’ intention to adopt digital technologies. JMIR Aging 2023 May 16;6:e44564. [doi: 10.2196/44564] [Medline: 37191976]
27. Kim S, Park C, Park S, et al. Measuring digital health literacy in older adults: development and validation study. J Med Internet Res 2025 Feb 5;27:e65492. [doi: 10.2196/65492] [Medline: 39908081]
28. Liang X, Alam N, Sultana T, Bandara E, Shetty S. Designing a blockchain-empowered telehealth artifact for decentralized identity management and trustworthy communication: interdisciplinary approach. J Med Internet Res 2024 Sep 25;26:e46556. [doi: 10.2196/46556] [Medline: 39320943]
29. Koo JH, Park YH, Kang DR. Factors predicting older people’s acceptance of a personalized health care service app and the effect of chronic disease: cross-sectional questionnaire study. JMIR Aging 2023 Jun 21;6:e41429. [doi: 10.2196/41429] [Medline: 37342076]
30. Tan E, Mahula S, Crompvoets J. Blockchain governance in the public sector: a conceptual framework for public management. Gov Inf Q 2022 Jan;39(1):101625. [doi: 10.1016/j.giq.2021.101625]
31. Barrera C. A framework for blockchain governance design: the Prysm group wheel.: Medium; 2019. URL: https://medium.com/prysmeconomics/a-framework-for-blockchain-governance-design-the-prysm-group-wheel-703279c1b0dd [accessed 2025-06-01]
32. Ibrahim AHH. Decentralization and its impact on improving public services. Int J Soc Sci 2024;7(2):45-53. [doi: 10.21744/ijss.v7n2.2278]
33. Rizky A, Lutfiani N, Sri Mariyati W, Atika Sari A, Febrianto KR. Decentralization of information using blockchain technology on mobile apps e-journal. B-FronT 2022;1(2):1-10 [FREE Full text] [doi: 10.34306/bfront.v1i2.37]
34. Asmawi A, Saifulbahri EM, Mohd Ariffin NA. Development of BlockScholar as an educational mobile application on blockchain technology. ARASET 2023;34(1):15-23 [FREE Full text] [doi: 10.37934/araset.34.1.1523]
35. Gumilar DWA, Sangka KB, Totalia SA. Digital financial literacy and digital financial inclusion in the era of digital disruption: systematic literature review. FJMR 2024;3(5):1563-1576 [FREE Full text] [doi: 10.55927/fjmr.v3i5.9213]
36. Musa HS, Krichen M, Altun AA, Ammi M. Survey on blockchain-based data storage security for android mobile applications. Sensors (Basel) 2023 Oct 26;23(21):8749. [doi: 10.3390/s23218749] [Medline: 37960449]
37. Alanzi H, Alkhatib M. Towards improving privacy and security of identity management systems using blockchain technology: a systematic review. Appl Sci (Basel) 2022;12(23):12415. [doi: 10.3390/app122312415]
38. Jang H, Han SH. User experience framework for understanding user experience in blockchain services. Int J Hum Comput Stud 2022 Feb;158:102733. [doi: 10.1016/j.ijhcs.2021.102733]

Abbreviations

DT: design thinking
SSI: self-sovereign identity
SUS: System Usability Scale

Edited by S Munce; submitted 24.Jun.2025; peer-reviewed by C Wrigley, SB Guo; revised version received 20.Nov.2025; accepted 03.Dec.2025; published 02.Feb.2026.

Please cite as:
Arana-Ramos W, Pastrana-Leon AF, Morales-Arevalo JC
Blockchain-Based Mobile App for Digital Identification of Older Adults in Rural Peru: Design and Usability Evaluation Study
JMIR Rehabil Assist Technol 2026;13:e79553
URL: https://rehab.jmir.org/2026/1/e79553
doi:10.2196/79553

Therapeutic Use of Virtual Reality for Patients With Fibromyalgia and Chronic Neck Pain: Randomized Controlled Trial

Edurne Úbeda-D'Ocasar[1,2*], PhD, Prof Dr; Yaiza Moreno-Crespo[1,2*], PT; Eduardo Cimadevilla-Fernández-Pola[1,2*], PhD, Prof Dr; Juan Hernández-Lougedo[1,2*], PhD, Prof Dr; Álvaro Navas-Mosqueda[1,2*], PT; Mario Caballero-Corella[1,2*], PT; Noemí Mayoral-Gonzalo[1,2*], PT; Blanca Pedauyé-Rueda[1,2*], RDN; María Jesús Fernández-Aceñero[3*], Dr med, PhD; Juan Pablo Hervás-Pérez[4*], PhD, Prof Dr; Cristina Ojedo-Martín[1,2*], PT

[1]Physiotherapy and Health Research Group (FYSA), Faculty of Health Sciences – HM Hospitals, Camilo José Cela University, Villanueva de la Cañada, Madrid, Spain

[2]Instituto de Investigación Sanitaria HM Hospitales, Madrid, Spain

[3]Department of Legal Medicine, Psychiatry and Pathology, Facultad de Medicina, Universidad Complutense de Madrid, Fundación de Investigación Clínico San Carlos (IDiSCC), Madrid, Spain

[4]Department of Chemistry in Pharmaceutical Sciences, Analytical Chemistry Unit, Faculty of Pharmacy, Complutense University of Madrid, Plaza Ramón y Cajal s/n, Ciudad Universitaria, Madrid, Spain

[*]all authors contributed equally

Corresponding Author:
Juan Pablo Hervás-Pérez, PhD, Prof Dr
Department of Chemistry in Pharmaceutical Sciences, Analytical Chemistry Unit, Faculty of Pharmacy, Complutense University of Madrid, Plaza Ramón y Cajal s/n, Ciudad Universitaria, Madrid, Spain

Abstract

Background: Fibromyalgia (FM) causes widespread pain, fatigue, and cognitive abnormalities. Cervical pain is a common and debilitating symptom.

Objective: This study aims to evaluate the effectiveness of virtual reality (VR) as a treatment for chronic cervical pain experienced by patients with FM.

Methods: A single-blind randomized clinical trial was conducted. A total of 56 women were randomly assigned to 3 groups: G1 (VR+cervical mobility exercises), G2 (cervical mobility exercises), and the control group. Therapy was administered twice a week for 4 weeks. Variables such as disease impact, quality of life, kinesiophobia, pain, range of motion, fatigue, and treatment adherence were measured.

Results: The mean age of the participants was 54.26 (SD 7.7) years. Participants were overweight, with a mean BMI of 28.7 (SD 7.8). The mean visual analog scale value was 6.72 (SD 1.8). The baseline values for age, BMI, visual analog scale, algometric measures, and functional capacity (measured using the Timed Up and Go test, cervical rotation, and lateral displacement) were similar across the 3 groups. Following the intervention therapy, the control group did not exhibit notable improvement (mean 3.5, SD 1.4; differences of mean values –0.46, 95% CI –1.1 to 0.2; *P*=.15), particularly in pain perception, while both therapy groups did show improvements (G1: mean 3.8, SD 1.1; differences of mean values 1.2, 95% CI 0.78-1.54; *P*<.001; G2: mean 2.8, SD 1.8; differences of mean values 1.2, 95% CI 0.66-1.7; *P*<.001). Both intervention groups improved significantly compared to control postintervention in FM impact (CG vs G1: differences of mean values 9.31, 95% CI 14.7-3.8; *P*<.001; CG vs G2: differences of mean values 8.4, 95% CI 13.84-3.06; *P*<.001), central sensitization (CG vs G2: differences of mean values 7.53, 95% CI 12.12-2.95; *P*<.001), and cervical disability (CG vs G2: differences of mean values 6.44, 95% CI 9.93-2.94; *P*<.001). However, at 1 month, only G1 maintained superior improvements across all measures, including a reduction in kinesiophobia (G2: differences of mean values 6.2, 95% CI 4.7-9.8; *P*<.001), indicating a more sustained effect of the combined approach.

Conclusions: The combination of VR with cervical mobility and strengthening exercises produced superior and sustained improvements in women with FM compared to exercise alone or control. Significant benefits were observed in disease impact, central sensitization, cervical disability, and kinesiophobia, with effects maintained at 1 month only in the VR group. These findings support VR as an effective adjunct to enhance symptom management and treatment adherence in FM.

Trial Registration: ClinicalTrials.gov NCT05933941; https://clinicaltrials.gov/study/NCT05933941

(JMIR Rehabil Assist Technol 2026;13:e81158) doi:10.2196/81158

KEYWORDS

chronic neck pain; exercise; fibromyalgia; virtual reality; randomized controlled trial

Introduction

Fibromyalgia (FM) is a disease that causes widespread chronic pain and intense fatigue along with many other symptoms [1-3]. Historically, there was some skepticism among health care professionals regarding the existence of this disease and its classification as a mental disorder. Such skepticism may have arisen from concerns about the optimal use of resources in providing care for individuals with FM [3,4]. It is estimated that this disease affects between 2% and 4% of the Spanish population, making it the most common cause of diffuse chronic musculoskeletal pain. Notably, the prevalence is much higher in women, with a ratio of 8:1 [5-7]. Its incidence is increasing because of the improvements in diagnostic criteria and advances in research that are bringing this pathology to light [7,8]. Given its high prevalence, this disease represents a significant public health concern [9,10].

Currently, there are no specific diagnosis tests available for confirming an FM diagnosis. Historically, an FM diagnosis was based on the presence of widespread pain in at least 11 of 18 designated tender points for a duration of 3 months. However, this criterion proved insufficient to capture the complexity of the disorder. Consequently, more comprehensive diagnostic frameworks have been developed, including a range of associated symptoms such as fatigue, joint stiffness, cervical and lumbar pain, muscle weakness, depressive symptoms, gastrointestinal issues, cognitive impairments, and balance alterations. Collectively, these manifestations significantly compromise the quality of life of individuals affected by FM. This pain can become disabling during flare-ups and even become chronic, often being a cause of recurrent absences from work [11].

Unlike pain associated with a sedentary lifestyle or poor posture, neck pain in FM results from a complex interplay between physiological, neurological, and psychological factors. Chronic pain often involves the cervical spine, leading to acute and prolonged pain episodes, along with muscle tension in the suboccipital, trapezius, and elevator scapulae muscles. It is thought that several factors may contribute to this chronic pain, including central pain sensitization, nervous system dysregulation, sleep disturbances, musculoskeletal changes, and psychological and emotional factors [11-13].

In recent years, new therapeutic approaches have been explored to improve symptomatology and thus patients' quality of life. One of the main challenges in treating patients with chronic pain, particularly those with FM, is their low adherence to physical therapy exercises. This may be due to factors such as severe pain and fatigue, a fear of worsening symptoms, a lack of understanding about exercise benefits, inadequate support and supervision, and frustration with slow improvement. To address these issues, it would be beneficial to adopt a comprehensive approach that involves educating patients regarding the benefits of exercise, customizing programs according to individual needs, providing ongoing support via health care professionals, managing pain and fatigue during exercise, setting realistic expectations, and recognizing achievements [14-16]. It may be beneficial to consider a treatment plan that incorporates aerobic activities, strength training, and stretching exercises [17-19]. However, this type of therapy may be limited by kinesiophobia. One possible solution to this issue could be the use of virtual reality (VR) technology. Exercise interventions delivered through VR have shown superior efficacy in alleviating the core symptoms of FM, including pain, fatigue, and stiffness, while also promoting improvements in balance and postural control [20]. The affordability of portable VR devices, combined with their sustained effectiveness as a nonpharmacological intervention for chronic pain management, positions VR as a promising tool with potential future applications as an analgesic modality [21]. Numerous randomized controlled trials have demonstrated that VR constitutes an alternative and accessible therapeutic approach for pain management [22].

Immersive VR is an innovative and interactive technology that generates 3D scenarios, enabling patients to experience highly realistic simulations while interacting with the virtual environment using their own hands. VR has been researched and applied in various contexts for pain management, demonstrating considerable potential as a therapeutic tool. The effectiveness of this therapeutic approach has been demonstrated for patients with chronic pain, as it addresses several factors, such as fatigue, kinesiophobia, and range of motion (ROM) [23]. This therapy can be adapted to suit the specific requirements of each patient, providing the option to adjust the level of difficulty from low to intermediate or high. Additionally, it offers a variety of game modes that could be beneficial in targeting different clinical conditions, such as low back pain, neck pain, and balance disorders [24,25].

The therapeutic mechanisms and components of VR are distraction, activity management and behavioral activation, skills-based cognitive behavioral therapy, relaxation training and biofeedback, positive emotion induction, neuromodulation, physical rehabilitation, and reduction of kinesiophobia [8,23,26].

VR can be categorized into 2 main types: immersive and nonimmersive. Immersive VR involves the use of head-mounted displays or VR goggles that fully cover the visual field and may include motion sensors for the hands or feet. In contrast, nonimmersive VR is accessed through conventional screens or computers, without any device that isolates the user from the external environment [27] .

Chronic pain has been extensively investigated over the past decades, and research efforts persist in identifying novel strategies to mitigate its impact and enhance patients' quality of life. VR in both modalities has demonstrated efficacy in reducing chronic musculoskeletal pain. These findings support the integration of VR as a therapeutic tool in clinical populations affected by conditions associated with this type of pain [28].

Patients may perceive amplified motion in the virtual world, depending on the configuration of the headset used. This could result in small neck flexion in real life being rewarded with greater apparent motion in VR, which might increase kinesiophobia [26]. Studies have indicated that for patients with chronic low-back pain, kinesiophobia may be addressed through VR interventions [20,29,30]. VR helps to distract an individual from painful stimuli while exercising in a way that improves

their perception of pain [31]. In addition, there have been indications that improvements in fatigue, sleep quality, and ROM may also be achieved [9].

In this study, we aimed to evaluate the effectiveness of VR as a therapeutic option for neck pain in patients with FM. We hypothesized that VR could decrease cervical pain and kinesiophobia and increase ROM, facilitating improvements in quality of life and adherence to treatment.

Methods

Trial Design

A single-blind, randomized experimental clinical trial was conducted among women diagnosed with FM recruited through the Association of Fibromyalgia, "AFINSFACRO," in Móstoles, Spain. Following the selection of patients who met the inclusion criteria and the collection of informed consent forms and information sheets, according to ethical standards, the patients were randomly assigned to 1 of the 3 groups. The first group performed 20 minutes of exercise plus 10 minutes of VR (G1), the second group performed 30 minutes of exercise (G2), and the control group (CG) did not perform any exercise or VR treatment.

This study was registered at ClinicalTrials.gov (NCT05933941) and approved by the Clinical Research Committee of the Hospital Clínico San Carlos (Madrid, Spain; ID: VRTCNPPFM-13/07/2023). The research was conducted from April 1, 2024, to January 30, 2025. All participants gave informed consent, and the general scope of this study was explained to them via a participant information sheet. This clinical trial was designed and reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials) guidelines, with specific adherence to the CONSORT-Harms (Consolidated Standards of Reporting Trials Harms Extension) 2022 statement (Checklist 1) to ensure transparent and comprehensive reporting of adverse events and safety outcomes.

Participants

The QuestionPro survey software (QuestionPro, Inc) was used to calculate the required sample size, considering a 95% CI and a 5% α error. The sample size calculation was based on the study by Gulsen et al [8]. A total of 56 women were enrolled, of whom 2 withdrew for personal reasons unrelated to the study. The inclusion criteria comprised female patients aged 20 to 65 years with a confirmed diagnosis of FM and chronic cervical pain lasting more than 3 months. The exclusion criteria included vertigo, claustrophobia, epilepsy, pregnancy, or refusal to provide informed consent.

Type of Sampling

A consecutive, nonprobabilistic convenience sampling strategy was applied to participants who met the inclusion criteria.

Variables and Outcomes

Sociodemographic variables included age, height, weight, BMI, marital status, pain medication use, employment status, smoking history, and comorbid conditions such as restless legs syndrome, chronic fatigue syndrome, temporomandibular joint dysfunction, migraines, irritable bowel syndrome, multiple chemical sensitivity, anxiety, and depression. All these sociodemographic variables were collected using a self-administered survey.

The following tools were used in this study:

1. Pain: we used a visual analog scale (VAS) to assess self-reported pain intensity, ensuring maximum reproducibility among different observers [32]. Furthermore, we used an analog pressure FPK 20 algometer (Wagner algometer, Force Dial FDK 20; Wagner Instruments) to evaluate pain among individuals diagnosed with FM. This instrument allows the assessment of a patient's central sensitivity to pain. Measurements focused on tender points in the occipital and upper trapezius regions bilaterally [33-35].
2. Subjective intensity of effort was assessed using the Borg Category-Ratio Scale (CR-10), which quantifies the subjective intensity of effort experienced during physical exercise or functional testing. Participants were asked to rate their exertion immediately after each test or exercise session on a 0 to 10 scale, where 0 indicates "no exertion at all" and 10 represents "maximal exertion" [36].
3. Neck Disability Index (NDI): the validated Spanish version of this questionnaire was used to assess pain and neck-related disability [37].
4. Fear of movement: the Tampa Scale for Kinesiophobia (TSK) was used to assess the fear of movement [38].
5. Exercise adherence: we used the Exercise Adherence Rating Scale, which is a validated questionnaire with 2 sections, one assessing exercise performance and another evaluating frequency, motivation, and consistency [39].
6. Impact of FM: the Fibromyalgia Impact Questionnaire (FIQ) was used to assess the impact of FM on health-related quality of life [40-42].
7. Quality of life: we used EQ-5D (EuroQol 5-Dimensions) questionnaire, which is a tool that allows the evaluation of a patient's overall quality of life in primary care settings [43-45].
8. Symptoms of central sensitization: we used the validated Spanish version of the Central Sensitization Inventory (CSI). This questionnaire assesses symptoms using a scale of 0 ("never") to 4 ("always"). The total score ranges from 0 to 100. A score above 40 indicates the presence of central sensitization [46].
9. ROM: cervical flexion, extension, lateral flexion, and rotation were assessed using a goniometer [47]. This instrument is used to evaluate one's degree of joint mobility, thereby facilitating the determination of an individual's restrictions.
10. Time Up Go test: the test consisted of measuring the time it took participants to get up from a chair with a height of 46 cm, walking 3 meters, turning around a cone, and sitting down again. This test was performed to assess physical performance, gait, and dynamic balance [48,49].

Procedure

The intervention was conducted over a period of 1 month. G1 and G2 participated in 2 sessions per week. G1 performed 20 minutes of cervical mobility and strengthening exercises

(Multimedia Appendix 1), followed by 10 minutes of immersive VR therapy. G2 completed 30 minutes of cervical mobility and strengthening exercises. The CG did not perform any cervical mobility or strengthening exercises nor participate in immersive VR therapy. The immersive VR therapy was delivered using Meta Quest 2000 headsets, using the game "Interkosmos 2000." In the game, participants assumed the role of a spacecraft pilot navigating through a series of rings while avoiding meteoroids by performing neck movements according to the game's instructions. The difficulty level was adjusted as follows: the first 3 sessions were carried out in "easy" mode, allowing a maximum cervical mobility of 30°. The subsequent 3 sessions were conducted in "medium" mode, allowing up to 60° of cervical mobility. The final 2 sessions were performed in "hard" mode, which not only increased the spacecraft's speed but also enabled a full cervical range of motion. All groups underwent assessments of cervical ROM including flexion, extension, right and left lateral inclination, and right and left rotation. In addition, pain intensity in the upper trapezius (bilaterally) and in the right and left occipital regions was quantified using an algometer. Evaluations were conducted at 4 time points: baseline (preintervention), immediately after the intervention, 15 days postintervention, and 1 month postintervention. All follow-up assessments were performed at the association's premises. After each session, participants from G1 and G2 reported their perceived levels of cervical pain and fatigue. All outcome measures were collected by a research assistant blinded to the study objectives and group allocations.

Evaluations were conducted at 4 time points: baseline (preintervention), immediately postintervention, 15 days postintervention, and 1-month postintervention. All follow-up assessments were conducted at the association's headquarters. After each session, participants in G1 and G2 reported perceived levels of cervical pain and fatigue.

All outcome measures were collected by a research assistant who was blinded to both the study's objectives and the participants' group assignments.

Statistical Analysis

Statistical analyses were performed using SPSS 29.0 (IBM Corp). The normality of quantitative variables was assessed with the Kolmogorov-Smirnov test. Depending on distribution, data were described using means and SDs or medians, IQRs, and ranges. Qualitative variables were expressed as percentages and absolute values. Baseline comparisons between the groups were conducted using chi-squared tests or ANOVA with Bonferroni post hoc analysis. Temporal changes in outcomes were analyzed with paired tests, and intergroup differences were analyzed using ANOVA, Bonferroni post hoc, or Kruskal-Wallis tests, as appropriate. Correlations between quantitative variables were assessed using the Pearson or Spearman tests, and associations between qualitative variables were assessed using chi-squared tests. Statistical significance was set at $P<.05$.

Ethical Considerations

This study received approval from the institutional review board of Hospital Clínico San Carlos (23 - 458 EC) and was conducted in accordance with the Spanish legislation, including Law 41/2002 on patient autonomy and Organic Law 3/2018 on data protection and digital rights. These laws prohibit the processing of sensitive personal data such as racial or ethnic origin, political or religious beliefs, biometric identifiers, health information, and sexual orientation. The study also adhered to the ethical principles outlined in the World Medical Association's Declaration of Helsinki 2014 [50]. All procedures conducted in this study complied with applicable ethical standards. Informed consent was obtained from all participants prior to their inclusion in the study. Participant privacy and confidentiality were strictly protected, and all data were handled and stored in accordance with relevant data protection regulations. No personally identifiable information was collected or disclosed. Participants did not receive any financial or material compensation for their participation.

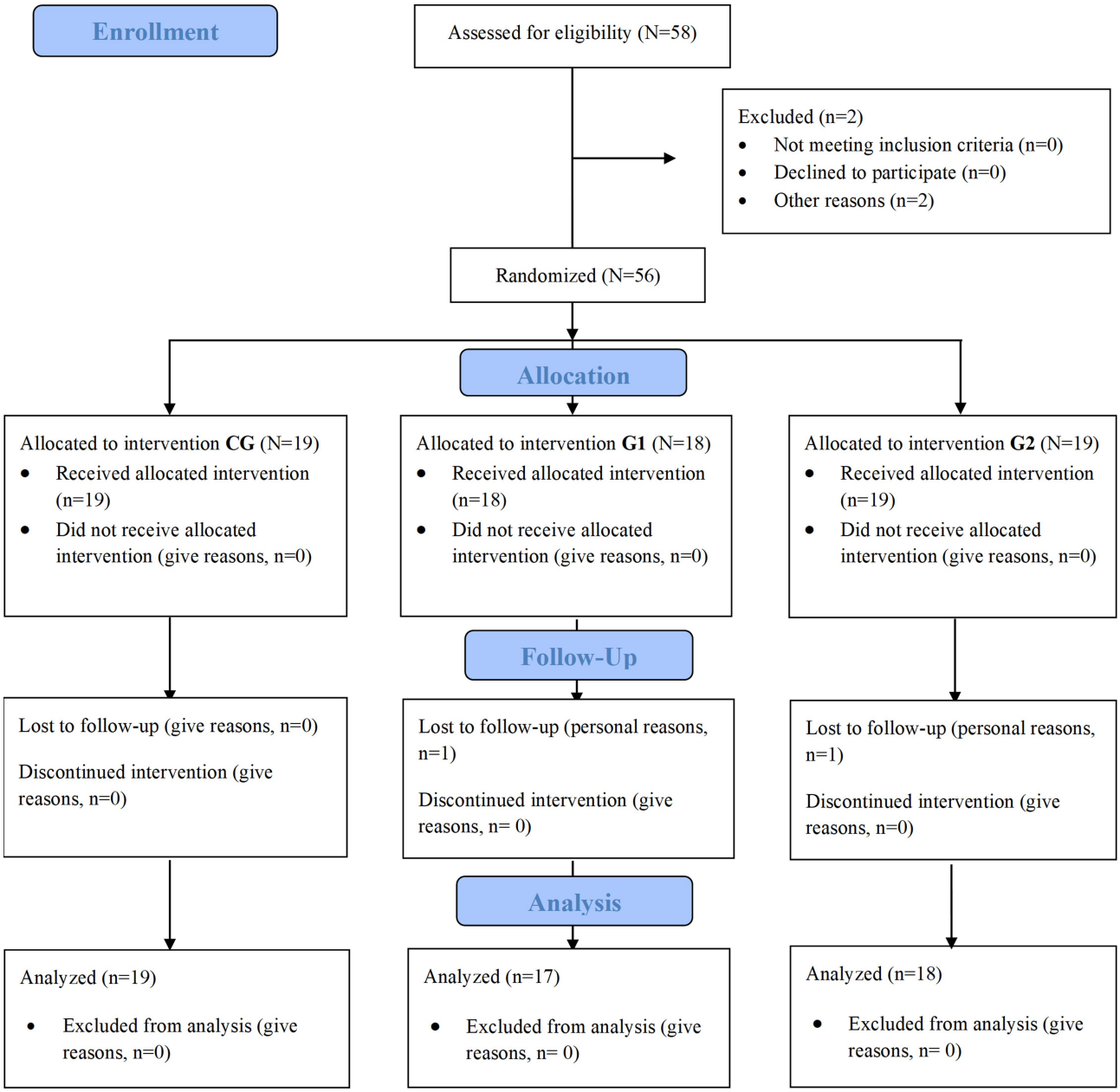
Results

Participant and Baseline Characteristics

The final sample included 54 women with a mean age of 54.26 (SD 7.7) years. Participants were distributed among the groups by computer-generated random number sequence (Figure 1).

Among the participants, 33 (61%) presented chronic fatigue syndrome, 37 (68%) had temporomandibular joint dysfunction, and 16 (33%) reported restless legs syndrome. Regarding the psychological comorbidities, 33 (61%) participants presented anxiety and 13 (24%) depression. Baseline comparisons revealed no statistically significant differences among groups in age, BMI, or other demographic variables. However, subjective measures showed group-level differences in several baseline scales, including FIQ, EQ-5D, CSI, NDI, and Borg scale. Almost all variables at baseline start from similar values, although G1 has a value of 1 point lower on the VAS scale, indicating that they have less pain at the outset. However, in variables such as algometry, the averages reflect similar values compared to the other 2 groups (Multimedia Appendix 2).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. CG: control group.



Postintervention Outcomes

Table 1 shows the differences in the mean values measured before therapy and immediately after its end. We found significant differences in all groups that received intervention in FIQ, CSI, NDI, right and left trapezius, occipital algometer, and in all movements in which ROM was evaluated. For the right and left trapezius and occipital muscles, algometric measurements were obtained, and ROM was assessed across all evaluated movements. About the treatment adherence variable, G1 obtained an average of 94 points out of 100, compared to G2, whose post-treatment average was 74.72. The summary of the comparison between baseline measures and measures immediately after intervention (intragroup comparisons). Shown are the differences in the mean values between both time points (Multimedia Appendix 3).

Table . Intragroup (G1, G2, and CG[a]) analysis of variables measured using questionnaires between baseline and immediately after intervention.

Value and intervention group	Differences of mean values (95% CI)	*P* value
FIQ[b] (1-100)		
Whole series	5.4 (3.2 to 7.5)	<.001
G1	9 (4.6 to 13.3)	<.001
G2	8.1 (5.5 to 10.6)	<.001
CG	0.36 (−3.2 to 2.51)	.80
EQ-5D[c] (–0.59 to 1)		
Whole series	4.1 (2.6 to 5.6)	<.001
G1	3.1 (0.6 to 5.6)	.01
G2	7.7 (5.6 to 9.8)	<.001
CG	−1.7 (−4.4 to 1)	.20
TSK[d] (17-37)		
Whole series	3.2 (0.8 to 5.6)	.01
G1	1.2 (−5.2 to 7.6)	.70
G2	6.2 (4.7 to 9.8)	<.001
CG	−2.04 (−5.9 to 1.86)	.30
CSI[e] (0-100)		
Whole series	4.2 (2.4 to 6)	<.001
G1	6 (3.5 to 8.5)	<.001
G2	7.2 (4.7 to 9.8)	<.001
CG	0.32 (−3.5 to 2.9)	.80
NDI[f] (0-50)		
Whole series	4.5 (3.1 to 5.9)	<.001
G1	5.35 (3.2 to 7.5)	<.001
G2	7.4 (5.1 to 9.7)	<.001
CG	−0.95 (−2.9 to 1.05)	.33

[a]CG: control group.

[b]FIQ: Fibromyalgia Impact Questionnaire.

[c]EQ-5D: EuroQol 5-Dimensions.

[d]TSK: Tampa Scale for Kinesiophobia.

[e]CSI: Central Sensitization Inventory.

[f]NDI: Neck Disability Index Questionnaire.

Follow-Up at 1 Month

Summary of the comparison between baseline and 1-month post-intervention scores for FIQ, EQ-5D, TSK, CSI, and NDI (intragroup comparisons). Shown are the differences in the mean values between both time points (Table 2). The differences in the measurement times of the other variables can be observed in Multimedia Appendix 4.

Table . Intragroup (G1, G2, and CG[a]) analysis of variables measured using questionnaires between baseline and 1 month after intervention.

Value and intervention group	Differences of mean values (95% CI)	*P* value
FIQ[b]		
Whole series	4.4 (2.5 to 76.3)	<.001
G1	10.71 (6.6 to 14.8)	<.001
G2	2.71 (0.55 to 4.87)	.02
CG	0.33 (–2.08 to 1.42)	.70
EQ-5D[c]		
Whole series	4.7 (2.6 to 5.7)	<.001
G1	7.7 (5.46 to 9.97)	<.001
G2	4.17 (1.7 to 6.63)	.002
CG	1.07 (–3.87 to 1.73)	.43
TSK[d]		
Whole series	3.3 (1.4 to 5.1)	.001
G1	8.41 (5.25 to 11.58)	<.001
G2	2.1 (0.06 to 4.13)	.04
CG	0.2 (–2.87 to 3.28)	.89
CSI[e]		
Whole series	3.9 (2.18 to 5.8)	<.001
G1	9.78 (7.35 to 12.22)	<.001
G2	2.94 (1.25 to 4.64)	.02
CG	0.26 (–2.76 to 3.3)	.85
NDI[f]		
Whole series	4.03 (2.5 to 6.31)	<.001
G1	10.65 (8.35 to 12.94)	<.001
G2	4.22 (2.39 to 6.06)	<.001
CG	2.05 (–0.33 to 4.43)	.08

[a]CG: control group.

[b]FIQ: Fibromyalgia Impact Questionnaire.

[c]EQ-5D: EuroQol 5-Dimensions.

[d]TSK: Tampa Scale for Kinesiophobia.

[e]CSI: Central Sensitization Inventory.

[f]NDI: Neck Disability Index Questionnaire.

After confirming the effect of both interventions, we compared the intervention groups to see whether VR had a significant influence on the effect of the therapy. Table 3 shows that the ANOVA between groups was significant for FIQ, CSI, and NDI at both postintervention and 1-month follow-up (*P*<.001). Immediately after treatment, both intervention groups (G1 and G2) improved significantly compared with the CG, with no differences between G1 and G2. At 1 month, G1 maintained superior outcomes across all measures, showing significantly better scores than both GC and G2 (*P*<.001), while G2 no longer differed from GC for FIQ and CSI. For NDI, all pairwise comparisons were significant (*P*<.001), confirming that both interventions reduce neck disability, with the VR-enhanced program providing the greatest and most sustained benefit.

Table . Summary of the comparison between the groups for the change in FIQ[a], CSI[b], and NDI[c], both immediately after intervention and 1 month after.

Variable and intervention group	*P* value for ANOVA	Mean difference between time points (95% CI)	*P* value
FIQ			
Immediate	<.001		
CG[d] vs G1		9.31 (14.7 to 3.8)	<.001
CG vs G2		8.4 (13.84 to 3.06)	.001
G2 vs G1		0.85 (6.39 to −4.68)	>.99
1 month	<.001		
CG vs G1		10.38 (5.8 to 14.9)	<.001
CG vs G2		2.38 (–2.12 to 6.9)	.59
G2 vs G1		7.99 (3.35 to 12.63)	<.001
CSI			
Immediate	<.001		
CG vs G1		6.31 (10.96 to 1.66)	.04
CG vs G2		7.53 (12.12 to 2.95)	<.001
G2 vs G1		–1.22 (3.4 to −5.93)	.51
1 month	<.001		
CG vs G1		10.04 (14.16 to 5.93)	<.001
CG vs G2		3.20 (7.26 to −0.84)	.16
G2 vs G1		6.84 (11.01 to 2.67)	.001
NDI			
Immediate	<.001		
CG vs G1		4.40 (7.95 to 0.85)	.01
CG vs G2		6.44 (9.93 to 2.94)	<.001
G2 vs G1		–2.03 (1.55 to −5.62)	.50
1 month	<.001		
CG vs G1		12.69 (16.33 to 9.06)	<.001
CG vs G2		6.27 (9.85 to 2.69)	<.001
G2 vs G1		6.42 (10.10 to 2.74)	<.001

[a]FIQ: Fibromyalgia Impact Questionnaire.

[b]CSI: Central Sensitization Inventory.

[c]NDI: Neck Disability Index Questionnaire.

[d]CG: control group.

Discussion

Principal Findings

This study evaluated the effectiveness of VR as an adjunctive therapy for cervical pain in women with FM. The main findings revealed that combining VR with mobility and strengthening exercises produced significant improvements in pain perception, ROM, and functional performance compared with physical therapy alone. These benefits were also sustained at 1-month follow-up, emphasizing the potential of immersive VR interventions to enhance therapeutic outcomes in individuals with FM.

In the intergroup analysis, the G1 consistently outperformed both G2 and CG in those outcomes most closely related to global disease impact and NDI. Specifically, G1 showed significantly greater and more sustained improvements in FIQ and CSI scores than both CG and G2, whereas improvements in G2 tended to converge toward control values at 1-month follow-up. Similarly, NDI scores decreased in both intervention groups immediately after treatment, but at follow-up, a clear gradient GC>G2>G1 was observed, indicating the largest and most persistent reduction in neck-related disability in the combined intervention. These effects were paralleled by a recurrent pattern of superior gains in cervical ROM and higher-pressure pain thresholds in the trapezius and occipital in G1, while the Time Up Go test

improved in both intervention groups compared with CG and TSK decreased more markedly in G1 at 1 month. By contrast, the VAS and Borg scale showed only limited discrimination between the groups, suggesting that the added value of VR was more evident in multidimensional impact and sensitization-related and functional domains than in isolated pain ratings.

This intergroup pattern is consistent with previous evidence in FM and chronic neck pain, indicating that VR, particularly when used as an adjunct to active exercise, preferentially enhances global impact, disability, and pain-modulation outcomes. In patients with FM, a study reported that fully immersive VR combined with aerobic and pilates training produced greater improvements in pain, kinesiophobia, fatigue, physical activity levels, and mental quality of life than exercise alone, while both groups improved in FM impact, supporting an adjunctive role of immersive VR in comprehensive rehabilitation programs [8]. Other studies showed that, in individuals with chronic neck pain, VR-based cervical training led to larger gains in pressure pain thresholds at the upper cervical levels and greater reductions in functional limitation compared with motor control exercises, despite the absence of between-group differences in pain intensity and quality of life, which parallels our finding of more robust intergroup differences in pressure pain thresholds, ROM, and disability than in VAS score [51]. Furthermore, a randomized crossover trial in women with FM found that VR increased cold pain thresholds and tolerance in both FM patients and pain-free controls, whereas effects on pain intensity were limited, reinforcing the notion that VR may primarily modulate pain processing rather than consistently decreasing reported pain intensity [52]. Taken together, these converging results suggest that combining immersive VR with targeted cervical exercise may primarily potentiate mechanisms related to central sensitization, motor control, and the global impact of FM rather than merely amplifying short-term analgesic effects.

VR has emerged as a promising nonpharmacological intervention for the symptom management of FM. Its immersive and interactive features provide cognitive distraction, emotional engagement, and motor stimulation, which together enhance adherence and reduce pain intensity. Previous studies have demonstrated that VR can improve mood, motivation, and functional capacity when applied alongside traditional rehabilitation or psychological therapies. This approach not only benefits patients with FM but may also be applicable to other chronic pain conditions [49,50,53].

The use of VR has been demonstrated to significantly influence pain relief, motor function, and joint mobility among patients with a range of chronic pathologies [53]. The primary applications of VR in health care include the management of pain and anxiety as well as the enhancement of patient motivation [47,54,55].

While it is not possible to make direct comparisons between this study and research conducted by other authors, the described implementation of an 8-session intervention resulted in statistically significant outcomes for the variables under investigation. This suggests VR is effective when combined with active exercise therapy.

The comparative analysis concerning VR and physical exercise indicated that there was minimal divergence in the measured variables. In a separate study [56], 44 patients underwent an intervention for 4 weeks, with 2 sessions per week. The intervention was divided into 2 groups: G1 (whose members performed only cervical mobility exercises) and G2 (whose members conducted a treatment based solely on VR exercises). The findings indicated that there were no statistically significant differences between the 2 groups across the different variables, including pain, ROM, neck disability, incapacitating pain, and anxiety. However, in this study, the combination of VR with cervical mobility exercises yielded a significantly better outcome. Furthermore, these improvements were maintained over time (1 month) to a greater degree in the group subjected to both therapies combined when compared to the group that performed only mobility exercises. It is important to highlight that the study with which this investigation was compared implemented its intervention in a population without FM. Consequently, the observed discrepancy in outcomes could be due to this factor. A similar study was conducted [51], with an equivalent number of sessions implemented. The outcomes observed in this research were analogous to those obtained in our research, as the investigators combined active therapy and VR exercises for patients with FM [55]. However, the FM cohort in that report comprised only 20 patients. In addition, the active component focused on aerobic training and Pilates instead of the cervical exercises used in our research.

Another study assessed the efficacy of VR for patients with chronic neck pain versus a CG that performed motor control exercises. In this case, the study spanned a period of 6 weeks, comprising 3 weekly sessions. Their findings revealed that the utilization of VR was advantageous when measuring pain due to pressure with an algometer at C1-C2 and C5-C6, and it was also beneficial when considering functional limitations. However, no significant differences were observed between the groups with respect to pain intensity, muscle performance, or quality of life (36-Item Short Form Health Survey). It is evident that the results of this study deviate from those previously observed. However, it should be noted that the mentioned study did not include the population with FM, and no group underwent combined VR and active therapy [51].

Following a thorough review of the scientific literature, we found no studies addressing the use of VR in the cervical region using this protocol. Other studies have applied VR in FM use, games, and global exercises, focusing on observing aspects such as anaerobic capacity, balance, or fatigue [7,54,55]. In this study, the focus was pain in the cervical region and other more global variables such as fatigue, VAS, and FIQ. Therefore, our study offers a novel contribution to this field of research by addressing this gap in the existing literature. Given that patients with FM experience cervical pain, headaches, and tender points in the occipital regions, it is imperative to reduce the associated symptoms.

The limitations of this study lie in the fact that the results are not representative of a heterogeneous population with FM, since the cohort of participants in this research consisted solely of women. Due to the challenging nature of the disease, it was not possible to consider medication used by participants, which

may have influenced the results. Finally, the study did not include a group that used only VR, so we cannot conclude that the results are attributable to the VR component or its combination with exercise.

This approach not only benefits patients with FM but may also be applicable to other chronic pain conditions [49,50,53].

Conclusions

Virtual reality, when added to cervical mobility and strengthening exercises, produced greater and more sustained improvements in disease impact, neck disability, central sensitization, kinesiophobia, and functional outcomes in women with FM than exercise alone or no intervention. The improvements were maintained for 1 month in the G1 group in the variables evaluated. These findings support VR as an effective adjunct to active therapy, with potential benefits for symptom reduction and warrant further research on long-term effects, cost-effectiveness, and applicability to broader chronic pain populations.

Funding

This research received external funding for the FIBYSAR project of Camilo Jose Cela University.

Data Availability

The data presented in this study are available upon reasonable request from the corresponding author.

Authors' Contributions

Conceptualization: AN-M, CO-M, EÚ-D, YM-C
Data curation: BP-R, EC-F-P, EÚ-D, JPH-P
Formal analysis: JH-L, JPH-P, MJF-A, NM-G
Investigation: BP-R, EC-F-P, EÚ-D, JH-L, JPH-P, NM-G, YM-C
Methodology: EC-F-P, JH-L, JPH-P, MC-C
Resources: CO-M, EC-F-P, EÚ-D, JPH-P
Software: AN-M, JH-L, YM-C
Supervision: CO-M, EÚ-D, JPH-P
Validation: EÚ-D, EC-F-P, JPH-P, MC-C, NM-G
Visualization: EÚ-D, JPH-P, MC-C, YM-C
Writing – original draft: AN-M, CO-M, EÚ-D, NM-G, YM-C
Writing – review & editing: BP-R, EC-F-P, EÚ-D, JH-L, JPH-P, MJF-A

Conflicts of Interest

None declared.

Multimedia Appendix 1
Recommended exercises.
[DOCX File, 15 KB - rehab_v13i1e81158_app1.docx]

Multimedia Appendix 2
Baseline measurements of pain and functional capacity (mean and SD).
[DOCX File, 23 KB - rehab_v13i1e81158_app2.docx]

Multimedia Appendix 3
Intragroup analysis of baseline and immediately after intervention.
[DOCX File, 20 KB - rehab_v13i1e81158_app3.docx]

Multimedia Appendix 4
Intragroup analysis of baseline and 1 month after intervention.
[DOCX File, 26 KB - rehab_v13i1e81158_app4.docx]

Checklist 1
CONSORT-Harms-2022 checklist.
[PDF File, 208 KB - rehab_v13i1e81158_app5.pdf]

References

1. Richard JY, Hurley RA, Taber KH. Fibromyalgia: Centralized pain processing and neuroimaging. J Neuropsychiatry Clin Neurosci 2019;31(3):A6-187. [doi: 10.1176/appi.neuropsych.19050107] [Medline: 31322995]
2. Häuser W, Fitzcharles MA. Facts and myths pertaining to fibromyalgia. Dialogues Clin Neurosci 2018 Mar;20(1):53-62. [doi: 10.31887/DCNS.2018.20.1/whauser] [Medline: 29946212]
3. Solà JF. El síndrome de fatiga crónica [Article in Spanish]. Medicina Integral 2002;40(2):56-63 [FREE Full text]
4. Sumpton JE, Moulin DE. Fibromyalgia. Handb Clin Neurol 2014;119:513-527. [doi: 10.1016/B978-0-7020-4086-3.00033-3] [Medline: 24365316]
5. Ruschak I, Montesó-Curto P, Rosselló L, Aguilar Martín C, Sánchez-Montesó L, Toussaint L. Fibromyalgia syndrome pain in men and women: a scoping review. Healthcare (Basel) 2023 Jan 11;11(2):223. [doi: 10.3390/healthcare11020223] [Medline: 36673591]
6. Liang D, Guo X, Zhang J, Jiang R. Pain characteristics of patients with fibromyalgia: a comparison between gender and different emotional states. Pain Physician 2024 Jan 20;27(1):E109-E118. [doi: 10.36076/ppj.2024.27.E109] [Medline: 38285043]
7. Cabo-Meseguer A, Cerdá-Olmedo G, Trillo-Mata JL. Fibromyalgia: prevalence, epidemiologic profiles and economic costs. Med Clin 2017 Nov;149(10):441-448. [doi: 10.1016/j.medcli.2017.06.008]
8. Gulsen C PT, MSc, Soke F PT, PhD, Eldemir K PT, MSc, et al. Effect of fully immersive virtual reality treatment combined with exercise in fibromyalgia patients: a randomized controlled trial. Assist Technol 2022 May 4;34(3):256-263. [doi: 10.1080/10400435.2020.1772900] [Medline: 32543290]
9. Villanueva VL, Valía JC, Cerdá G, et al. Fibromialgia: diagnóstico y tratamiento. El estado de la cuestión [Article in Spanish]. Rev Soc Esp Dolor 2004;11(7):430-443 [FREE Full text]
10. Borchers AT, Gershwin ME. Fibromyalgia: A critical and comprehensive review. Clinic Rev Allerg Immunol 2015 Oct;49(2):100-151. [doi: 10.1007/s12016-015-8509-4]
11. Wang JJ, Tam KW, Hsiao HY, Liou TH, Rau CL, Hsu TH. Effect of resistance exercises on function and pain in fibromyalgia. Am J Phys Med Rehabil 2024;103(4):275-283. [doi: 10.1097/PHM.0000000000002318]
12. Giesecke T, Gracely RH, Grant MAB, et al. Evidence of augmented central pain processing in idiopathic chronic low back pain. Arthritis Rheum 2004 Feb;50(2):613-623. [doi: 10.1002/art.20063]
13. Yunus MB. The prevalence of fibromyalgia in other chronic pain conditions. Pain Res Treat 2012;2012:584573. [doi: 10.1155/2012/584573] [Medline: 22191024]
14. Prikhodkina M, Melnikov S. Factors that influence medication adherence in women with fibromyalgia: a path analysis. J Clin Nurs 2024 Oct;33(10):3943-3953. [doi: 10.1111/jocn.17044] [Medline: 38284436]
15. Desai R, Marlow N. Effect of non-adherence on healthcare resource utilization in fibromyalgia patients: a retrospective cohort study. Value Health 2018 May;21:S255. [doi: 10.1016/j.jval.2018.04.1776]
16. Dobkin PL, Sita A, Sewitch MJ. Predictors of adherence to treatment in women with fibromyalgia. Clin J Pain 2006;22(3):286-294. [doi: 10.1097/01.ajp.0000173016.87612.4b] [Medline: 16514330]
17. Iglesias AR, Soria Ayuda RE, Martinez AB, Sánchez AJ, Villaroya Bielsa E, Callejero Guillen AJ. La efectividad de un programa de ejercicio terapéutico en pacientes con fibromialgia [Web page in Spanish]. Revista Sanitaria de Investigacion. URL: https://revistasanitariadeinvestigacion.com/la-efectividad-de-un-programa-de-ejercicio-terapeutico-en-pacientes-con-fibromialgia/ [accessed 2025-12-30]
18. Winkelmann A, Häuser W, Friedel E, et al. Physiotherapy and physical therapies for fibromyalgia syndrome. Systematic review, meta-analysis and guideline [Title in German]. Schmerz 2012 Jun;26(3):276-286. [doi: 10.1007/s00482-012-1171-3] [Medline: 22760460]
19. Sarac AJ, Gur A. Complementary and alternative medical therapies in fibromyalgia. Curr Pharm Des 2006;12(1):47-57. [doi: 10.2174/138161206775193262] [Medline: 16454724]
20. Collado-Mateo D, Dominguez-Muñoz FJ, Adsuar JC, Merellano-Navarro E, Gusi N. Exergames for women with fibromyalgia: a randomised controlled trial to evaluate the effects on mobility skills, balance and fear of falling. PeerJ 2017;5:e3211. [doi: 10.7717/peerj.3211] [Medline: 28439471]
21. Ahmadpour N, Randall H, Choksi H, Gao A, Vaughan C, Poronnik P. Virtual reality interventions for acute and chronic pain management. Int J Biochem Cell Biol 2019 Sep;114(September):105568. [doi: 10.1016/j.biocel.2019.105568] [Medline: 31306747]
22. Huang Q, Lin J, Han R, Peng C, Huang A. Using virtual reality exposure therapy in pain management: a systematic review and meta-analysis of randomized controlled trials. Value Health 2022 Feb;25(2):288-301. [doi: 10.1016/j.jval.2021.04.1285] [Medline: 35094802]
23. Cortés-Pérez I, Zagalaz-Anula N, Ibancos-Losada MDR, Nieto-Escámez FA, Obrero-Gaitán E, Osuna-Pérez MC. Virtual reality-based therapy reduces the disabling impact of fibromyalgia syndrome in women: systematic review with meta-analysis of randomized controlled trials. J Pers Med 2021 Nov 9;11(11):1167. [doi: 10.3390/jpm11111167] [Medline: 34834518]
24. Carvalho MSD, Carvalho LC, Menezes FDS, Frazin A, Gomes EDC, Iunes DH. Effects of exergames in women with fibromyalgia: a randomized controlled study. Games Health J 2020 Oct 1;9(5):358-367. [doi: 10.1089/g4h.2019.0108]
25. Gómez-Pérez L, López-Martínez AE, Ruiz-Párraga GT. Psychometric properties of the Spanish version of the Tampa Scale for Kinesiophobia (TSK). J Pain 2011 Apr;12(4):425-435. [doi: 10.1016/j.jpain.2010.08.004] [Medline: 20926355]

26. Sato K, Fukumori S, Matsusaki T, et al. Nonimmersive virtual reality mirror visual feedback therapy and its application for the treatment of complex regional pain syndrome: an open-label pilot study. Pain Med 2010 Apr;11(4):622-629. [doi: 10.1111/j.1526-4637.2010.00819.x] [Medline: 20202141]
27. Freeman D, Reeve S, Robinson A, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol Med 2017 Oct;47(14):2393-2400. [doi: 10.1017/S003329171700040X] [Medline: 28325167]
28. Lo HHM, Zhu M, Zou Z, et al. Immersive and nonimmersive virtual reality-assisted active training in chronic musculoskeletal pain: systematic review and meta-analysis. J Med Internet Res 2024 Aug 19;26:e48787. [doi: 10.2196/48787] [Medline: 39159449]
29. Harvie DS, Broecker M, Smith RT, Meulders A, Madden VJ, Moseley GL. Bogus visual feedback alters onset of movement-evoked pain in people with neck pain. Psychol Sci 2015 Apr;26(4):385-392. [doi: 10.1177/0956797614563339] [Medline: 25691362]
30. Trost Z, Zielke M, Guck A, et al. The promise and challenge of virtual gaming technologies for chronic pain: the case of graded exposure for low back pain. Pain Manag 2015;5(3):197-206. [doi: 10.2217/pmt.15.6] [Medline: 25971643]
31. Sarig Bahat H, Takasaki H, Chen X, Bet-Or Y, Treleaven J. Cervical kinematic training with and without interactive VR training for chronic neck pain—a randomized clinical trial. Man Ther 2015 Feb;20(1):68-78. [doi: 10.1016/j.math.2014.06.008] [Medline: 25066503]
32. Ubillos-Landa S, García-Otero R, Puente-Martínez A. Validation of an instrument for measuring chronic pain in nursing homes. An Sist Sanit Navar 2019 Apr 25;42(1):19-30. [doi: 10.23938/ASSN.0390] [Medline: 30706908]
33. Úbeda-D'Ocasar E, Valera-Calero JA, Hervás-Pérez JP, Caballero-Corella M, Ojedo-Martín C, Gallego-Sendarrubias GM. Pain intensity and sensory perception of tender points in female patients with fibromyalgia: a pilot study. Int J Environ Res Public Health 2021 Feb 4;18(4):1461. [doi: 10.3390/ijerph18041461] [Medline: 33557288]
34. Ayala Pastorino S, Varaldi Azcoytia G, Illescas Calegari L, Castroman Espasandín Espasandin PJ. Test cuantitativo sensorial con algometría de presión: experiencia preliminar en un servicio de anestesiología [Article in Spanish]. Rev Soc Esp Dolor 2022;28(6):311-318. [doi: 10.20986/resed.2022.3947/2021]
35. Ge HY, Nie H, Madeleine P, Danneskiold-Samsøe B, Graven-Nielsen T, Arendt-Nielsen L. Contribution of the local and referred pain from active myofascial trigger points in fibromyalgia syndrome. Pain 2009 Dec 15;147(1-3):233-240. [doi: 10.1016/j.pain.2009.09.019] [Medline: 19819074]
36. Walthall H, Jenkinson C, Boulton M. Living with breathlessness in chronic heart failure: a qualitative study. J Clin Nurs 2017 Jul;26(13-14):2036-2044. [doi: 10.1111/jocn.13615] [Medline: 27731919]
37. Jovicic MD, Konstantinovic LM, Grgurevic AD, et al. Validation of the Neck Disability Index in Serbian patients with cervical radiculopathy. J Manipulative Physiol Ther 2018;41(6):496-502. [doi: 10.1016/j.jmpt.2017.10.018] [Medline: 30107938]
38. Roelofs J, Sluiter JK, Frings-Dresen MHW, et al. Fear of movement and (re)injury in chronic musculoskeletal pain: evidence for an invariant two-factor model of the Tampa Scale for Kinesiophobia across pain diagnoses and Dutch, Swedish, and Canadian samples. Pain 2007 Sep;131(1-2):181-190. [doi: 10.1016/j.pain.2007.01.008] [Medline: 17317011]
39. Newman-Beinart NA, Norton S, Dowling D, et al. The development and initial psychometric evaluation of a measure assessing adherence to prescribed exercise: the Exercise Adherence Rating Scale (EARS). Physiotherapy 2017 Jun;103(2):180-185. [doi: 10.1016/j.physio.2016.11.001] [Medline: 27913064]
40. Gauffin J, Hankama T, Kautiainen H, Arkela-Kautiainen M, Hannonen P, Haanpää M. Validation of a Finnish version of the Fibromyalgia Impact Questionnaire (Finn-FIQ). Scand J Pain 2012 Jan 1;3(1):15-20. [doi: 10.1016/j.sjpain.2011.10.004] [Medline: 29913759]
41. Monterdea S, Salvata I, Montulla S, Fernández-Ballart J. Validación de la versión española del Fibromyalgia Impact Questionnaire [Article in Spanish]. Rev Esp Reumatol 2004;31:507-513 [FREE Full text]
42. Esteve-Vives J, Rivera Redondo J, Isabel Salvat Salvat M, de Gracia Blanco M, de Miquel CA. Propuesta de una versión de consenso del Fibromyalgia Impact Questionnaire (FIQ) para la población española. Reumatol Clín 2007 Jan;3(1):21-24. [doi: 10.1016/S1699-258X(07)73594-5] [Medline: 21794391]
43. Herdman M, Badia X, Berra S. El EuroQol-5D: una alternativa sencilla para la medición de la calidad de vida relacionada con la salud en atención primaria. Aten Primaria 2001;28(6):425-429. [doi: 10.1016/S0212-6567(01)70406-4]
44. Wolfe F, Michaud K, Li T, Katz RS. EQ-5D and SF-36 quality of life measures in systemic lupus erythematosus: comparisons with rheumatoid arthritis, noninflammatory rheumatic disorders, and fibromyalgia. J Rheumatol 2010 Feb;37(2):296-304. [doi: 10.3899/jrheum.090778] [Medline: 20032098]
45. Carralero García P, Rosa Hoyos Miranda F, Deblas Sandoval Á, López García M. Calidad del sueño según el Pittsburgh Sleep Quality Index en una muestra de pacientes recibiendo cuidados paliativos. Med Paliat 2013;20(2):44-48. [doi: 10.1016/j.medipa.2012.05.005]
46. Salaffi F, Farah S, Mariani C, Sarzi-Puttini P, Di Carlo M. Validity of the Central Sensitization Inventory compared with traditional measures of disease severity in fibromyalgia. Pain Pract 2022 Nov;22(8):702-710. [doi: 10.1111/papr.13162] [Medline: 36097821]
47. Prushansky T, Dvir Z. La prueba de la movilidad cervical: metodología e implicaciones clínicas [Article in Spanish]. Osteopat Cient 2008;3:108-114. [doi: 10.1016/S1886-9297(08)75759-X]

48. Collado-Mateo D, Domínguez-Muñoz FJ, Adsuar JC, Merellano-Navarro E, Olivares PR, Gusi N. Reliability of the timed up and go test in fibromyalgia. Rehabil Nurs 2018;43(1):35-39. [doi: 10.1002/rnj.307] [Medline: 27781288]
49. Beaudart C, Rolland Y, Cruz-Jentoft AJ, et al. Assessment of muscle function and physical performance in daily clinical practice: a position paper endorsed by the European Society for Clinical and Economic Aspects of Osteoporosis, Osteoarthritis and Musculoskeletal Diseases (ESCEO). Calcif Tissue Int 2019;105(1):1-14. [doi: 10.1007/s00223-019-00545-w]
50. Declaración de Helsinki de la AMM [Web page in Spanish]. Asociación Médica Mundial. 2024. URL: https://www.wma.net/es/policies-post/declaracion-de-helsinki-de-la-amm-principios-eticos-para-las-investigaciones-medicas-en-seres-humanos/ [accessed 2026-01-05]
51. Cetin H, Kose N, Oge HK. Virtual reality and motor control exercises to treat chronic neck pain: a randomized controlled trial. Musculoskelet Sci Pract 2022 Dec;62(August):102636. [doi: 10.1016/j.msksp.2022.102636] [Medline: 35952621]
52. Christensen SWM, Almsborg H, Vain TS, Vaegter HB. The effect of virtual reality on cold pain sensitivity in patients with fibromyalgia and pain-free individuals: a randomized crossover study. Games Health J 2023 Aug;12(4):295-301. [doi: 10.1089/g4h.2022.0138] [Medline: 36454199]
53. Garcia-Palacios A, Herrero R, Vizcaíno Y, et al. Integrating virtual reality with activity management for the treatment of fibromyalgia: acceptability and preliminary efficacy. Clin J Pain 2015 Jun;31(6):564-572. [doi: 10.1097/AJP.0000000000000196] [Medline: 25551475]
54. Darnall BD, Krishnamurthy P, Tsuei J, Minor JD. Self-administered skills-based virtual reality intervention for chronic pain: randomized controlled pilot study. JMIR Form Res 2020 Jul 7;4(7):e17293. [doi: 10.2196/17293] [Medline: 32374272]
55. Herrero R, García-Palacios A, Castilla D, Molinari G, Botella C. Virtual reality for the induction of positive emotions in the treatment of fibromyalgia: a pilot study over acceptability, satisfaction, and the effect of virtual reality on mood. Cyberpsychol Behav Soc Netw 2014 Jun;17(6):379-384. [doi: 10.1089/cyber.2014.0052] [Medline: 24892201]
56. Tejera DM, Beltran-Alacreu H, Cano-de-la-Cuerda R, et al. Effects of virtual reality versus exercise on pain, functional, somatosensory and psychosocial outcomes in patients with non-specific chronic neck pain: a randomized clinical trial. Int J Environ Res Public Health 2020 Aug 16;17(16):5950. [doi: 10.3390/ijerph17165950] [Medline: 32824394]

Abbreviations

CG: control group
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-Harms: Consolidated Standards of Reporting Trials—Harms Extension
CSI: Central Sensitization Inventory
EQ-5D: EuroQol Dimensions
FIQ: Fibromyalgia Impact Questionnaire
FM: fibromyalgia
NDI: Neck Disability Index Questionnaire
ROM: range of motion
TSK: Tampa Scale for Kinesiophobia
VAS: visual analog scale
VR: virtual reality

Edited by S Munce; submitted 23.Jul.2025; peer-reviewed by D Parmak, GGD Alencar; revised version received 10.Dec.2025; accepted 19.Dec.2025; published 23.Jan.2026.

Please cite as:
Úbeda-D'Ocasar E, Moreno-Crespo Y, Cimadevilla-Fernández-Pola E, Hernández-Lougedo J, Navas-Mosqueda Á, Caballero-Corella M, Mayoral-Gonzalo N, Pedauyé-Rueda B, Fernández-Aceñero MJ, Hervás-Pérez JP, Ojedo-Martín C
Therapeutic Use of Virtual Reality for Patients With Fibromyalgia and Chronic Neck Pain: Randomized Controlled Trial
JMIR Rehabil Assist Technol 2026;13:e81158
URL: https://rehab.jmir.org/2026/1/e81158
doi:10.2196/81158

Linguistic Validation and Cross-Cultural Adaptation of the Shoulder Telehealth Assessment Tool for Filipino Patients with Musculoskeletal Shoulder Condition: Cross-Sectional Study

Jeffrey Arboleda[1], PTRP, MD, MBA; Sharon Ignacio[1,2], MD; Jose Alvin Mojica[1], MD, MHPEd; Carl Froilan Leochico[2,3], PTRP, MD, MHSc

[1]Department of Rehabilitation Medicine, Philippine General Hospital, University of the Philippines Manila, Taft Avenue, Ermita, Metro Manila, Philippines

[2]Department of Physical Medicine and Rehabilitation, St. Luke's Medical Center, Metro Manila, Philippines

[3]Department of Medicine, Division of Physical Medicine and Rehabilitation, Toronto Rehabilitation Institute, University Health Network, Toronto, ON, Canada

Corresponding Author:
Jeffrey Arboleda, PTRP, MD, MBA
Department of Rehabilitation Medicine, Philippine General Hospital, University of the Philippines Manila, Taft Avenue, Ermita, Metro Manila, Philippines

Abstract

Background: Telerehabilitation has been widely adopted to meet the growing rehabilitation demand, but it is often limited by unstable internet connection, poor audiovisual resolution, and difficult virtual assessment. The Shoulder Telehealth Assessment Tool (STAT), a comprehensive, patient-led, preconsultation shoulder physical examination pictorial guide, was developed to address these limitations by easing the communication of instruction during the consultation and potentially removing the need for video calls.

Objective: This study aimed to develop a linguistically valid and culturally appropriate Filipino version of STAT and to evaluate its content validity, internal consistency, understandability, and ease of use.

Methods: A cross-sectional study on the Filipino STAT was conducted in three phases: (1) linguistic validation by experts, (2) cross-cultural adaptation through pretesting of 12 participants diagnosed with a musculoskeletal shoulder condition at the Philippine General Hospital, and (3) pilot study on 47 participants of the same population.

Results: The Filipino STAT had an excellent content validity (scale validity index=0.80 - 0.97), excellent interrater reliability (κ coefficient=0.82 - 1.00), and good internal consistency (Cronbach α=0.87). Understandability was found to be excellent for pain and activity (98%), good for range of motion and special tests (85%), and poor for strength (37%). However, 24% (11/46) of participants perceived the tool difficult to understand with the use of some Tagalog words as the primary barrier, followed by non-familiarity with the tool and difficulty reading the text.

Conclusions: Development of the Filipino STAT through a rigorous linguistic validation and cultural adaptation has produced a culturally appropriate, valid, and reliable tool. Pain and activity, range of motions, and special test subdomains are suitable for clinical assessment, while strength subdomain needs further improvement in understandability.

(JMIR Rehabil Assist Technol 2026;13:e67974) doi:10.2196/67974

KEYWORDS

cross-cultural adaptation; linguistic validation; musculoskeletal rehabilitation; shoulder examination; Shoulder Telehealth Assessment Tool; telerehabilitation; STAT

Introduction

Telemedicine is the use of telecommunication technologies to deliver health care, public health, and health education services remotely [1]. Telerehabilitation is a branch of telemedicine specifically aimed at delivering rehabilitation services. Being an archipelago with limited health care professionals and facilities, the Philippines uses these innovative health care delivery services to improve health access for Filipinos. Telerehabilitation in the Philippines was first adopted in the context of community-based rehabilitation in 2017 [2] and was widely used in 2020 during the COVID-19 pandemic [3]. With the improving COVID-19 situation, telerehabilitation remains a viable solution to delivering rehabilitation services in far-flung areas of the country [4].

However, telerehabilitation is not without limitations, such as unstable internet connection, lack of confidence in establishing clinical diagnosis virtually, limited time allotment per patient, and poor audiovisual resolution [5]. Specifically, in the administration of virtual physical examination, rehabilitation professionals face difficulties in the conduct of special tests, range of motion (ROM) assessment, and strength testing, among others [3]. These challenges have opened opportunities for innovation in telerehabilitation.

One of these innovations was the development of the Shoulder Telehealth Assessment Tool (STAT), which is a comprehensive patient-led shoulder physical examination pictorial guide done prior to the actual teleconsultation, aimed to improve clinical efficiency [6]. It is the first published patient-reported outcome measure to simulate the performance of in-person physical examination, which includes special tests for screening of different shoulder pathologies [7]. A validated visual analog scale (VAS) [8], single assessment numeric evaluation (SANE) [9], and motion analysis and range of motion studies on activities of daily living [10] have also been integrated into the tool. The first subcategory of the STAT is composed of 3 questions for pain and activity. Both VAS and SANE are patient-reported. The VAS ranges from 0=no pain to 10=maximum pain, while the SANE is scored from 0% to 100% of normal. Meanwhile, the current level of daily activity is a nominal score (ie, unaffected sleep, full work, and full recreation or sport). The ROM testing has 9 questions that are answered with either yes (movement completed) or no (movement not completed). There are 5 items for the shoulder special tests answerable by yes or no, and finally 3 items for strength that are answered with either painful, weak, both, or none. Overall, there are a total of 20 questions in the STAT that can be completed in 30 to 45 minutes. In the Philippines and numerous other countries where Filipinos are found, the STAT can potentially improve the accuracy of virtual physical examination techniques for the shoulder, provide better rehabilitation care for patients who do not have stable internet connections or video call capacity, and optimize the delivery of telerehabilitation services. This may also be used by providers from various health care disciplines, such as physiatry, physical therapy, occupational therapy, orthopedics, rheumatology, and pain management, among others, which routinely conduct musculoskeletal examination of the shoulder.

In the Philippines, shoulder conditions have been found to be prevalent among Filipino office workers [11], as well as migrant workers [12]. Currently, the STAT is available in English [6] and has not been translated into any other language. Language and cultural differences call for a careful adaptation of health outcome measures to accurately reflect the cultural nuances and context of the target language version [13]. Hence, this study aimed to develop a linguistically valid and culturally appropriate Filipino version of the STAT and determine its content validity, internal consistency, understandability, and ease of use.

Methods

Study Design

This cross-sectional study used a mixed methods research design on the linguistic validation and cross-cultural adaptation of the Filipino STAT.

Participants

The target population was Filipino adults (aged at least 18 years) with unilateral shoulder pain for at least 6 weeks who consulted in-person or through telemedicine at the Philippine General Hospital (PGH)—Outpatient Department of Rehabilitation Medicine from January 2023 to June 2023. Based on the Department census during that period, there were 304 patients diagnosed with shoulder pathology, such as adhesive capsulitis, rotator cuff injury, and myofascial pain syndrome. The participants should have access to a stable internet and a device with video call capacity and be able to understand written instructions in Filipino or Tagalog, which is the Philippines' most commonly used language. Participants with a history of trauma, suspicion or diagnosis of upper extremity fracture, severe cognitive impairment, known psychiatric comorbidity, cerebrovascular disease, cervical radiculopathy, brachial plexus injury, upper extremity peripheral nerve injury, complex regional pain syndrome, and prior shoulder, neck, and breast surgery were excluded. The participants were allowed to withdraw from the study for any reason.

Sampling

The sample size for the pretest was based on a generally recommended sample size of 12 from the target population based on feasibility and precision of estimates for succeeding studies [14]. The sample size for a pilot study to determine internal consistency through Cronbach α was 46 [15]. Systematic randomized sampling was done where every seventh patient was selected.

Study Procedure

Overview

Adapting the recommendations on linguistic validation and cross-cultural adaptation of self-reported measures, the study procedure was divided into two phases [16,17]: (1) translation of the STAT from the original English to the Filipino version; and (2) cross-cultural adaptation through pretesting. An additional phase was added for internal consistency testing of the final version of the Filipino STAT (Figure 1). Meetings, where necessary, were all done virtually, and permission to record each online meeting was sought from all participants.

Figure 1. This is a flowchart adapted from Lorca et al [17] for linguistic validation and cross-cultural adaptation [6]. STAT: Shoulder Telehealth Assessment Tool.

Phase 1: Linguistic Validation

Step 1: Permission From the Instrument's Authors

Permission for linguistic validation and cross-cultural adaptation into the Filipino language was obtained from the STAT developers prior to initiation of translation.

Step 2: Forward Translation

The forward translation was independently done by 2 bilingual translators. The first forward translator (T1) was a licensed physical therapist with a Master's degree in physical therapy who provided a "reliable equivalence from a clinical perspective," while the second forward translator (T2) was a university instructor and a representative from the *Sentro ng Wikang Filipino* (National Center of the Filipino Language) with no medical background who served to "reflect the language used by the population."

Step 3: Review and Synthesis

The translators, together with a recording observer, convened to synthesize their versions and arrive at a consensus version. Translation and synthesis forms were used throughout the linguistic validation process by the translators and recording observer to document the translated version, rationale for changes, and any disputes or challenges in doing the translation or synthesis.

Step 4: Backward Translation

Backward translation was independently done by another pair of bilingual translators (B1 and B2), who happened to be high school Filipino language teachers and were not familiar with the original version. This step was done to prevent information bias and to detect any errors in the forward translation.

Step 5: Content Validation

The first forward- and backward-translated version of the Filipino STAT was then assessed for content validity independently by clinical experts using a content validation form. The experts were composed of 6 board-certified local physiatrists with expertise in the fields of sports medicine, musculoskeletal ultrasound, interventional physiatry, telerehabilitation, clinical anatomy, and kinesiology. The form was composed of a yes or no scale for clarity, 4-point Likert scale for relevance, and a comment section for each item.

Step 6: Synthesis and Conciliation

All versions of the Filipino STAT available so far and the completed content validation forms (Multimedia Appendix 1) were then reviewed by a consensus committee of experts to arrive at a prefinal version. The consensus committee consisted of all 4 translators (T1, T2, B1, and B2) from the previous steps and the 6 clinical experts who participated in the content validation step. They all met virtually and collectively decided and agreed on each item through the Expert Consensus Committee Guide on Equivalence Form (Multimedia Appendix 2), wherein semantic, idiomatic, experiential, and conceptual equivalence were assessed between the forward- and backward-translated English and Filipino versions. Semantic equivalence pertains to the use of words that have similar, unique meaning in both cultures. Idiomatic equivalence pertains to the use of equivalent colloquialisms and idioms in both cultures. Conceptual equivalence pertains to the use of words or phrases with equivalent conceptual meanings in both cultures, and experiential equivalence pertains to experiences elicited that are consistent or equivalent in both cultures. The form contained a yes or no scale on each type of equivalence, a column for rewording suggestions, and another column for other comments and suggestions to meet cultural equivalence.

Phase 2: Cross-Cultural Adaptation

Step 1: Pretesting

A pretest on 12 adult Filipino patients with shoulder pain was done using the prefinal version of the translated tool through video consultation to simulate the actual STAT procedure and to avoid unnecessary risk of contagion during the pandemic (time of study).

A think-aloud protocol during each video call with the principal investigator or research assistant using an encrypted platform, such as Zoom (Zoom Communications, Inc) or Viber (whichever the participant preferred), was used. The think-aloud protocol entailed each participant to read, perform, and answer the Filipino STAT while verbalizing their thoughts. This was done twice for each question: first without the pictorial guide; and second with the pictorial guide. The trial without the pictorial guide was done to assess the feasibility of the Filipino STAT Text Version for patients with no access to smartphones. An observational checklist (Multimedia Appendix 3) was filled out by the principal investigator to document the participant response, understandability of the tool, and ease and accuracy in performing the Filipino STAT, noting participants' quality of movements and compensatory movements. The principal investigator also asked each participant open-ended interview questions (Multimedia Appendix 4) on their experience (including perceived barriers and facilitators) in using the tool and their suggestions in improving it.

Step 2: Review of Results

General and question-specific errors and suggestions from the pre-test participants were reviewed by the committee in developing the final version of the Filipino STAT.

Step 3: Proofreading

The resulting version was then proofread by a Filipino language teacher to correct any spelling and grammatical errors. All translation forms and STAT versions were sent to the original authors for their review and feedback.

Phase 3: Internal Consistency Testing

Finally, all 46 participants underwent the same process as pretesting using the final version of the Filipino STAT for the pilot study to determine the internal consistency of the tool.

Data Analysis

Content validation was determined through computation of item-level and scale-level content validity indices based on average and universal agreement methods. Sociodemographic data were reported using descriptive statistics. The internal consistency of participant responses was determined through Cronbach α. This was evaluated separately for items in VAS,

SANE, ROM, strength tests, and special tests using STATA 16.1/IC (StataCorp LLC). Considering ≥90% as excellent, 89% to 70% as good, 69% to 50% as fair, and ≤49% as poor, understandability was assessed using the worst-performing item in the observational checklist.

Interview responses were uploaded to NVivo 12 (Lumivero) for organization and thematic analysis of unstructured data pertaining to understandability. An inductive approach was employed where the data content directed the development of themes. The members of the research team were all physiatrists and had broad experience in musculoskeletal evaluation and telerehabilitation in clinical practice and research. JA was a rehabilitation medicine resident trained during the height of telerehabilitation during the COVID-19 pandemic. CFL spearheaded several research endeavors on telerehabilitation in the Philippines. SI and JM were the chairs of the Department of Rehabilitation Medicine with extensive teaching, research, and clinical experiences.

Ethical Considerations

The study was approved by the University of the Philippines Manila Research Ethics Board (UPMREB Code 2022-0516-01). Informed consent was secured through Google Forms, written and explained in Filipino either by the principal investigator or research assistant. Study introduction, objectives, procedure, duration, risks, benefits, incentives, and contact information were discussed with the participants. Participants were also assured of the confidentiality of their information, the voluntary nature of participation, and their right to refuse at any point of the study. The translators, recording observers, and Filipino teacher were appropriately compensated with Philippine 1000 each (US $20). All phase 2 participants were compensated with Philippine 200 (US $4) worth of prepaid load as reimbursement for internet fee for the video call interview and answering of online forms. The clinical experts did not receive remuneration and were acknowledged in this paper.

Results

Forward Translation

Both forward translators had challenges in translating the strength and special tests questions due to absence of direct translation of some English words to Filipino. Hence, both translators resorted to rephrasing some English sentences in Filipino. During synthesis, T1 shared the technical context of the subtests, while T2 suggested the use of understandable albeit literal translations of some words. For item Q9 (ROM) (Table 1), the use of the phrase "*itupi ang siko*" was selected over "*itupi ang braso*" based on T1's input to pertain to the correct anatomical joint that does flexion. On the other hand, for item Q12 (special test), "*idiin ang kamay sa tiyan*" was selected over "*pindutin ang tiyan*" based on T2's input on the context of pressing the belly. Items that resulted in differences between T1 and T2, as well as the final terminology agreed upon during synthesis, are presented in Table 1.

Table . Results of the forward translation and synthesis steps.

Item	T1[a]	T2[b]	Prefinal STAT[c]
Pain and activity			
Q1	*Lebel* *Pinakamasakit na naramdaman sa iyong buhay*	*Antas* *Pinakamasakit*	*Lebel* *Pinakamasakit*
Q3	*Dalas ng pisikal na aktibidad* • *Hindi apektado ang pagtulog* • *Nagtatrabaho nang buong araw* • *Nagagawa ang panlibangang aktibidad o isports*	*Antas ng pang-araw-araw na aktibidad* • *Di naaabalang pagtulog* • *Panay trabaho* • *Panay lingan/ isports*	*Antas ng pang-araw-araw na aktibidad* • *'Di naaabalang pagtulog* • *Nakakapagtrabaho nang buong araw* • *Nagagawa ang panlibangang aktibidad o isports*
Range of motion	*Saklaw ng galaw*	*Saklaw ng paggalaw*	*Saklaw ng paggalaw*
Instructions	*Kilos*	*Paggalaw*	*Paggalaw*
Q4	*Taas ng ulo*	*Tuktok ng ulo*	*Taas ng ulo*
Q7	*Bulsa ng pantalon sa parehong panig*	*Bulsa sa likod*	*Bulsa sa likod ng apektadong balikat*
Q8	*Ibabang likod*	*Ibabang bahagi ng iyong likuran*	*Ibabang bahagi ng iyong likuran*
Q9	*Braso ay nakaharap sa pader* *Itupi ang siko* *Ilapat*	*Nakatagilid sa pader* *Itupi ang braso* *Idikit*	*Nakatagilid sa pader* *Itupi ang braso* *Ilapat*
Strength			
Instructions	*Bigyan ng pwersa* *Mahina ba ang pakiramdam?* *Wala sa mga nabanggit*	*Habang nilalabanan ang bigat* *Nanghihina ba?* *Wala*	*Bigyan ng pwersa* *Mahina ba ang pakiramdam?* *Wala*
Q10	*Gamit ang kamao* *Idiin ang palad ng kamay sa masakit na braso*	*Nakakuyom ang kamao* *Idiin ang mga ito sa isa't isa*	*Gamit ang kamao* *Idiin ang palad ng kamay sa masakit na braso*
Q11	*Hilahin* *Direksyon ng kabilang braso*	*Kapitan* *Gumalaw pakanan at pakaliwa*	*Hilahin* *Direksyon ng normal na braso*
Q12	*Labanan* *Ilayo*	*Kapitan* *Palayo*	*Kapitan* *Ilayo*
Special tests			
Instructions	*Nakasaad na kilos at sasabihin mo*	*Maniobra na makakapagsabi sa amin*	*Mga kilos na makakapagsabi sa amin*
Q13	*Iangat ang buong braso*	*Itaas ang kamay*	*Itaas ang kamay*
Q14	*Umabot lagpas sa kabilang balikat*	*Ilagay ang inyong apektadong braso sa harap ng inyong dibdib*	*Umabot lagpas sa kabilang balikat*
Q15	*Pindutin*	*Idiin ang kamay*	*Idiin ang kamay*
Q16	*Ilapat*	*Idikit*	*Ilapat*
Q17	*Itaas ang braso na parang may hawak na plato* *Dahan-dahang itulak pababa*	*Iangat ang braso sa harap ng inyong dibdib at ilahad ang palad* *Bahagyang diinan*	*Itaas ang braso na parang may hawak na plato* *Dahan-dahang itulak pababa*

[a]T1: forward translator 1.

[b] T2: forward translator 2.

[c]STAT: Shoulder Telehealth Assessment Tool.

Backward Translation

Similar to the forward translators, the back translators did not apply word-for-word translation in strength and special test questions to simplify the items and make them easier to understand. The back translation versions were generally consistent with the original English version, with no significant differences, as agreed upon by the consensus committee.

Content Validation of Prefinal STAT

The instructions in all subdomains, except for strength, were clear to the experts. The items on pain and activity (Q1) and range of motion (Q4, Q5, and Q6) were clear to all experts. Content validity through the scale validity index was measured using the item-level content validity index (0.97) and using universal agreement (0.80). κ values across all items ranged from 0.82 to 1.00, indicating excellent inter-rater reliability.

Cultural Equivalence

Pertinent findings from the discussion of the consensus experts on equivalence are summarized in Table 2.

Table . Pertinent results from the expert consensus discussion on equivalence[a].

	SE[b]	IE[c]	CE[d]	EE[e]	Comments
Pain and activity					
Q1	✓	✓	✓	✓	• Inconsistent: not a question
Q2	✓	✓	✓	✓	• Inconsistent: not a question
Q3	✓	✓	✓	X	• Inconsistent: not a question • Vague: choices not in one spectrum • EE: did not include informal work
Range of motion					
Q7	✓	✓	X	✓	• CE: not relatable to everyone
Special test					
Instruction	✓	X	✓	✓	• IE: Test demands active patient response
Q13	✓	✓	✓	✓	• Safety: for severely painful shoulder • Unclear: confusing literal translation
Q17	✓	✓	X	✓	• CE: incomplete instruction for the test

[a]Checks (✓) indicate equivalence, while cross marks (X) indicate non-equivalence.

[b]SE: semantic equivalence.

[c]IE: idiomatic equivalence.

[d]CE: conceptual equivalence.

[e]EE: experiential equivalence.

To facilitate consistency of the translations throughout the tool, items Q1 to Q3 (pain and activity) were converted from declarative statements into questions. Choices for item Q3 (sleep, work, and recreation or sports) were deemed to pertain to different aspects of activities and were thus converted to stand-alone questions for each activity (Q3A, Q3B, and Q3C). To facilitate clarity, the literal translation of item Q13 was rephrased for easier understanding. To ensure patient safety when using the tool, precautions for severely painful shoulders were added to the instructions for the special test subdomain.

To facilitate experiential equivalence, item Q3B's use of the term "*nakakapagtrabaho*" (able to work) was supplemented by the modifier "*gawaing bahay*" (house chores) to be more encompassing of the different types of work in the Filipino culture. To facilitate conceptual equivalence, item Q7's (ROM) use of "*bulsa sa likod ng apektadong balikat*" (back pocket on the affected side) was changed to "*pigi*" (buttock) as it pertains to the same area and is more relatable and easily understandable. Likewise, item Q17 was clarified with the phrase "*paharap sa lebel ng braso*" (forward raising to arm level) to demonstrate shoulder flexion to 90 degrees as in the Speed test. To facilitate idiomatic equivalence, instruction for the special test subdomain was improved from "*na makakapagsabi sa amin*" (that may tell us) to "*upang malaman namin*" (so that we will know) as the test demands active patient response. Finally, the expert committee all suggested taking new pictures for the tool with a Filipino-looking model for the pictorial guide for it to be more relatable and also to facilitate minor improvements in the angle of the shots and designation of the task of each arm.

Pretest

There were 12 participants in the pretest, with the majority being females (10/12, 83%), aged 50 - 59 years (5/12, 42%) with a mean age of 53 (SD 12) years, married or cohabiting (5/12, 42%), and having finished tertiary education (8/12, 67%; Table 3). Most were unemployed (6/12, 50%), while those employed were engaged in nonhealth-related work (8/12, 67%). The monthly family income of participants was positively skewed, with the majority (7/12, 58%) earning Philippine 5000 to 10,000 (US $100 - $200).

Table . Sociodemographic profile of the participants.

Characteristics	Pretest (n=12)	Pilot (n=47)
Age group (y), n (%)		
19 - 29	0 (0)	2 (4)
30 - 39	2 (2)	5 (11)
40 - 49	2 (2)	3 (6)
50 - 59	5 (42)	21 (45)
≥60	3 (25)	16 (34)
Age (y), mean (SD)	53 (12)	55 (14)
Sex, n (%)		
Female	10 (83)	35 (75)
Male	2 (17)	12 (26)
Civil status, n (%)		
Single	4 (33)	10 (21)
Married or cohabiting	5 (42)	29 (62)
Separated or divorced	1 (8)	2 (4)
Widowed	2 (17)	6 (13)
Educational status, n (%)		
Primary	0 (0)	4 (9)
Secondary	4 (33)	15 (32)
Tertiary	8 (67)	26 (55)
Postgraduate	0 (0)	2 (4)
Employment, n (%)		
Student	0 (0)	2 (4)
Employed	4 (33)	15 (32)
Unemployed	6 (50)	23 (49)
Retired	2 (17)	7 (15)
Type of work, n (%)		
Health-related work	4 (33)	7 (15)
Nonhealth-related work	8 (67)	40 (85)
Family monthly income (Philippine [US $]), n (%)		
<5000 (<100)	2 (17)	21 (45)
5000-10,000 (100 - 200)	7 (58)	9 (19)
10,001-20,000 (201 - 400)	2 (18)	8 (17)
20,001-50,000 (401 - 1000)	1 (8)	6 (13)
>50,000 (>1000)	0 (0)	3 (6)

Items on pain and activity (Q1 and Q3A-Q3C), ROM (Q2-Q6), and special tests (Q15) were performed correctly by at least 75% (9/12) of the participants independently. Up to 25% (3/12) of the participants needed cueing in the whole pain and activity subdomain and special tests (Q14). Less than 25% of the participants were able to perform ROM items Q8 and Q9, all strength items, and half of the items in the special test subdomain.

Six out of 12 (50%) participants reported an increase in pain upon performance of certain maneuvers (eg, items Q4, Q5, Q7, Q9, and Q14), although all were still able to complete the tasks. They were advised to seek outpatient consultation at the Rehabilitation Medicine Outpatient Clinic if the pain persisted.

Six (50%) participants found the tool easy to follow, while 5 (42%) found it difficult. Visual aids in the form of pictures and arrows were most helpful in making the tool easily

understandable. The presence of a caregiver was helpful for two of the participants.

The use of some Tagalog words (such as *pigi* or buttocks in English) was the primary barrier for understanding the tool (Table 4). Five (42%) participants were content with the prefinal version of the tool and had no suggested changes. One (8%) participant suggested that the presence of a physician or a caregiver (at the least) was still necessary to guide patients and ensure accuracy and safety in following the tool. Finally, 6 (50%) participants reiterated the need for pictures, and 1 (8%) participant suggested improvement in the portrayal of movement in the pictures.

Table . Results from the thematic analysis related to understandability of the Filipino Shoulder Telehealth Assessment Tool (STAT) and factors that make it easy or difficult to understand.

	Pretest (n=12), n (%)	Pilot (n=46), n (%)
Ease of understanding		
Tool is easy	6 (50)	13 (28)
Use of pictures and arrows	10 (83)	11 (24)
Presence of caregiver	2 (17)	3 (6)
Tool is difficult	5 (42)	11 (24)
Use of some Tagalog words	3 (25)	7 (15)
Nonfamiliarity with the tool	1 (8)	3 (6)
Difficulty reading the text	2 (17)	3 (6)
Poor internet connection	2 (17)	2 (4)
Poor audiovisual setup	1 (8)	1 (2)
Not adept with technology	1 (8)	1 (2)
Absence of caregiver	0 (0)	1 (2)
Nonideal venue	1 (8)	1 (2)
Use of arrows	1 (8)	0 (0)

Lessons From the Pretest and Development of the Final Filipino STAT

Difficulty reading the text of the tool and tendency to skip a specific question (Q3A) in a predominantly older adult population suggested inappropriate user interface design to the study team. While readers with advancing age have varying levels of possible age-related cognitive and visual decline, inclusivity was ensured in redesigning the tool [18]. These lessons were applied to the final version of the Filipino STAT (Multimedia Appendix 5) as follows: breaking down chunks of texts into bullet points, using Sans Serif typefaces with at least a 16-point font size, and providing white spaces and appropriate contrasts [18,19].

The conduct of special tests revealed an average of 73% (SD 14%) positive test findings (painful) that might have contributed to the increase in shoulder pain among pretest participants. This was addressed in the final version by placing the special test subdomain last as in clinical examination so any pain would not get in the way of the conduct of the other parts of the assessment.

Specific pretest errors could be classified as (1) misinterpretations observed in the subdomains of pain and activity (items Q2 and Q3C) and ROM (Q8); (2) differences in the semantic understanding of specific body parts, evident in ROM items Q1, Q4, Q5, Q7, and Q8; (3) difficulty understanding long instructions, which were evident in ROM item Q9 and all strength and special test items; and (4) inconsistent experiential equivalence in special test item Q17. Text qualifiers were added to misinterpreted items such as Q2, labeling 0=“normal” and 100=“hindi normal” (not normal), to avoid reversal of the scale. Items with different potential semantic meanings were significantly improved when pictorial guides were provided. Long, unclear instructions were broken down into short, step-by-step instructions with pictorial guides in each step. Finally, in item Q17 (Speed test), instead of holding a plate forward, most participants held a plate on the side, as a waiter would. The phrase was therefore changed from “*parang may hawak na plato*” (like holding a plate) to “*parang nanghihingi habang nakaunat ang siko*” (like reaching out or asking for something with an outstretched arm) to improve participants’ accuracy in performing the task and the relatability of the task in Filipino culture.

Final Filipino STAT Pilot Study

There were 47 participants in the pilot study; one of them did not complete the video consultation and withdrew due to technical difficulties (ie, unstable internet) on the patient’s end. The majority of the participants were females (35/47, 75%), aged 50 - 59 years (21/47, 45%) with a mean age of 55 (SD 14) years, married or cohabiting (29/47, 62%), and finished tertiary education (26/47, 55%; Table 3). Most were unemployed (23/47, 49%), while those employed were mostly engaged in nonhealth-related work (40/47, 85%). The monthly family income of the participants was positively skewed, with the majority (21/47, 45%) earning less than Philippine 5000 (US $<100).

For each of the items under the pain and activity subdomains, 35 (75%) to 43 (91%) participants were able to perform the tasks correctly. Meanwhile, some needed cueing from their caregiver to perform the tasks correctly, particularly for item Q2, where 11 (23%) participants needed help. With items Q3A and Q3C, 1 out of the 47 (2%) participants was not able to perform the task completely.

The ROM, strength, and special test subdomains were tested both without and with pictorial guides. For ROM-related items, a lot of participants were not able to perform the tasks correctly without a picture (Q2: chin, Q7: back pocket, Q8: lower back, and Q9: wall touching). In contrast, for those items with picture guides, there was a greater number of patients who were able to perform the tasks correctly, with little to no need of cueing from their caregivers. Four (8%) participants showed compensatory movements with some items, yet they answered yes when asked if they could do the tasks. Hence, they were considered incorrect task performances, with item Q8 being the most common item that resulted in compensatory movements.

Without pictorial guides, more than half of the participants were not able to perform all the tasks under the strength subdomain. Slightly more participants correctly performed the tasks for items Q11 and Q12 when shown pictorial guides. Moreover, the same scenario was observed for all items in the special test subdomain—having pictorial guides improved the number of participants able to perform the tasks correctly from 34 (74%) to 39 (87%) participants, while 2 (4%) to 6 (13%) participants needed cueing from their caregivers to be able to perform the tasks correctly.

Among the 46 participants who shared their experience with the final Filipino STAT, 13 (28%) found the tool easy to understand, while 11 (24%) found it difficult. The pictures and arrows were found to be most helpful for participants. Participant 47 remarked, "*maayos po binigay ang panuto at dahil sa larawan ay naintindihan po nang mabuti*" (the instructions were given properly, and because of the pictures, they were better understood). The use of some Tagalog words, however, was considered the primary barrier by the majority of the participants, followed by their nonfamiliarity with the tool and difficulty reading the text (Table 4). The same participant also remarked, "*May kahinaan po ako sa Pilipino, tulad po ng pigi po, hindi ko siya naintindihan*" (I have some difficulty in understanding Tagalog, for example, I did not understand the Tagalog word for buttock").

Content Validity and Internal Consistency Testing of the Final Filipino STAT

Content validity through scale validity indices was deemed excellent at 0.97 using the item-level content validity index and 0.80 using the universal agreement. The κ coefficients were excellent across all items, supporting that the degree of agreement among the experts was beyond chance [20]. A Cronbach α score of 0.87 indicated that all test items were unidimensional, or that performance on the items could be explained in terms of a single underlying factor (Multimedia Appendix 6).

Discussion

Principal Findings

The Filipino STAT has an excellent content validity (scale validity index=0.80 - 0.97), excellent interrater reliability (κ coefficient=0.82 - 1.00), and good internal consistency (Cronbach α=0.87). Its understandability is excellent for pain and activity (98%), good for ROM and special tests (85%), and poor for strength (37%).

Comparison With Prior Work

The findings of this study are consistent with the existing literature on shoulder teleconsultation psychometric properties. Internet assessment of rehabilitation outcomes such as pain, ROM, muscle strength, and functional assessment had good concurrent validity. There is a strong agreement between virtual and in-person examination for ROM (87.4% agreement; x^2=30.782; P<.001), and diagnosis (85.1% agreement; κ=0.82, 95% CI) [21-23]. Further studies on the Filipino STAT compared against in-person examination or imaging modalities may be done to ascertain concurrent validity of the tool.

Strengths and Limitations

Since a direct translation of a tool from an original English version may not reflect the cultural nuances and context of another culture [13], this study observed a rigorous and standard linguistic validation and cross-cultural adaptation to ensure a culturally appropriate, valid, and reliable translation. Although the original STAT has not been validated as a whole tool, the Filipino-translated tool has been subjected to content validity and internal consistency testing in this study. The collective inputs from the consensus committee of experts, the understandability questionnaire, and qualitative data from the participants ensured that the final tool can be used in the clinical setting.

There are some limitations to this study. First, the assessment of ease of understanding of the tool using the open-ended questionnaire may have been confounded by the consecutive performance of the tool without and with a pictorial guide, as well as the videoconsultation study setup. The test-taking that was twice as long, with no visual aid provided at the start, could have made the consultation more difficult for the participants. The barriers to the videoconsultation study setup were consistent with previous literature, such as unstable internet connection and poor audiovisual resolution [5]. In the envisioned clinical performance of the Filipino STAT, only the pictorial guide will be performed before the actual consultation, thereby eliminating these concerns.

Second, understandability per subdomain was excellent for pain and activity, good for ROM and special tests, and poor for strength. The incidence of compensatory or trick movement was accounted for in this study and was most frequently seen in the ROM subdomain, particularly item Q8 (50 degrees of internal rotation).

Finally, such as in an in-person physical examination, incorrect performance of a virtual test, from poor understandability or compensatory movement, makes its results invalid and

questionable. This shall serve as a reminder that the Filipino STAT is not meant to replace actual in-person consultations and should just aid the clinician in assessing the patient and make clinical work efficient. The clinician must be prudent in confirming the initial Filipino STAT findings during the actual consultation, as necessary. Further studies may consider the use of instructional videos to improve understandability of the tool, especially for the strength subdomain.

Future Directions

Both pretest and pilot study findings reveal significant improvements in the correct performance of the Filipino STAT tasks with pictorial guides, compared to none. The majority of the participants from both phases of the study also shared a positive perception of the use of pictures and arrows and the presence of a caregiver. Thus, it is the study's recommendation to use the Filipino STAT with a pictorial guide, as intended in the original tool [6]. The association of the presence of a caregiver in the successful performance of the Filipino STAT was beyond the scope of this present study. Nonetheless, having a caregiver around may help with the conduct of the Filipino STAT but should not discourage those who do not have an available caregiver.

Conclusions

The development of the Filipino STAT through a rigorous linguistic validation and cultural adaptation ensured a culturally appropriate, valid, and reliable translation. Understandability and ease of understanding by end-users are also critical to assess in patient-reported outcome measures. The pain and activity, ROM, and special test subdomains of the Filipino STAT may be used for clinical assessment, while the strength subdomain needs further improvement on understandability.

Acknowledgments

The authors would like to acknowledge the clinical experts Dorothy Dy Ching Bing-Agsaoay, Monalisa Lim-Dungca, Christopher Constantino, Nathaniel Napa, Mitchelle Gabuya, and Frances Carlos, as well as our Filipino STAT model Nancy Cheng, for their generous contributions to this study. They would also like to acknowledge the STAT original authors, Gregory Sprowls, Jaycen Brown, and Brett Robin, for allowing translation of their tool.

Funding

This research study was funded by the Philippine General Hospital—Expanded Hospital Research Office.

Data Availability

The data that support the findings of this study are uploaded online as supplementary files and may be requested from the corresponding author, JA, upon reasonable request.

Authors' Contributions

Conceptualization: JA, SI, JAM, CFL
Data curation: JA
Funding acquisition: SI
Implementation: JA
Methodology: JA, SI, JAM, CFL
Supervision: SI, JAM, CFL
Writing – original draft: JA
Writing – reviewing & editing: SI, JAM, CFL

Conflicts of Interest

None declared.

Multimedia Appendix 1
The content validation form.
[PDF File, 2546 KB - rehab_v13i1e67974_app1.pdf]

Multimedia Appendix 2
The expert consensus committee guide on equivalence.
[PDF File, 114 KB - rehab_v13i1e67974_app2.pdf]

Multimedia Appendix 3
The observational checklist.
[PDF File, 50 KB - rehab_v13i1e67974_app3.pdf]

Multimedia Appendix 4
The open-ended interview questions.
[PDF File, 48 KB - rehab_v13i1e67974_app4.pdf]

Multimedia Appendix 5
The final Filipino Shoulder Telehealth Assessment Tool (STAT).
[PDF File, 9454 KB - rehab_v13i1e67974_app5.pdf]

Multimedia Appendix 6
Supplemental dataset to a full manuscript published in the *J Med Internet Res*.
[DOCX File, 29 KB - rehab_v13i1e67974_app6.docx]

References

1. Bos WH, van Tubergen A, Vonkeman HE. Telemedicine for patients with rheumatic and musculoskeletal diseases during the COVID-19 pandemic; a positive experience in the Netherlands. Rheumatol Int 2021 Mar;41(3):565-573. [doi: 10.1007/s00296-020-04771-6] [Medline: 33449162]
2. D Leochico CF. Adoption of telerehabilitation in a developing country before and during the COVID-19 pandemic. Ann Phys Rehabil Med 2020 Nov;63(6):563-564. [doi: 10.1016/j.rehab.2020.06.001] [Medline: 32544528]
3. Leochico CFD, Mojica JAP, Rey-Matias RR, Supnet IE, Ignacio SD. Role of telerehabilitation in the rehabilitation medicine training program of a COVID-19 referral center in a developing country. Am J Phys Med Rehabil 2021 Jun;100(6):526-532. [doi: 10.1097/PHM.0000000000001755] [Medline: 33998606]
4. Leochico CFD, Montemayor GAM, Obeles AJT, Ong BAG. Telerehabilitation in a developing country toward the tail end of the COVID-19 pandemic: is it here to stay? Acta Med Philipp 2022 Nov. [doi: 10.47895/amp.vi0.5905]
5. Leochico CFD, Espiritu AI, Ignacio SD, Mojica JAP. Challenges to the emergence of telerehabilitation in a developing country: a systematic review. Front Neurol 2020;11:1007. [doi: 10.3389/fneur.2020.01007] [Medline: 33013666]
6. Sprowls GR, Brown JC, Robin BN. The Shoulder Telehealth Assessment Tool in transition to distance orthopedics. Arthrosc Tech 2020 Nov;9(11):e1673-e1681. [doi: 10.1016/j.eats.2020.07.008] [Medline: 33294325]
7. McLean JM, Awwad D, Lisle R, Besanko J, Shivakkumar D, Leith J. An international, multicenter cohort study comparing 6 shoulder clinical scores in an asymptomatic population. J Shoulder Elbow Surg 2018 Feb;27(2):306-314. [doi: 10.1016/j.jse.2017.08.016] [Medline: 29221757]
8. Reed MD, Van Nostran W. Assessing pain intensity with the visual analog scale: a plea for uniformity. J Clin Pharmacol 2014 Mar;54(3):241-244. [doi: 10.1002/jcph.250] [Medline: 24374753]
9. Retzky JS, Baker M, Hannan CV, Srikumaran U. Single Assessment Numeric Evaluation scores correlate positively with American Shoulder and Elbow Surgeons scores postoperatively in patients undergoing rotator cuff repair. J Shoulder Elbow Surg 2020 Jan;29(1):146-149. [doi: 10.1016/j.jse.2019.05.039] [Medline: 31401127]
10. Oosterwijk AM, Nieuwenhuis MK, van der Schans CP, Mouton LJ. Shoulder and elbow range of motion for the performance of activities of daily living: a systematic review. Physiother Theory Pract 2018 Jul;34(7):505-528. [doi: 10.1080/09593985.2017.1422206]
11. Capio CM, Maranion DZ, Pascua HE. Musculoskeletal pain and dysfunction among University of the Philippines Manila personnel. Univ Philippines Manila J 2001;6(4):25-23 [FREE Full text]
12. Labao HC, Faller EM, Bacayo MFD. “Aches and pains” of Filipino migrant workers in Malaysia: a profile of work-related musculoskeletal disorders. Ann Glob Health 2018 Aug;84(3):474-480. [doi: 10.29024/aogh.2331] [Medline: 30835403]
13. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1993 Dec;46(12):1417-1432. [doi: 10.1016/0895-4356(93)90142-n] [Medline: 8263569]
14. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharm Stat 2005 Oct;4(4):287-291. [doi: 10.1002/pst.185]
15. Bonett DG. Sample size requirements for testing and estimating coefficient alpha. J Educ Behav Stat 2002 Dec;27(4):335-340. [doi: 10.3102/10769986027004335]
16. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. SPINE (Phila Pa 1976) 2000 Dec;25(24):3186-3191. [doi: 10.1097/00007632-200012150-00014] [Medline: 11124735]
17. Lorca LA, Torres-Castro R, Ribeiro IL, et al. Linguistic validation and cross-cultural adaptation of the post–COVID-19 functional status scale for the Chilean population. Am J Phys Med Rehabil 2021 Apr 1;100(4):313-320. [doi: 10.1097/PHM.0000000000001706] [Medline: 33496442]
18. Office of Disease Prevention and Health Promotion. Section 5.3 use a readable font that’s at least 16 pixels. Health Literacy Online, 3rd Edition. 2025. URL: https://odphp.health.gov/healthliteracyonline/design-easy-scanning/use-readable-font-thats-least-16-pixels [accessed 2025-12-01]

19. Hou G, Anicetus U, He J. How to design font size for older adults: a systematic literature review with a mobile device. Front Psychol 2022;13:931646. [doi: 10.3389/fpsyg.2022.931646] [Medline: 35978796]
20. Rodrigues IB, Adachi JD, Beattie KA, MacDermid JC. Development and validation of a new tool to measure the facilitators, barriers and preferences to exercise in people with osteoporosis. BMC Musculoskelet Disord 2017 Dec;18(1):540. [doi: 10.1186/s12891-017-1914-5] [Medline: 29258503]
21. Steele L, Lade H, McKenzie S, Russell TG. Assessment and diagnosis of musculoskeletal shoulder disorders over the internet. Int J Telemed Appl 2012;2012:945745. [doi: 10.1155/2012/945745] [Medline: 23193395]
22. Rabin A, Dolkart O, Kazum E, et al. Shoulder assessment by smartphone: a valid alternative for times of social distancing. Arch Orthop Trauma Surg 2022 Jun;142(6):979-985. [doi: 10.1007/s00402-021-03762-x] [Medline: 33439302]
23. Mani S, Sharma S, Omar B, Paungmali A, Joseph L. Validity and reliability of internet-based physiotherapy assessment for musculoskeletal disorders: a systematic review. J Telemed Telecare 2017 Apr;23(3):379-391. [doi: 10.1177/1357633X16642369] [Medline: 27036879]

Abbreviations

ROM: range of motion
SANE: single assessment numeric evaluation
STAT: Shoulder Telehealth Assessment Tool
VAS: visual analog scale

Edited by N Cahill; submitted 25.Oct.2024; peer-reviewed by T Lienou, V Prieto; revised version received 04.Nov.2025; accepted 11.Nov.2025; published 20.Jan.2026.

Please cite as:
Arboleda J, Ignacio S, Mojica JA, Leochico CF
Linguistic Validation and Cross-Cultural Adaptation of the Shoulder Telehealth Assessment Tool for Filipino Patients with Musculoskeletal Shoulder Condition: Cross-Sectional Study
JMIR Rehabil Assist Technol 2026;13:e67974
URL: https://rehab.jmir.org/2026/1/e67974
doi:10.2196/67974

Telerehabilitation Trends in Australian Physiotherapy and an Exploration of Factors That Influence Use After COVID-19 Restrictions: Qualitative Content Analysis

Megan H Ross[1], PhD; Joshua Simmich[1], PhD; Belinda J Lawford[2], PhD; Kim L Bennell[2], PhD; Rana S Hinman[2], PhD; Trevor Russell[1], PhD

[1]RECOVER Injury Research Centre, Faculty of Health, Medicine and Behavioural Sciences, The University of Queensland, Surgical, Treatment and Rehabilitation Service (STARS), 296 Herston Rd, Herston, Queensland, Australia

[2]Centre for Health, Exercise and Sports Medicine, Department of Physiotherapy, School of Health Sciences, Faculty of Medicine, Dentistry & Health Sciences, The University of Melbourne, Melbourne, Australia

Corresponding Author:
Megan H Ross, PhD
RECOVER Injury Research Centre, Faculty of Health, Medicine and Behavioural Sciences, The University of Queensland, Surgical, Treatment and Rehabilitation Service (STARS), 296 Herston Rd, Herston, Queensland, Australia

Abstract

Background: Telerehabilitation is a safe and effective means of delivering physiotherapy services, but implementation in clinical practice has not been widespread.

Objective: This study aimed to explore the shifts in telerehabilitation use throughout the COVID-19 pandemic and the key factors that influenced telerehabilitation caseload after restrictions were eased.

Methods: Between September and November 2023, physiotherapists practicing in Australian private practice, hospital outpatient, or community settings completed an online survey. Data were collected regarding participants' use of telerehabilitation before, during, and after the COVID-19 pandemic restrictions to in-person physiotherapy. Qualitative content analysis of open-text questions was performed to garner more nuanced information about the use of telerehabilitation in clinical practice, and quantitative data were analyzed descriptively.

Results: The proportion of participants using telerehabilitation rose from 30% (44/148) before the pandemic to 94% (138/147) when restrictions to in-person physiotherapy were in place. Although 82% (118/144) of the sample continued to deliver telerehabilitation after COVID-19 restrictions were eased, telerehabilitation accounted for only 14% of the total caseload. Exploratory analyses suggest that despite increased confidence, satisfaction, and perceptions about the effectiveness of telerehabilitation, reduced patient demand, physiotherapists' perceptions about patient preference for in-person consultations, and the perception that in-person physiotherapy is easier continue to influence the use of telerehabilitation in the post-COVID era.

Conclusions: Despite increased uptake during the pandemic, telerehabilitation caseload after restrictions were eased was low. Physiotherapists' perceptions about telerehabilitation in clinical practice remain a substantial barrier to sustained adoption.

(JMIR Rehabil Assist Technol 2026;13:e81008) doi:10.2196/81008

KEYWORDS

telehealth; COVID-19; videoconferencing; survey; physiotherapy; qualitative

Introduction

Background

Evidence suggests that telerehabilitation is a safe, feasible, and effective means of delivering physiotherapy care that is at least as good as in-person physiotherapy in terms of patient outcomes [1,2]. Despite two decades of evidence supporting the effectiveness of telerehabilitation in the management of musculoskeletal [3], neurological [4], cardiorespiratory [5], and postsurgical rehabilitation [6], clinician acceptance and adoption have been low [7]. While telerehabilitation is a viable alternative to traditional in-person physiotherapy with the potential to overcome geographical barriers, improve access, and facilitate continuity of treatment, integration into routine physiotherapy practice before the COVID-19 pandemic remained limited [8].

Before the COVID-19 pandemic, physiotherapists were slow and reluctant to adopt telerehabilitation as a standard model of care [9,10]. Several barriers contributed to this limited uptake of telerehabilitation in physiotherapy practice, including limited acceptance of and low confidence in using telehealth technology,

perceived limitations in conducting physical assessments remotely, and reduced capacity to deliver hands-on interventions that are central to traditional physiotherapy practice [8,11,12]. Additional challenges included perceptions that telerehabilitation was less effective for certain clinical presentations and concerns about developing rapport and patient engagement [11,12].

The COVID-19 pandemic brought unprecedented challenges to health care systems worldwide, including the delivery of physiotherapy services. In response to the pandemic-related restrictions on in-person consultations, many physiotherapy practices turned to telerehabilitation [13] as an alternative method to continue providing necessary care [14]. Although physiotherapy was recognized as an essential health care service in Australia during the COVID-19 pandemic, practice was still notably restricted. Stringent infection control measures, such as mandatory use of personal protective equipment, rigorous patient screening, physical distancing requirements, density limits within clinical spaces, and group size limitations on group therapy sessions [15], impacted the delivery of care across multiple clinical settings. During initial lockdowns, some states, such as Victoria, further restricted in-person physiotherapy services, permitting face-to-face consultations only for urgent cases [16]. Community and aged care physiotherapy faced further barriers, including restrictions on therapists attending multiple sites and outright bans on external providers entering residential facilities [17]. Consequently, physiotherapy practice during the pandemic was markedly disrupted, forcing providers to rapidly transition to providing telerehabilitation services to adhere to public health guidelines and ensure continuity of care.

With the rapid transition to telerehabilitation in response to the pandemic came changes in regulatory frameworks to fund telerehabilitation [18], position statements advocating for the use of telerehabilitation [19], and increased infrastructure and clinical training to support the integration of telerehabilitation into clinical care [20]. During this period, uptake of telerehabilitation increased substantially, reflecting the necessity to maintain continuity of care. Research conducted at the time suggested that physiotherapists intended to continue offering services via telerehabilitation after the easing of restrictions to in-person physiotherapy [20,21]. However, international evidence suggests that uptake and usage have generally decreased from the pandemic peak [22].

Objectives

With the rapid transition to telerehabilitation in response to the pandemic came changes in regulatory frameworks to fund telerehabilitation [18], position statements advocating for the use of telerehabilitation [19], and increased infrastructure and clinical training to support the integration of telerehabilitation into clinical care [20]. During this period, uptake of telerehabilitation increased substantially, reflecting the necessity to maintain continuity of care. Research conducted at the time suggested that physiotherapists intended to continue offering services via telerehabilitation after the easing of restrictions to in-person physiotherapy [20,21]. However, international evidence suggests that uptake and usage have generally decreased from the pandemic peak [22].

The aim of this study was to investigate the use of telerehabilitation in Australian physiotherapy clinical practice throughout the COVID-19 pandemic, with a focus on telerehabilitation use after restrictions were eased.

The specific research questions for this study were as follows: (1) How did the use of telerehabilitation vary in physiotherapy clinical practice in Australia before, during, and after COVID-19 restrictions to in-person consultations? and (2) What are the key factors that influence physiotherapists' telerehabilitation caseload in the postrestrictions period?

Methods

Design

A descriptive, cross-sectional survey was conducted online with physiotherapists currently practicing in Australia. The study was primarily quantitative, with a small qualitative component to supplement descriptive analyses.

Ethical Considerations

The study was approved by The University of Queensland Human Research Ethics Committee (approval number: 2023/HE001802) and reported following the consensus-based CROSS (Checklist for Reporting of Survey Studies) [23]. Participants provided electronic informed consent after reviewing an information sheet and before completing the survey. Participants were entered into a draw for a AUD $1000 (US $667) gift voucher upon completion of the survey. Participants' privacy and confidentiality were maintained by storing nonidentifiable survey data separately from contact details on the University of Queensland Research Data Management System.

Participants

Participants were physiotherapists recruited from the community via online advertisements on social media (eg, Facebook, X, and LinkedIn), via targeted emails, and through Australian Physiotherapy Association member communications (eg, eComms). Physiotherapists were eligible to participate if they were registered with the Australian Health Practitioner Regulation Agency and currently practicing in an Australian private practice, hospital, or community setting. Participants who had not delivered telerehabilitation services were eligible to complete a short version of the questionnaire to explore reasons for not engaging with telerehabilitation and the circumstances that might influence uptake.

Procedure

An online survey was designed to capture information that was relevant to stakeholders and ensure readability and credibility (Multimedia Appendix 1). The survey was developed by the authors using Bennell [20] as a guide and adapted to capture information relevant to the different phases of the COVID-19 pandemic restrictions. The 3 phases were "Prior to the pandemic restrictions," "During the period of restrictions to in-person physiotherapy" (from the introduction of restrictions in 2020 to 2022), and "After restrictions were eased" (2022 onward). Questions were primarily multiple choice (checkbox questions), numerical rating scales (0 - 10), and 5-point Likert scales.

Respondents were asked to estimate their telerehabilitation caseload (individual video, group video, and telephone) for each phase using a sliding scale (0% - 100%). Free-text responses were sought for some questions to ascertain more nuanced and in-depth information about physiotherapists' perceptions of using telerehabilitation in clinical practice.

The survey was administered via an online secure platform (Qualtrics, LLC) and hosted by The University of Queensland. Participants were first invited to complete the online consent form and screening and, if eligible, proceeded to the survey. Participants were asked to provide demographic information, details of clinical practice, and experience with telerehabilitation. The second primary section of the survey comprised questions pertaining to the use of telerehabilitation during each phase of the COVID-19 pandemic restrictions (before, during, and after). All data were collected between September 15 and November 8, 2023.

Data Analysis

Data were exported from the online platform for analysis in R (version 4.3.3; R Core Team). Descriptive statistics, including frequencies (percentages) and means and standard deviations, were used to summarize the data. All responses (including partial responses) meeting eligibility criteria were included in analyses. When an "other" field was provided for additional response options, 2 researchers reviewed free-text responses and either aligned them with existing response options or designated them as unique responses that were added to the final list of response options. Any discrepancies in coding were resolved via discussion.

Responses to free-text questions were analyzed qualitatively using inductive content analysis in Microsoft Excel [24]. First, 2 researchers (MHR and JS) independently read the entire dataset, conducted open coding, and identified topics and initial patterns. The unit of analysis was meaning units, identified within individual responses. Codes were subsequently categorized and combined to form main categories or themes (abstraction), with both authors returning to the dataset to check that codes made sense in relation to the raw data. The 2 authors then met to compare and discuss their coding frameworks, and discrepancies were resolved through discussion. An audit trail was maintained to document coding decisions and category development. Themes with the highest number of individual data points were identified, reported, and described. To enhance trustworthiness, reflexivity was considered throughout the process, and attention was paid to credibility and transparency in coding and interpretation.

To explore which factors influenced physiotherapists' use of telerehabilitation in the postpandemic restrictions period, the total proportion of videoconferencing telerehabilitation caseload (individual and group consultations) was examined. Specifically, this proportion was plotted against the following five key postpandemic variables: confidence, satisfaction, and perceived effectiveness of telerehabilitation; physiotherapists' perception about how much patients like telerehabilitation; and how often patients are requesting it. Locally estimated scatterplot smoothing curves were fitted using the full span of the data (span=1). These smoothed trends, along with their corresponding 95% CIs, were used to visually explore apparent associations. No statistical correlation or regression analyses were performed on these trends.

Results

Sample Characteristics

A total of 222 physiotherapists responded to the survey, with 152 (68%) meeting eligibility criteria and providing sufficient data to be included in analyses (58/222, 26%, excluded for not being an Australian Health Practitioner Regulation Agency–registered physiotherapist currently practicing in an eligible setting [eg, private practice, hospital outpatient, or community] and 12/222, 5% not providing sufficient data to determine eligibility). Most participants (107/152, 70%) completed the survey in less than 20 minutes.

Respondents were primarily women (87/152, 57%); working in musculoskeletal (105/152, 69%) private practice (84/152, 55%) in Queensland (42/152, 28%), Victoria (42/152, 28%), or New South Wales (38/152, 25%); and held either a Bachelor's (70/152, 46%) or Master's (58/152, 38%) degree in physiotherapy. Physiotherapists primarily used Zoom (66/142, 47%), a telephone (59/142, 42%) or Microsoft Teams (47/142, 33%) to conduct telerehabilitation consultations. Only 40% (n=56) of respondents indicated that they had participated in telerehabilitation training. Additional participant characteristics are provided in Table 1.

Table . Participant characteristics (total N=152 unless otherwise specified).

Characteristic	Values, n (%)[a]
Gender	
Woman	87 (57)
Man	63 (41)
Prefer not to say	2 (1)
State or territory	
Queensland	42 (28)
Victoria	42 (28)
New South Wales	38 (25)
Western Australia	13 (9)
South Australia	10 (7)
Australian Capital Territory	6 (4)
Tasmania	1 (1)
Area of practice	
Private practice (primary care)	83 (55)
Public health outpatient center	45 (30)
Community health center	29 (19)
Private hospital	10 (7)
Other	20 (13)
Clinical focuses	
Musculoskeletal or orthopedic	105 (69)
Sports and exercise	43 (28)
Neurology	30 (20)
Gerontology	24 (16)
Pediatric	12 (8)
Other	35 (23)
Highest education	
Bachelor’s degree	70 (46)
Master’s by coursework	58 (38)
Masters by research	3 (2)
Postgraduate diploma	11 (7)
PhD	6 (4)
Other	4 (3)
Prior training in telehealth	
No	96 (63)
Yes, <6 mo ago	5 (3)
Yes, between 6 and 12 mo ago	6 (4)
Yes, between 12 mo and 2 y ago	17 (11)
Yes, between 2 and 3 y ago	17 (11)
Yes, longer than 3 y ago	11 (7)
Telerehabilitation software (recently used; n=142)	
Zoom	66 (47)
Telephone	59 (42)

Characteristic	Values, n (%)[a]
Microsoft Teams	47 (33)
Physitrack	30 (21)
Other	89 (63)

[a]Percentages may not sum to 100% because respondents could select multiple options.

Shifts in Telerehabilitation Use Through the Phases of the COVID-19 Pandemic

Thirty percent (44/148) of respondents indicated that they were using telerehabilitation in clinical practice before the pandemic. This rose to 94% (138/147) during the period of COVID-19 restrictions and reduced to 82% (118/144) after restrictions were lifted. Only 3% (4/152) of the sample indicated that they had never provided telerehabilitation consultations (individual or group videoconferencing, or telephone consultations). Total telerehabilitation caseload rose to account for almost 47% of the total caseload during the period of restrictions but dropped substantially to 14% once restrictions were lifted, but still remained above the prepandemic level of 4% (Figure 1). This pattern was fairly consistent across areas of practice over the COVID-19 pandemic (Figure S1 in Multimedia Appendix 2).

Reasons for not providing telerehabilitation consultations during each phase of the pandemic are provided in Figure 2. Across all 3 phases, the primary reasons were the perception that patients prefer in-person consultations (83/139, 60%) and that it was easier to do in-person consultations (55/139, 55%; Figure 2).

Figure 1. Shift in estimated telerehabilitation caseload before, during, and after the COVID-19 pandemic restrictions (values <2% are plotted without labels).

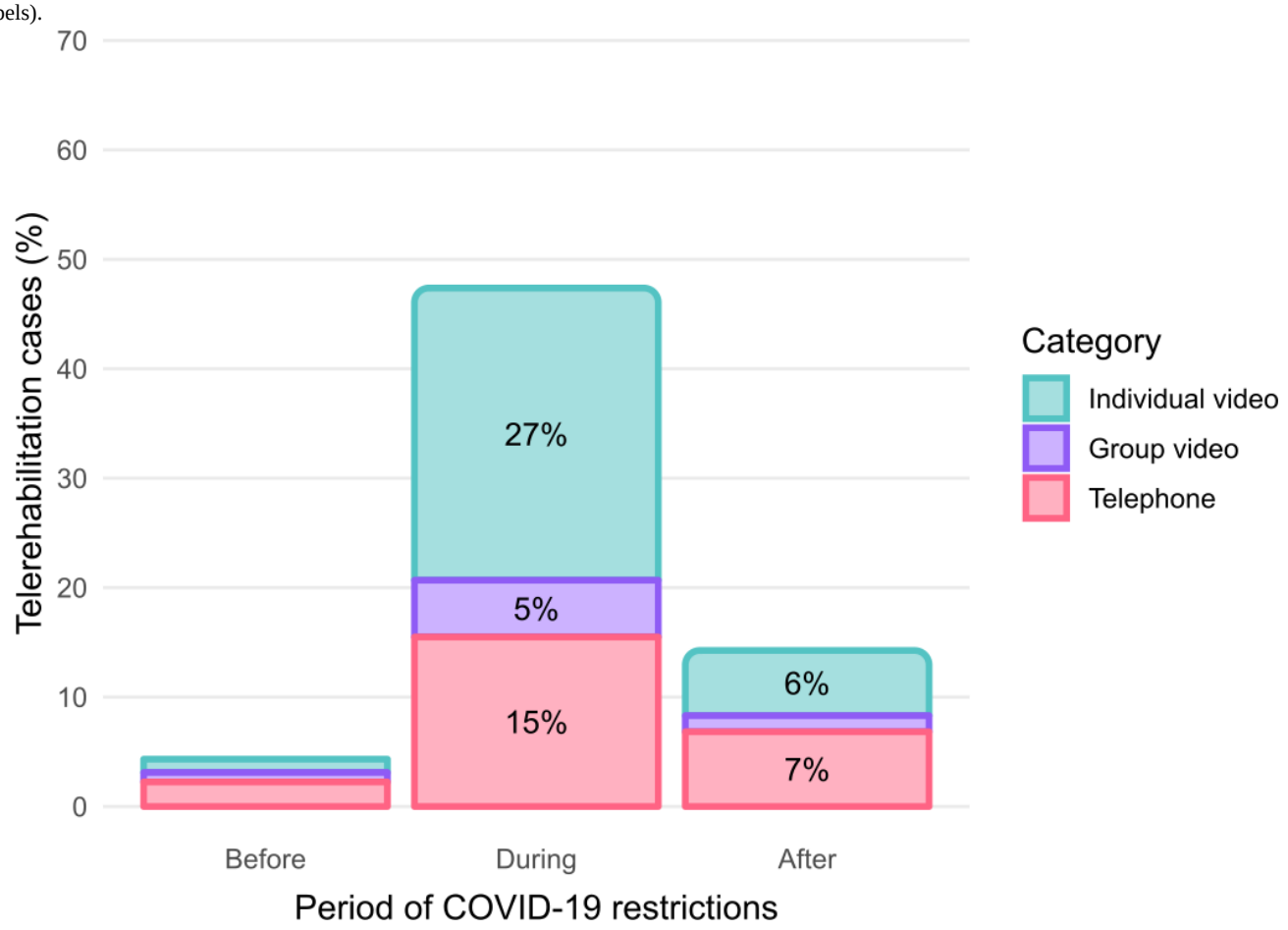
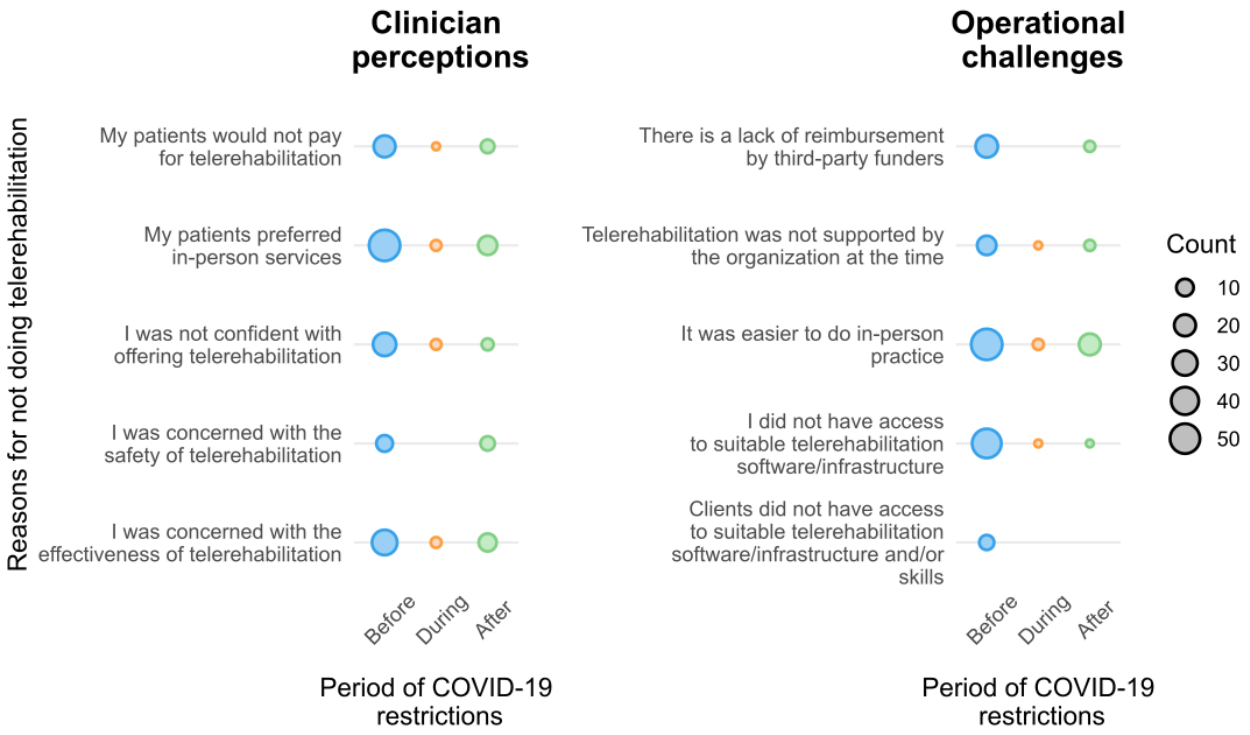


Figure 2. Reasons for not providing telerehabilitation throughout the phases of the COVID-19 pandemic, grouped by clinician perceptions and operational challenges. The area of circles represents the total count of respondents listing that reason for that point in time.

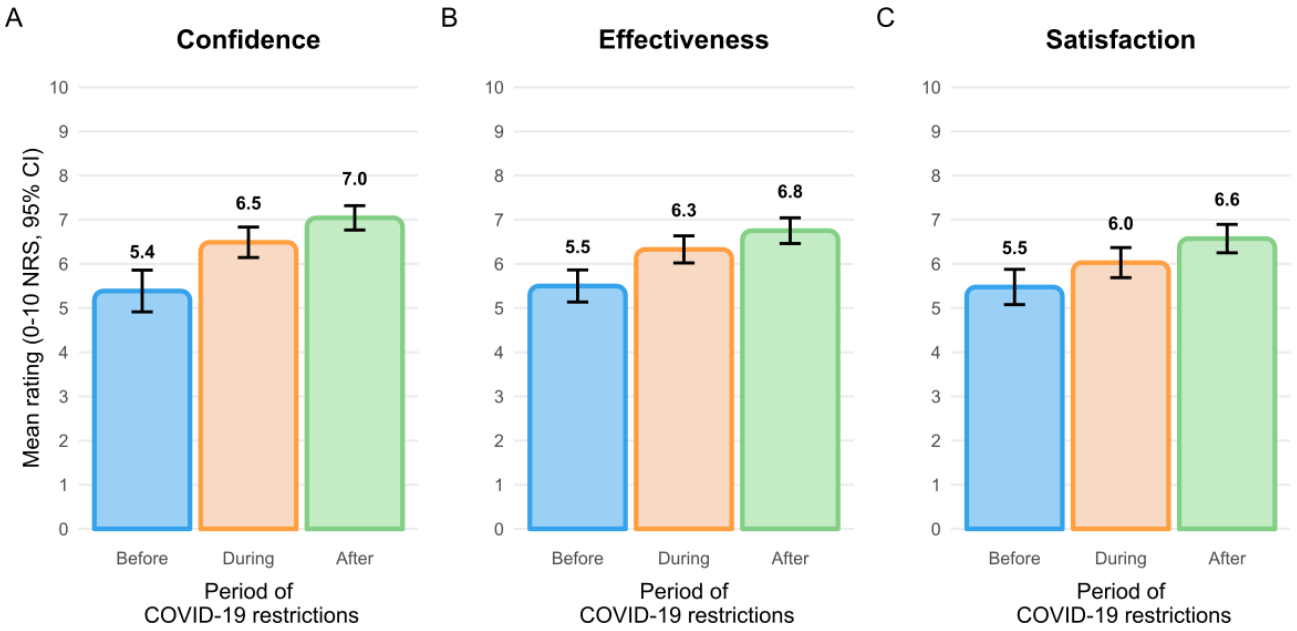


Before the pandemic, additional reasons for not offering telerehabilitation were primarily the perception that there was no need for telerehabilitation (57/104, 55%) or that physiotherapists did not have access to suitable telerehabilitation software or infrastructure (47/104, 45%; Table S1 in Multimedia Appendix 3). After restrictions were eased, the primary reasons for not providing telerehabilitation services were that respondents were concerned about the effectiveness of telerehabilitation (13/26, 50%) and did not like providing care via telerehabilitation (11/26, 42%; Table S1 in Multimedia Appendix 3).

Shifts in Confidence, Effectiveness, and Satisfaction With Telerehabilitation

Physiotherapist ratings of confidence in providing care via telerehabilitation, perceived effectiveness of telerehabilitation, and satisfaction with telerehabilitation progressively increased from before, during, to after restrictions associated with the COVID-19 pandemic (Figure 3). Almost 85% (120/142) of respondents indicated that providing telerehabilitation had become easier over time.

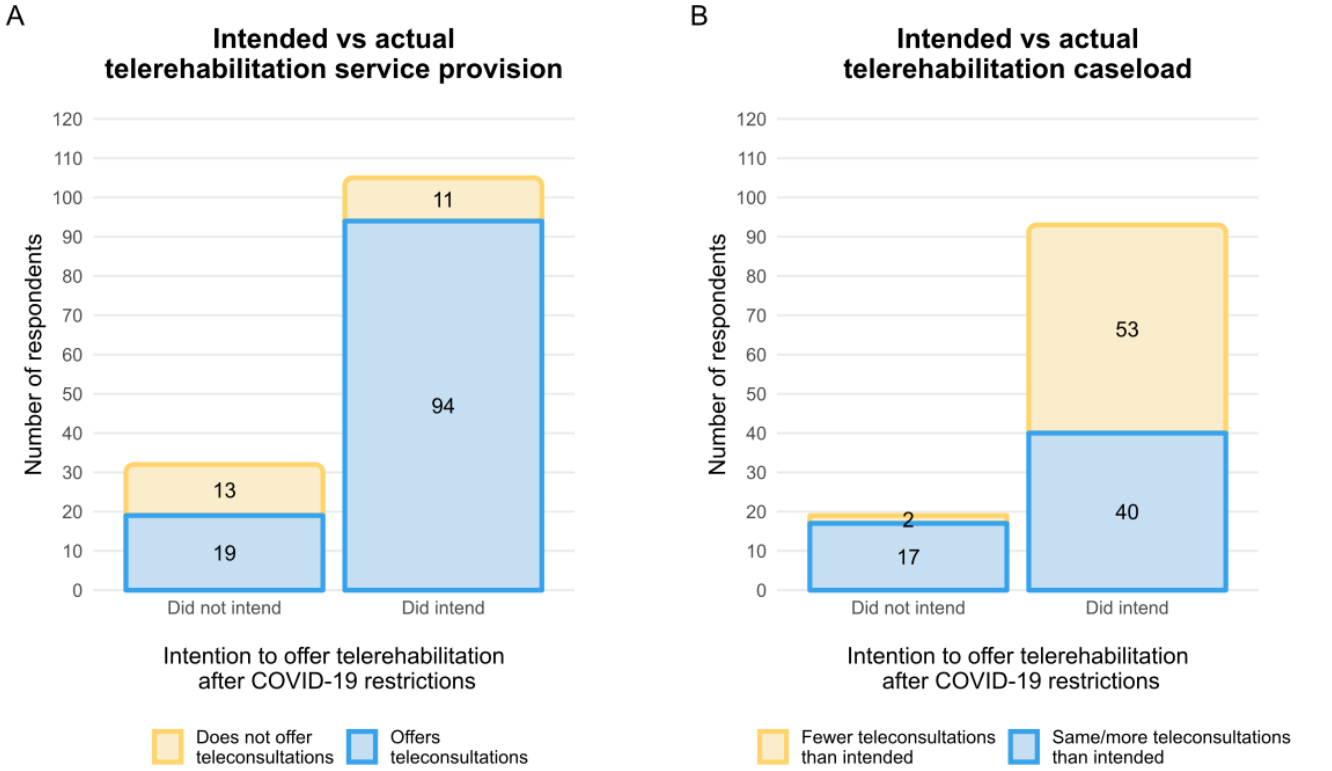
Figure 3. Participant ratings of (A) confidence, (B) perceived effectiveness, and (C) satisfaction with telerehabilitation across the pandemic. NRS: numerical rating scale.



Intended Versus Actual Telerehabilitation Use

Most respondents (105/137, 77%) intended to offer telerehabilitation after the easing of COVID-19 restrictions (Figure 4). Of these, only 10% (11/105) did not offer telerehabilitation despite intending to do so (Figure 4A). Primary reasons for not intending to offer telerehabilitation were because it was "easier to do in-person" (23/32, 72%) and because "patients prefer in-person" (22/32, 69%; Table S2 in Multimedia Appendix 3). Of those who did not intend to, more than half (19/32, 59%) did continue to offer telerehabilitation after restrictions were eased (Figure 4A). Approximately 50% of respondents who intended to continue offering telerehabilitation consultations (53/93) or who actually continued offering them (58/118) after restrictions were eased were providing fewer consultations than initially intended (Figure 4B). Respondents indicated that this was because patients "prefer in-person services" (44/58, 76%), "patient demand reduced more than expected" (32/38, 55%) and because it was "easier to do in-person consultations" (27/38, 47%; Table S3 in Multimedia Appendix 3). Primary reasons for continuing to offer telerehabilitation services included that telerehabilitation allowed physiotherapists to offer services to patients who would not usually be able to attend their clinic (84/118, 71%), that patients like the option of receiving care via telerehabilitation (76/118, 64%), and that patients find telerehabilitation convenient (74/118, 63%; Table S4 in Multimedia Appendix 3).

Figure 4. Intended versus actual telerehabilitation service provision and caseload. (A) Number of respondents who intended to offer telerehabilitation after easing of pandemic restrictions compared to whether they do offer telerehabilitation now. (B) Number of respondents who intended to offer telerehabilitation after easing of pandemic restrictions compared to whether the frequency of telerehabilitation met intentions.



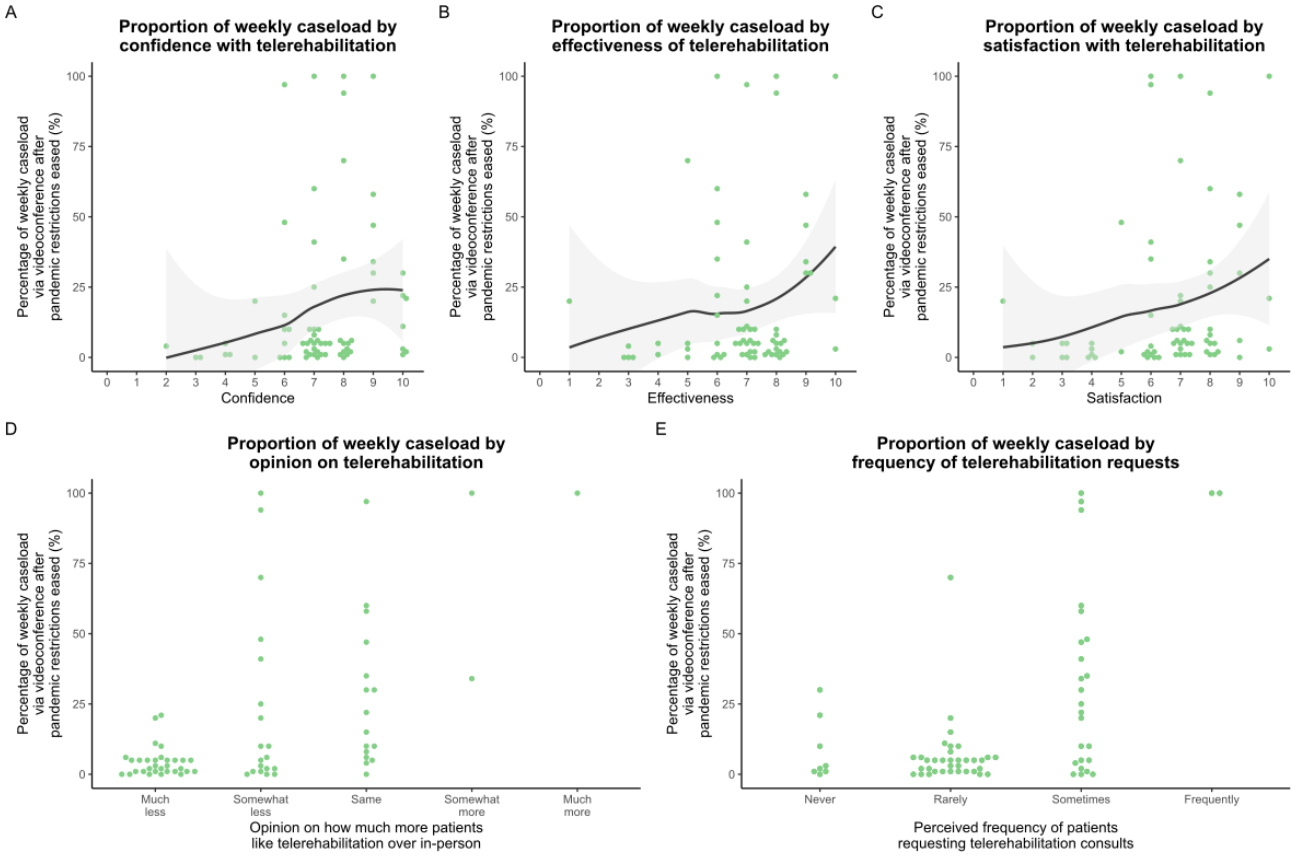
Factors That Influence Telerehabilitation Use Postpandemic Restrictions

Positive correlations were noted between a higher proportion of weekly caseload conducted via telerehabilitation and higher ratings of confidence in using telerehabilitation (Figure 5A), perceived effectiveness of telerehabilitation (Figure 5B), and satisfaction with telerehabilitation (Figure 5C).

Almost half of the respondents (69/142, 49%) indicated that patients were “rarely” requesting telerehabilitation since the easing of restrictions, and in the opinion of approximately half of the respondents (70/142, 49%), patients like telerehabilitation “much less than in-person consultations.” Physiotherapists who believed that patients liked telerehabilitation much less than in-person consultations appeared to have a lower proportion of their weekly caseload conducted via telerehabilitation (Figure 5D). Similarly, physiotherapists who reported that their patients requested telerehabilitation at least sometimes seemed more likely to have a higher proportion of weekly cases conducted via telehealth (Figure 5E).

The median (IQR) percentage of patients considered unsuitable for telerehabilitation by the respondents was 50.5% (50). Patient complexity and conditions requiring hands-on treatment were the primary reasons that patients were “often” considered unsuitable (Figure S2 in Multimedia Appendix 4 and Table S5 in Multimedia Appendix 3). Additional reasons respondents provided for deeming patients unsuitable are provided in Table S6 in Multimedia Appendix 3, with the largest proportion being patient preference for in-person consultations (6/30, 20%), physical examination being indicated (4/30, 13%), and complex patient presentations (6/30, 14%).

Figure 5. noTotal weekly telerehabilitation caseload versus (A) confidence, (B) effectiveness, (C) satisfaction, (D) patient liking for telerehabilitation, and (E) patient requests for telerehabilitation. Dark gray lines represent locally estimated scatterplot smoothing fits (using the whole span of the data) with 95% CIs (light gray shading).



Telerehabilitation Clinical Practice Considerations

Although only 5% (8/142) of respondents reported never experiencing technical issues themselves, most (120/142, 84%) indicated encountering these issues rarely or sometimes, and just 9% (14/142) experienced them often. Likewise, only one respondent (1/142, 1%) reported that their patients had never encountered technical issues, whereas the majority (112/142, 79%) reported that patients experienced technical issues rarely or sometimes, and 19% (29/142) reported that patients often experienced technical issues. When technical issues were encountered, 74% (105/142) reported only moderate or less disruption to the consultation. Only 5% (7/142) reported that technical issues were extremely disruptive, and just 4% (6/142) reported often having to cancel or reschedule appointments due to technical issues (Figure S3 in Multimedia Appendix 5). To support the delivery of telerehabilitation consultations, physiotherapists used text message reminders (109/142, 77%); written or digital educational material about the condition (67/142, 47%); and written instructions, diagrams, or booklets (63/142, 44%; Table S7 in Multimedia Appendix 3).

Almost three-quarters (104/142, 73%) of respondents indicated that they used similar parameters of care for telerehabilitation as for in-person consultations (eg, similar consultation frequency, duration, and similar content). For the respondents indicating that parameters of care were different (38/142, 27%), the primary reasons were that physical assessment or treatment was limited via telerehabilitation (15/38, 39%), consultations were shorter (10/38, 26%), and consultations were more focused on exercise or education (8/38, 21%). Additional reasons are provided in Table S8 in Multimedia Appendix 3. Similarly, most respondents (128/142, 90%) indicated that telerehabilitation consultations were about the same duration or shorter than in-person consultations (Table S9 in Multimedia Appendix 3) and that consultation frequency was "about the same" as in person (72/142, 51%) or less often than in person (53/142, 37%; Table S9 in Multimedia Appendix 3).

Considering the cost of telerehabilitation consultations, almost three-quarters of physiotherapists indicated that they charged "about the same" as an in-person consultation (104/142, 73%), with very few respondents (5/142, 4%) charging more than in-person consultations. Responses about the cost to the business of providing telerehabilitation were similar, with 50% (71/142) of respondents considering telerehabilitation to cost about the same and 36% (51/142) indicating that telerehabilitation consultations cost the business less than in-person consultations (Table S9 in Multimedia Appendix 3).

The median proportion of patients offered hybrid care in a current weekly caseload was 5% (minimum=0, Q1=1, Q3=20, maximum=100). In hybrid models of care, 42% (48/113) of physiotherapists indicated that patients typically receive many more in-person than telerehabilitation consultations, 27% (30/113) receive the same, and 18% (21/113) receive fewer in-person visits compared to telerehabilitation consultations (Table S9 in Multimedia Appendix 3). Other ways in which telerehabilitation models of care differ from in-person models were coded qualitatively and provided in Table S10 in

Multimedia Appendix 3. When respondents used a hybrid model, 27% (8/27) offered telerehabilitation only after an initial in-person consult, 15% (4/27) described limiting the physical assessment or treatment component of consultations, and 11% (3/27) said telerehabilitation consultations in hybrid models had a greater case management focus (3/27, 11%).

Inductive content analysis of free-text responses identified four key themes that reflected respondents' perspectives on using telerehabilitation in clinical practice postpandemic (Table S11 in Multimedia Appendix 3): (1) concerns about telerehabilitation, (2) perceived benefits of telerehabilitation, (3) how telerehabilitation is used in practice, and (4) physiotherapists' willingness to provide telerehabilitation services.

Theme 1: Concerns About Telerehabilitation (n=28)

Physiotherapists expressed a range of concerns about the suitability and practicality of telerehabilitation in postpandemic physiotherapy care. The most commonly reported issue was that clients prefer or actively seek in-person consultations (n=12, 43%). For example, one participant said that despite telerehabilitation remaining available for their patients, they "*often prefer face-to-face*" and that "*people wanted to revert back to the 'usual' ways and leave the changes of COVID behind them moving forward once restrictions eased*" (musculoskeletal physiotherapist). Some respondents (n=4, 14%) emphasized that telerehabilitation is not suitable for all clients, particularly those with complex conditions, communication difficulties, or low digital literacy, and that they were selecting suitable clients for telerehabilitation, and "*not offering [it] for those not 'tech-savvy'*" (neurological physiotherapist). Concerns were also raised about the limitations of assessment via videoconference (n=2, 7%) and challenges related to internet connectivity and software reliability (n=3, 11%). Additional issues included payment and reimbursement barriers (n=2, 7%), difficulties building rapport remotely (n=2, 7%), and reduced referrals and attendance for telerehabilitation compared to in-person care.

Theme 2: Perceived Benefits of Telerehabilitation (n=20)

Despite concerns, respondents acknowledged several advantages of using telerehabilitation in their clinical practice postpandemic restrictions. The most frequently cited benefit was that telerehabilitation improved patient access to care, particularly for those in rural or remote areas or those with difficulties with travel or limited time (n=8, 40%). One musculoskeletal physiotherapist said that telerehabilitation "*has made physiotherapy much more accessible to a wider population and allows people greater flexibility with appointments*." Participants also noted an increased acceptance of telerehabilitation (among patients and providers; n=6, 30%), with some suggesting that it "*has become common practice now*" (musculoskeletal physiotherapist) and that it can be effective for certain presentations (eg, chronic musculoskeletal conditions; n=3, 15%), for supporting patient self-management (n=2, 10%) and providing greater flexibility in service delivery (n=1, 5%).

Theme 3: How Telerehabilitation Is Used in Practice (n=8)

Participants described integrating telerehabilitation into their clinical practice for subsequent consultations following initial in-person visits (n=2, 25%), for triaging (n=1, 12.5%) and case management (n=1, 12.5%), and as a tool for exercise prescription (n=1, 12.5%). Some participants (n=2, 25%) indicated that videoconferencing was preferred over telephone, and 1 (12.5%) participant noted that at times additional support is required at the patient end to effectively deliver telerehabilitation services. One cardiorespiratory, hospital-based physiotherapist described that telerehabilitation "*consultations have been effective in triaging patients and determining the appropriate level of care required*," whereas another described that telerehabilitation "*has been a great option for follow-up appointments, especially when you have already build rapport with patients...[and it]...has been a great way to check in with people who have busy schedules or live far away and find it difficult coming in*" (pelvic health and musculoskeletal physiotherapist).

Theme 4: Physiotherapists' Willingness to Provide Telerehabilitation Services (n=23)

Many participants (n=16, 70%) were willing to continue providing telerehabilitation services, driven by the perceived benefits and uses of telerehabilitation. For example, one private practice, musculoskeletal and mental health physiotherapist said that "*for the provision of exercise and movement based interventions, telerehabilitation has worked better than in-person as it provides easier access to more people given I live in a regional area*." Despite this willingness, some participants expressed low satisfaction with telerehabilitation (n=2, 9%) or a preference for in-person consultations (n=2, 9%). For example, a sports, exercise, and musculoskeletal physiotherapist working in private practice said that despite telerehabilitation "*opening up my practice to lots of different people around Australia and internationally... I still prefer to consult in-person...*" The need for ongoing education about the utility of telerehabilitation in physiotherapy was noted (n=1, 4%), despite the perception that education about how to deliver telerehabilitation had improved during the pandemic (n=1, 4%).

Discussion

Principal Findings

Despite an initial increase due to the COVID-19 pandemic restrictions and physiotherapists' intentions to continue offering telerehabilitation services, many physiotherapists were offering fewer telerehabilitation consultations than anticipated once restrictions were lifted. This was primarily due to a preference for in-person consultations, concerns about the effectiveness of telerehabilitation, and the perception that physiotherapy consultations are easier to conduct in person.

International postpandemic data across both physiotherapy and other health services show a similar "peak-to-plateau" pattern, where telehealth usage increased substantially during restrictions before falling and stabilizing at a lower level rather than returning to prepandemic levels. In a Polish national dataset,

telehealth in both outpatient health and rehabilitation services (excluding mental health) rose from prepandemic levels near zero to peak in 2020 before subsequently stabilizing at approximately one-fifth and one-third of their respective peak volumes [25]. Similarly, musculoskeletal physical therapists in the United States reported reduced telerehabilitation usage postpandemic, albeit at levels higher than prepandemic [22]. Likewise, across the US health system, overall telehealth usage peaked in 2020 and then declined but stabilized by 2023 [26,27]. Notably, however, telehealth usage was better sustained for services less dependent on hands-on care (eg, behavioral health and psychiatry) and in states where policies were put in place to ensure payment parity with comparable in-person services [26].

Although the perceived effectiveness of telerehabilitation had increased over the course of the pandemic, more than half of participants still identified concerns about the effectiveness of telerehabilitation as a primary reason to stop offering telerehabilitation consultations once able to resume in-person services. This is consistent with other studies conducted during the pandemic, where physiotherapists indicated concerns about the effectiveness of telerehabilitation for physiotherapy assessment and/or management [13,28].

Research indicates that outcomes for telerehabilitation are the same, if not better, than in-person physiotherapy for a range of conditions. For example, systematic reviews and randomized controlled trials in musculoskeletal, cardiac, and pulmonary populations demonstrate the noninferiority of telerehabilitation [29] and good validity for assessment conducted via telerehabilitation [30]. Physiotherapists' perceptions may be centered around occupational self-efficacy [31] or their own personal clinical experience of "effectiveness" rather than evidence of effectiveness in the published literature. However, our results suggest that physiotherapists were not concerned about their own ability to deliver services via telerehabilitation. Both perceived satisfaction with telerehabilitation and confidence in delivering telerehabilitation trended upward from the prepandemic to the postpandemic period, and "I was not confident with telerehabilitation" was not a key factor in our findings after restrictions were lifted (3/26, 12%) . However, large proportions of respondents who did not offer telerehabilitation at this stage said it was easier to do in-person consultations instead (20/26, 70%); they were concerned with the effectiveness of telerehabilitation (13/26, 50%), and they did not like providing care via telerehabilitation (11/26, 42%). These findings suggest that there are additional factors influencing physiotherapists' perceptions about the superiority of "hands-on" or "in-person" physiotherapy [32-34] that have not been comprehensively explored, such as the professional identity of a physiotherapist [31,35].

A qualitative study describing a successful, rapid transition to telerehabilitation during the pandemic challenges the perception that physiotherapy requires "hands-on" approaches and needs to be in person [36]. This study identified that physiotherapists' readiness and willingness to modify their approach influenced the success of telerehabilitation. In our study, physiotherapists preferred in-person consultations themselves and perceived that their patients also preferred in-person consultations, which is likely to influence whether they offer telerehabilitation to patients. While systematic reviews suggest that patient satisfaction with telerehabilitation is comparable to and often higher than in-person care [37,38], many patients report a preference for in-person physiotherapy if given a choice [37,39]. Although physiotherapists might have thought during the pandemic that patient demand for telerehabilitation would remain (eg, explaining their intention to offer it), if patient demand for it decreased (as 55% of our sample indicated), physiotherapists would likely perceive that patients prefer in-person care (and indeed 76% of our sample did).

Clinician preferences for providing in-person physiotherapy have also been explored and reported on in the literature. Despite high levels of clinician satisfaction when providing telerehabilitation in clinical trials [40,41], this does not appear to be the case for *in-practice* preference for, or satisfaction with, telerehabilitation [21,22,28]. Although satisfaction and confidence with telerehabilitation increased over time, participants in this study still perceived in-person physiotherapy to be easier. The rigorous planning or structured training required for telerehabilitation delivery in a randomized clinical trial, rather than day-to-day clinical practice, may explain this difference in perceptions, highlighting a need for training specific to the clinical implementation of telerehabilitation. Studies examining barriers to implementing telerehabilitation in routine physiotherapy practice consistently identify insufficient training for conducting telerehabilitation consultations as a primary concern [42]. To address these challenges, international clinical practice guidelines provide evidence-based recommendations and strategies for overcoming barriers, guiding the training of clinicians and facilitating effective implementation of telerehabilitation into physiotherapy practice [43].

Clinicians have long identified the technological illiteracy of clients as a barrier to the adoption of telerehabilitation in physiotherapy [42]. Despite advances in technology infrastructure, when transitioning to telerehabilitation during the COVID-19 pandemic period, clinicians still identified "technology concerns" (including clinician concerns about client ability to use technology) as a barrier to telerehabilitation use in clinical practice [28,34,36,42,44]. In our study, concerns about technical issues or patients being unable to use or access technology were not identified as primary reasons physiotherapists determined patients were unsuitable for telerehabilitation. Additionally, technical issues were only slightly or moderately disruptive to consultations. This is consistent with findings from an evaluation of consultations delivered in a randomized controlled trial, which found that technical issues occurred but were infrequent and minimally disruptive [45]. This could potentially be because, at the time data were collected for this study (2023), physiotherapists and clients had greater experience with and exposure to the technology required for telerehabilitation and had become more comfortable over time [46]. In other studies where data were collected earlier in the pandemic, it is possible that fewer people were familiar with telerehabilitation technology, hence it being a bigger barrier to delivering telerehabilitation services at the time [20,21,36].

Strengths and Limitations

The findings of this study should be interpreted with the following limitations in mind. First, this was a small convenience sample, and findings may have been skewed by self-selection bias (with those with strong opinions, either positive or negative, electing to complete the survey). Second, we asked participants to recall what they were doing before and during the pandemic several years after the fact. Therefore, it should be acknowledged that participants' responses may have been influenced by recall bias. However, our data pertaining to before and during the pandemic were consistent with other studies conducted during the pandemic and their intentions to continue (eg, 69% in our study said that during the pandemic, they intended to offer telerehabilitation after restrictions were eased). In a study by Bennell et al [20], 81% intended to continue offering telerehabilitation consultations after the pandemic, and in a study by Peng et al [28], 55% and 68% intended to continue offering phone and videoconferencing, respectively. If the opportunity arises (ie, another period of restrictions to in-person consultations), researchers should consider using prospective study designs. Moreover, because our survey encompassed both phone calls and videoconferencing, our findings may not reflect modality-specific differences in perceptions reported elsewhere [28]. This survey also only sampled physiotherapists operating within Australia's health care system, so its findings may not fully translate to other countries with different telerehabilitation policies, funding models, or cultural attitudes toward remote care. Finally, due to an error, questions about confidence, satisfaction, and perceived effectiveness after the pandemic restrictions were eased were misworded and instead asked about experiences during the pandemic. It is likely that, given that all questions before this were about easing restrictions, most respondents still answered according to the intention of the question, but we cannot discount that some answered more literally, thereby skewing the data.

Conclusions

Although telerehabilitation use surged with pandemic restrictions, it has subsequently decreased significantly, with telerehabilitation accounting for only a small proportion of the total caseload. Despite increased confidence and satisfaction with telerehabilitation, clinician preference, and physiotherapists' perceptions of patient preference for in-person care, reduced demand and the ease of in-person practice influence the use of telerehabilitation postrestrictions and suggest persistent barriers to frequent use. Addressing these barriers is crucial to enhance the long-term viability and effectiveness of telerehabilitation physiotherapy in Australia.

Funding

This work was supported by the National Health and Medical Research Council (NHMRC Project Grant 1157977); RSH is supported by an NHMRC Investigator grant (2025733); KLB is supported by a NHMRC Investigator Grant (1174431); BJL is supported by a University of Melbourne CR Roper Fellowship and Dame Kate Campbell Fellowship. For the purposes of open access, the author has applied a CC BY public copyright licence to any author-accepted manuscript version arising from this submission. The funders have no role in conduct, analysis, or reporting of this study.

Data Availability

The datasets generated during and analyzed in this study are not publicly available due to lack of ethical clearance to disclose data to third parties.

Authors' Contributions

Conceptualization: MHR, JS, BL, KLB, RSH, TR Data curation: MHR, JS Formal analysis: MHR, JS Visualization: JS Writing – original draft: MHR, JS Writing – review & editing: MHR, JS, BL, KLB, RSH, TR Funding acquisition: KLB, RSH, TR

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey.
[DOCX File, 59 KB - rehab_v13i1e81008_app1.docx]

Multimedia Appendix 2
Telerehabilitation caseload over the COVID-19 pandemic for each physiotherapy area of practice.
[PNG File, 220 KB - rehab_v13i1e81008_app2.png]

Multimedia Appendix 3
Detailed quantitative (n, %) and qualitative survey findings on physiotherapists' telerehabilitation practice across COVID-19 restrictions: reasons for not offering, reducing, ceasing, or continuing telerehabilitation; reasons patients were considered unsuitable

and the resources used to support videoconferencing; differences in consultation parameters and hybrid models of care compared with in-person practice; and postrestriction perspectives and willingness to provide telerehabilitation.
[DOCX File, 43 KB - rehab_v13i1e81008_app3.docx]

Multimedia Appendix 4
Patient suitability for telerehabilitation.
[PNG File, 41 KB - rehab_v13i1e81008_app4.png]

Multimedia Appendix 5
Technical issues.
[PNG File, 42 KB - rehab_v13i1e81008_app5.png]

References

1. Brown RC, Coombes JS, Rodriguez KJ, Hickman IJ, Keating SE. Effectiveness of exercise via telehealth for chronic disease: a systematic review and meta-analysis of exercise interventions delivered via videoconferencing. Br J Sports Med 2022 Sep;56(18):1042-1052. [doi: 10.1136/bjsports-2021-105118]
2. Dias JF, Oliveira VC, Borges PRT, et al. Effectiveness of exercises by telerehabilitation on pain, physical function and quality of life in people with physical disabilities: a systematic review of randomised controlled trials with GRADE recommendations. Br J Sports Med 2021 Feb;55(3):155-162. [doi: 10.1136/bjsports-2019-101375] [Medline: 33060156]
3. Dario AB, Cabral AM, Almeida L, et al. Effectiveness of telehealth-based interventions in the management of non-specific low back pain: a systematic review with meta-analysis. Spine J 2017 Sep;17(9):1342-1351. [doi: 10.1016/j.spinee.2017.04.008] [Medline: 28412562]
4. Laver KE, Adey-Wakeling Z, Crotty M, Lannin NA, George S, Sherrington C. Telerehabilitation services for stroke. Cochrane Database Syst Rev 2020 Jan 31;1(1):CD010255. [doi: 10.1002/14651858.CD010255.pub3] [Medline: 32002991]
5. Hwang R, Bruning J, Morris NR, Mandrusiak A, Russell T. Home-based telerehabilitation is not inferior to a centre-based program in patients with chronic heart failure: a randomised trial. J Physiother 2017 Apr;63(2):101-107. [doi: 10.1016/j.jphys.2017.02.017] [Medline: 28336297]
6. Zhou Z, Zhou X, Cui N, et al. Effectiveness of tele-rehabilitation after total hip replacement: a systematic review and meta-analysis of randomized controlled trials. Disabil Rehabil 2024 Oct;46(20):4611-4616. [doi: 10.1080/09638288.2023.2280070] [Medline: 37990882]
7. Wade V, Soar J, Gray L. Uptake of telehealth services funded by Medicare in Australia. Aust Health Rev 2014 Nov;38(5):528-532. [doi: 10.1071/AH14090] [Medline: 25219655]
8. Gray LC, Edirippulige S, Smith AC, et al. Telehealth for nursing homes: the utilization of specialist services for residential care. J Telemed Telecare 2012 Apr;18(3):142-146. [doi: 10.1258/jtt.2012.SFT105] [Medline: 22362837]
9. Brewster L, Mountain G, Wessels B, Kelly C, Hawley M. Factors affecting front line staff acceptance of telehealth technologies: a mixed-method systematic review. J Adv Nurs 2014 Jan;70(1):21-33. [doi: 10.1111/jan.12196] [Medline: 23786584]
10. Gregory WJ, Burchett S, McCrum C. National survey of the current clinical practices of the UK rheumatology physiotherapists: a short report. Musculoskeletal Care 2021 Mar;19(1):136-141. [doi: 10.1002/msc.1516] [Medline: 32975369]
11. Cottrell MA, Hill AJ, O'Leary SP, Raymer ME, Russell TG. Service provider perceptions of telerehabilitation as an additional service delivery option within an Australian neurosurgical and orthopaedic physiotherapy screening clinic: a qualitative study. Musculoskelet Sci Pract 2017 Dec;32:7-16. [doi: 10.1016/j.msksp.2017.07.008] [Medline: 28787636]
12. Lawford BJ, Bennell KL, Kasza J, Hinman RS. Physical therapists' perceptions of telephone- and internet video-mediated service models for exercise management of people with osteoarthritis. Arthritis Care Res (Hoboken) 2018 Mar;70(3):398-408. [doi: 10.1002/acr.23260] [Medline: 28437566]
13. Reynolds A, Awan N, Gallagher P. Physiotherapists' perspective of telehealth during the COVID-19 pandemic. Int J Med Inform 2021 Dec;156:104613. [doi: 10.1016/j.ijmedinf.2021.104613] [Medline: 34688969]
14. Bettger JP, Thoumi A, Marquevich V, et al. COVID-19: maintaining essential rehabilitation services across the care continuum. BMJ Glob Health 2020 May;5(5):e002670. [doi: 10.1136/bmjgh-2020-002670] [Medline: 32376777]
15. COVID-19 submission 325. : Australian Physiotherapy Association; 2020 URL: https://www.aph.gov.au/DocumentStore.ashx?id=52e7027c-1187-4d62-8f2b-53fcbcceb939&subId=685789 [accessed 2025-12-31]
16. COVID-19 (coronavirus disease 2019). Victorian Government Department of Health. 2020 Apr 1. URL: https://www.health.vic.gov.au/infectious-diseases/covid-19-coronavirus-disease-2019 [accessed 2025-12-31]
17. Aged care royal commission impact of COVID-19 on aged care supplementary submission. : Australian Physiotherapy Association; 2020 URL: https://australian.physio/sites/default/files/submission/APA_Supplementary_Submission_Impact_of_COVID-19_on_Aged_Care_August_2020.pdf [accessed 2025-12-31]

18. Report from the telehealth working group. : Medicare Benefits Schedule Review Taskforce; 2020 URL: https://www.health.gov.au/sites/default/files/documents/2021/06/final-report-from-the-telehealth-working-group.pdf [accessed 2025-12-31]
19. Telehealth guidelines – Response to COVID-19. : Australian Physiotherapy Association; 2020 URL: https://australian.physio/sites/default/files/APATelehealthGuidelinesCOVID190420FA.pdf [accessed 2025-12-31]
20. Bennell KL, Lawford BJ, Metcalf B, et al. Physiotherapists and patients report positive experiences overall with telehealth during the COVID-19 pandemic: a mixed-methods study. J Physiother 2021 Jul;67(3):201-209. [doi: 10.1016/j.jphys.2021.06.009] [Medline: 34147399]
21. Malliaras P, Merolli M, Williams CM, Caneiro JP, Haines T, Barton C. "It's not hands-on therapy, so it's very limited": telehealth use and views among allied health clinicians during the coronavirus pandemic. Musculoskelet Sci Pract 2021 Apr;52:102340. [doi: 10.1016/j.msksp.2021.102340] [Medline: 33571900]
22. McLaughlin K, Minick KI, Fritz JM, et al. Physical therapists' perceptions of telerehabilitation for patients with musculoskeletal conditions in a post-pandemic world. medRxiv. Preprint posted online on Jan 18, 2025. [doi: 10.1101/2025.01.17.25320739] [Medline: 39867367]
23. Sharma A, Minh Duc NT, Luu Lam Thang T, et al. A consensus-based Checklist for Reporting of Survey Studies (CROSS). J Gen Intern Med 2021 Oct;36(10):3179-3187. [doi: 10.1007/s11606-021-06737-1] [Medline: 33886027]
24. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
25. Krawczyk A, Marszałek M. A comparative analysis of telerehabilitation and telemedicine utilization during the COVID-19 pandemic in Poland: trends, patterns, and implications. Int J Telerehabil 2024;16(1):e6627. [doi: 10.5195/ijt.2024.6627] [Medline: 39022440]
26. Gage AD, Knight MA, Bintz C, et al. Disparities in telemedicine use and payment policies in the United States between 2019 and 2023. Commun Med (Lond) 2025 Feb 26;5(1):52. [doi: 10.1038/s43856-025-00757-2] [Medline: 40011624]
27. Mandal S, Wiesenfeld BM, Mann DM, Nov O. The "new" new normal: changes in telemedicine utilization since COVID-19. Am J Manag Care 2025 Mar 1;31(3):e74-e78. [doi: 10.37765/ajmc.2025.89700] [Medline: 40053411]
28. Peng TH, Eng JJ, Harris A, et al. A survey of the experiences of delivering physiotherapy services through telerehabilitation during the COVID-19 pandemic. Front Rehabil Sci 2024;5:1486801. [doi: 10.3389/fresc.2024.1486801] [Medline: 39512760]
29. Suso-Martí L, La Touche R, Herranz-Gómez A, Angulo-Díaz-Parreño S, Paris-Alemany A, Cuenca-Martínez F. Effectiveness of telerehabilitation in physical therapist practice: an umbrella and mapping review with meta-meta-analysis. Phys Ther 2021 May 4;101(5):pzab075. [doi: 10.1093/ptj/pzab075] [Medline: 33611598]
30. Bernhardsson S, Larsson A, Bergenheim A, et al. Digital physiotherapy assessment vs conventional face-to-face physiotherapy assessment of patients with musculoskeletal disorders: a systematic review. PLOS ONE 2023;18(3):e0283013. [doi: 10.1371/journal.pone.0283013] [Medline: 36943857]
31. Roitenberg N, Pincus T, Ben Ami N. Physiotherapy services during the COVID-19 pandemic: a mediated model of physiotherapists' self-efficacy, tele-physiotherapy role stressors, and motivation to provide tele-physiotherapy. Physiother Theory Pract 2024 Jun;40(6):1140-1149. [doi: 10.1080/09593985.2022.2138662] [Medline: 36305357]
32. Hasani F, Malliaras P, Haines T, et al. Telehealth sounds a bit challenging, but it has potential: participant and physiotherapist experiences of gym-based exercise intervention for Achilles tendinopathy monitored via telehealth. BMC Musculoskelet Disord 2021 Feb 4;22(1):138. [doi: 10.1186/s12891-020-03907-w] [Medline: 33541314]
33. Fernandes LG, Oliveira RFF, Barros PM, Fagundes FRC, Soares RJ, Saragiotto BT. Physical therapists and public perceptions of telerehabilitation: an online open survey on acceptability, preferences, and needs. Braz J Phys Ther 2022;26(6):100464. [doi: 10.1016/j.bjpt.2022.100464] [Medline: 36410257]
34. Sia LL, Sharma S, Kumar S, Ajit Singh DK. Exploring physiotherapists' perceptions of telerehabilitation for musculoskeletal disorders: insights from focus groups. Digit Health 2024;10:20552076241248916. [doi: 10.1177/20552076241248916] [Medline: 38665882]
35. Hawley-Hague H, Gluchowski A, Lasrado R, et al. Exploring the delivery of remote physiotherapy during the COVID-19 pandemic: UK wide service evaluation. Physiother Theory Pract 2024 Oct;40(10):2241-2255. [doi: 10.1080/09593985.2023.2247069] [Medline: 37610255]
36. Ross MH, Nelson M, Parravicini V, et al. Staff perspectives on the key elements to successful rapid uptake of telerehabilitation in medium-sized public hospital physiotherapy departments. Physiother Res Int 2023 Jul;28(3):e1991. [doi: 10.1002/pri.1991] [Medline: 36540908]
37. Zischke C, Simas V, Hing W, Milne N, Spittle A, Pope R. The utility of physiotherapy assessments delivered by telehealth: a systematic review. J Glob Health 2021;11:04072. [doi: 10.7189/jogh.11.04072] [Medline: 34956637]
38. Simmich J, Ross MH, Russell T. Real-time video telerehabilitation shows comparable satisfaction and similar or better attendance and adherence compared with in-person physiotherapy: a systematic review. J Physiother 2024 Jul;70(3):181-192. [doi: 10.1016/j.jphys.2024.06.001] [Medline: 38879432]
39. Lawford BJ, Bennell KL, Kimp A, Campbell PK, Hinman RS. Understanding negative and positive feelings about telerehabilitation in people with chronic knee pain: a mixed-methods study. J Orthop Sports Phys Ther 2024 Sep;54(9):594-607. [doi: 10.2519/jospt.2024.12383] [Medline: 39207737]

40. Hinman RS, Nelligan RK, Bennell KL, Delany C. “Sounds a bit crazy, but it was almost more personal:” a qualitative study of patient and clinician experiences of physical therapist-prescribed exercise for knee osteoarthritis via Skype. Arthritis Care Res (Hoboken) 2017 Dec;69(12):1834-1844. [doi: 10.1002/acr.23218] [Medline: 28217864]
41. Amin J, Ahmad B, Amin S, Siddiqui AA, Alam MK. Rehabilitation professional and patient satisfaction with telerehabilitation of musculoskeletal disorders: a systematic review. Biomed Res Int 2022;2022:7366063. [doi: 10.1155/2022/7366063] [Medline: 35958819]
42. Sia LL, Sharma S, Ing JBM, Kumar S, Singh DKA. Physiotherapists’ perceptions, readiness, enablers, and barriers to use telerehabilitation: a scoping review. J Back Musculoskelet Rehabil 2024;37(6):1441-1454. [doi: 10.3233/BMR-240009] [Medline: 38905032]
43. Lee AC, Deutsch JE, Holdsworth L, et al. Telerehabilitation in physical therapist practice: a clinical practice guideline from the American Physical Therapy Association. Phys Ther 2024 May 1;104(5):pzae045. [doi: 10.1093/ptj/pzae045] [Medline: 38513257]
44. Albahrouh SI, Buabbas AJ. Physiotherapists’ perceptions of and willingness to use telerehabilitation in Kuwait during the COVID-19 pandemic. BMC Med Inform Decis Mak 2021 Apr 8;21(1):122. [doi: 10.1186/s12911-021-01478-x] [Medline: 33832473]
45. Ross MH, Russell T, Bennell KL, et al. Technical issues occur but are infrequent and have little impact on physiotherapist-delivered videoconferencing consultations for knee osteoarthritis: a descriptive study. Musculoskelet Sci Pract 2023 Aug;66:102782. [doi: 10.1016/j.msksp.2023.102782] [Medline: 37269590]
46. Tousignant M, Boissy P, Moffet H, et al. Patients’ satisfaction of healthcare services and perception with in-home telerehabilitation and physiotherapists’ satisfaction toward technology for post-knee arthroplasty: an embedded study in a randomized trial. Telemed J E Health 2011 Jun;17(5):376-382. [doi: 10.1089/tmj.2010.0198] [Medline: 21492030]

Abbreviations

CROSS: Checklist for Reporting of Survey Studies

Edited by S Munce; submitted 21.Jul.2025; peer-reviewed by D Ceprnja, I Wilson; accepted 23.Dec.2025; published 27.Jan.2026.

Please cite as:
Ross MH, Simmich J, Lawford BJ, Bennell KL, Hinman RS, Russell T
Telerehabilitation Trends in Australian Physiotherapy and an Exploration of Factors That Influence Use After COVID-19 Restrictions: Qualitative Content Analysis
JMIR Rehabil Assist Technol 2026;13:e81008
URL: https://rehab.jmir.org/2026/1/e81008
doi:10.2196/81008

Assessing the Role of Medical Caption Technology to Support Physician-Patient Communication for Patients With Hearing Loss: Mixed Methods Pilot Study

Sarah E Hughes[1*], BA; Liang-Yuan Wu[2*], MS; Lindsay J Ma[1], MD; Dhruv Jain[2], PhD; Michael M McKee[3,4], MD, MPH

[1]University of Michigan Medical School, Ann Arbor, MI, United States

[2]Department of Computer Science and Engineering, University of Michigan, Ann Arbor, MI, United States

[3]Department of Family Medicine, University of Michigan, 1018 Fuller St., Ann Arbor, MI, United States

[4]Department of Physical Medicine and Rehabilitation, University of Michigan Medical School, Ann Arbor, MI, United States

[*]these authors contributed equally

Corresponding Author:
Michael M McKee, MD, MPH
Department of Family Medicine, University of Michigan, 1018 Fuller St., Ann Arbor, MI, United States

Abstract

Background: Speech recognition technology is widely used by individuals who are Deaf/deaf and hard-of-hearing (DHH) in everyday communication, but its clinical applications remain underexplored. Communication barriers in health care can compromise safety, understanding, and autonomy for individuals who are DHH.

Objective: This study aimed to evaluate a real-time speech recognition system (SRS) tailored for clinical settings, examining its usability, perceived effectiveness, and transcription accuracy among users who are DHH.

Methods: We conducted a pilot study with 10 adults who are DHH participating in mock outpatient encounters using a custom SRS powered by Google's speech-to-text application programming interface. We used a convergent parallel mixed-methods design, collecting quantitative usability ratings and qualitative interview data during the same study session. These datasets were subsequently merged and jointly interpreted. Participants completed postscenario surveys and structured exit interviews assessing distraction, trust, ease of use, satisfaction, and emotional response. Caption accuracy was benchmarked against professional communication access real-time translation transcripts using word error rate (WER). Because WER assigns equal weight to all tokens, it does not differentiate between routine transcription errors and those involving safety-critical clinical terms (eg, medications or diagnoses). Therefore, WER may underestimate the potential impact of certain errors in medical contexts.

Results: Across 29 clinical scenario simulations, 86% (25/29) of participants found captions nondistracting, 90% (26/29) reported them easy to follow and trustworthy, and 76% (22/29) were satisfied with the experience. Participants described the SRS as intuitive, emotionally grounding, and preferable to lip reading in masked settings. WER ranged from 12.7% to 22.8%, consistent with benchmarks for automated SRSs. Interviews revealed themes of increased confidence in following clinical conversations and staying engaged despite masked communication. Participants reported less anxiety about missing critical medical information and expressed a strong interest in expanding the tool to real-world settings, especially for older adults or those with cognitive impairments.

Conclusions: Our findings support the potential of real-time captioning to enhance accessibility and reduce the cognitive and mental burden of communication for individuals who are DHH in clinical care. Participants described the SRS as both functionally effective and personally empowering. While accuracy for complex medical terminology remains a limitation, participants consistently expressed trust in the system and a desire for its integration into clinical care. Future research should explore real-world implementation, domain-specific optimization, and the development of user-centered evaluation metrics that extend beyond transcription fidelity to include trust, autonomy, and communication equity.

(JMIR Rehabil Assist Technol 2026;13:e79073) doi:10.2196/79073

KEYWORDS

health communication; hearing loss; deafness; speech recognition software; usability testing; health care accessibility

Introduction

Effective communication is foundational to safe, equitable, and high-quality health care [1]. However, individuals who are Deaf/deaf and hard-of-hearing (DHH) often face communication barriers that compromise understanding and autonomy [2]. These barriers contribute to poor health outcomes and reduced patient engagement in real-time clinical settings [2]. The scale of this issue highlights the need to understand which communication support tools are available and provided, and to whom. In the United States, an estimated 48 million people live with some degree of hearing loss (HL), and 1 in 3 adults older than 65 years experiences disabling age-related hearing loss [3,4]. Despite this growing population, access to communication supports remains inconsistent [5,6].

Deaf individuals who use American Sign Language often receive interpreter services [7]. In contrast, oral communicators with people with HL who normally rely on spoken English are less likely to receive accommodations such as captioning, assistive listening devices, or environmental modifications [7]. Especially in clinical workflows, interpreter services are systematically implemented, whereas accommodations for oral communicators are likely not [8-12]. This gap persists despite longstanding mandates under the Americans with Disabilities Act, which mandates effective communication in health care [13]. As a result, many patients who are DHH still receive incomplete or delayed health information [5]. These gaps undermine informed decision-making, autonomy, and overall care outcomes [14,15]. Far from logistical oversights, these structural inequities perpetuate persistent disparities in care for individuals who are DHH.

These long-standing disparities became even more visible during the COVID-19 pandemic [14]. Universal masking eliminated lip reading and facial cues, which were essential supports for many individuals who are DHH and rely on oral communication [16]. This shift underscored the need for scalable solutions to maintain accessible communication in high-stakes settings [14,17].

Real-time captioning is 1 solution for improving communication access for individuals who are DHH when traditional strategies (eg, lip reading or interpreters) are unavailable [18]. Captioning tools can be deployed quickly and readily support both in-person and virtual communication [19]. However, captioning accuracy of clinical conversations may be affected by terminology unique to the medical field or speaker attribution and is understudied [19,20]. This has left a critical gap in the development of effective and equitable access tools.

By allowing both conversation partners to see each other's faces while reading the same captions, transparent or dual-visibility captioning preserves the natural flow of spoken interaction and is a promising solution for clinical communication. Prior work, such as See-Through Captions [21], See-Through Captions in a Museum Guided Tour [22], and Wearable Subtitles [23], has primarily focused on general or educational settings. Our study extends this line of research into medical contexts, where communication accuracy can directly affect patient safety and outcomes. It also emphasizes the emotional and psychological impact of captioning during clinical interactions and addresses the unique technical challenges posed by medical vocabulary and workflow integration.

In summary, we developed and evaluated a real-time captioning tool using Google's speech-to-text engine to generate live captions during simulated clinical encounters. We tested this system in dynamic, medically relevant scenarios designed to simulate typical ambulatory care encounters. In this pilot study, we explored how individuals who are DHH experienced the captioning system in these simulated encounters, focusing on usability, accuracy, and communication access.

Methods

Background

The pilot took place in a patient room at one of the Department of Family Medicine clinics. The primary goal was to assess the feasibility and acceptability of a real-time captioning tool in a clinical setting. Secondary objectives included evaluating ease of use, distraction, trust, and satisfaction, factors critical to determining whether the tool supports communication access. Quantitative and qualitative data were collected concurrently within the same study session using a convergent parallel mixed-methods design. Participants completed postscenario surveys and a brief structured exit interview during the same visit, allowing us to analyze both datasets in parallel before merging findings during the interpretation phase.

Recruitment

We recruited participants who self-identified as DHH through internal email lists compiled from prior studies, social media, and snowball sampling. Inclusion criteria included people who were DHH, preferred to communicate in spoken English, and were at least 18 years old. Recruitment materials explained that the study evaluated a real-time captioning system in simulated medical scenarios.

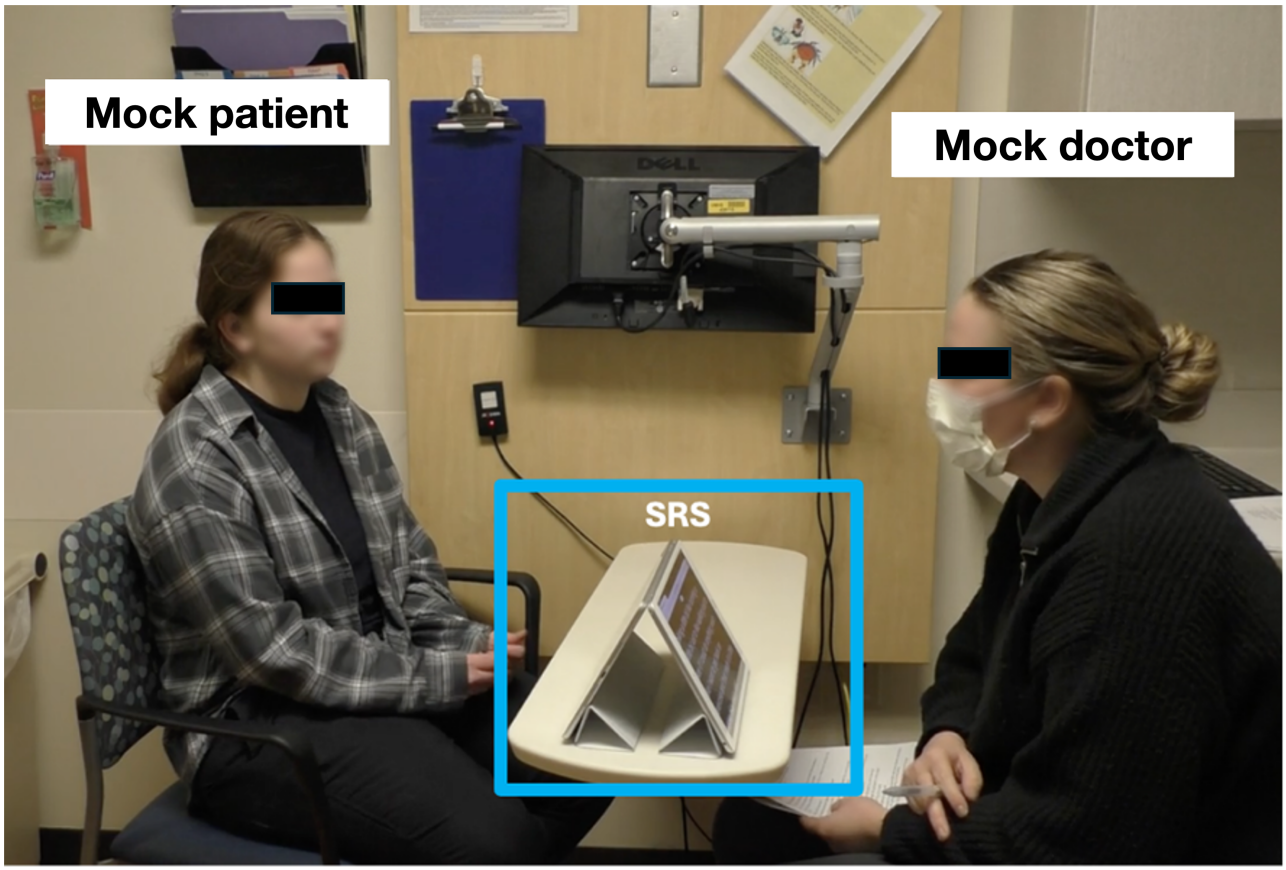
Mock Clinical Scenarios

Participants completed 3 mock clinical scenarios using the automated speech recognition system (SRS) which was developed by us. The SRS used Google's speech-to-text application programming interface to transcribe speech to text with low latency and competitive accuracy [24]. The setup included 2 iPads arranged in a tented position so that each device faced either the participant or the mock doctor. Both iPads displayed the generated captions simultaneously (Figure 1).

Before each experiment, we used a random number generator to assign scenario order for each participant. Two team members (both medical students) alternated between serving as the mock doctor (administering scenarios) or facilitator (administering postscenario surveys and exit interviews).

The scenarios were based on commonly reported primary care concerns: (1) back pain, (2) headache, and (3) high blood pressure. Scenario scripts were designed by trained medical students and a clinical faculty member to closely replicate real clinical conversations. The mock doctors wore surgical masks to simulate real-life communication barriers, such as muffled sound and loss of visual cues.

Figure 1. An example of a mock clinical scenario with the real-time speech recognition system set up on a table between the participant (left) and the mock doctor (right). A microphone on the iPad facing the mock doctor detects audio during interviews. Transcripts are displayed on both iPads in real time. SRS: speech recognition system.



Postscenario SRS Assessments

Following each scenario, participants provided feedback on the captioning system, rating it across 4 domains: distraction, ease of use, trust, and overall satisfaction (Multimedia Appendix 1). Scenario-specific questions included: “In your discussion with the mock doctor, how distracting were the captions?” “How easy or difficult was it to watch the caption while talking with the mock doctor?” “How much did you trust the accuracy of the generated captions?” “In this scenario, how satisfied were you with the captioning technology?” To reduce response bias, we alternated the direction of the scales: ease of use and trust rated from 1 (strong agreement) to 5 (strong disagreement), and satisfaction rated from 1 (strong disagreement) to 5 (strong agreement). Distraction was scaled separately from 1 (strong disagreement) to 3 (strong agreement).

Participant Survey Questions

To evaluate user experience with the SRS, participants completed a structured exit interview consisting of 9 questions (5 scalar and 4 open-ended items; Multimedia Appendix 1). To ensure accessibility, a study team member read all questions aloud while they were displayed on an iPad (Apple Inc). We audio-recorded and transcribed responses verbatim using a third-party service, then deidentified the transcripts. We reviewed audio files to clarify unclear segments. Given the brief interviews, we organized and analyzed responses in Microsoft Excel (version 16.77).

Open-ended responses were reviewed using a structured framework aligned with predefined domains: ease of use, comfort, satisfaction, trust, emotional response, and the captioning system’s ability to support or replace lip reading. Overall, 3 team members (SEH, LJM, and LW) independently applied initial codes to a subset of transcripts. Coding discrepancies were resolved through discussion, and the codebook was refined iteratively. Consistency was maintained through regular team meetings, and reflexive discussions were used to address potential bias.

Themes were identified based on frequency, relevance to study aims, and salience across participants. Representative participant comments were selected to illustrate key insights. Thematic saturation was reached when no new concepts emerged from successive interviews.

Mixed Methods Integration

To integrate quantitative and qualitative data, we used a convergent parallel approach in which both datasets were collected during the same phase, analyzed separately, and then merged during the interpretation phase. Integration occurred through (1) narrative weaving of findings across domains and (2) construction of a joint display that juxtaposed quantitative ratings with representative qualitative insights to generate meta-inferences. This approach allowed identification of areas of convergence and divergence between usability ratings and participants’ lived communication experiences.

Closed Captioning Accuracy

In addition to participant feedback, we analyzed the accuracy of the system's transcriptions. We compiled all transcripts generated by the mock doctors and compared them to professional communication access real-time translation transcripts.

We used word error rate (WER), a standard metric in automatic speech recognition (ASR) that calculates errors as the ratio of insertions, deletions, and substitutions required to align the system output with the reference [25,26]. We implemented WER calculations using the Python-based *jiwer* library, which provides standardized scoring for automated SRSs. This approach allowed us to assess how closely the SRS-generated captions matched professional-level transcription, validating the system's effectiveness in realistic use cases.

Statistical Analysis

We performed univariate analyses on demographic data and postscenario survey responses. Because of the small sample size, the study was not powered to detect subgroup differences. For transcript analysis, we segmented transcripts from mock sessions into 3 distinct scenarios. To focus on the primary use case, captioning clinician speech, we excluded utterances from participants who are DHH and analyzed only the mock doctors' speech.

Ethical Considerations

This study was approved by the University of Michigan Institutional Review Board (IRB; HUM00240244). All participants provided informed consent prior to participation. Participants were informed of the study purpose, procedures, potential risks, and their right to withdraw at any time without penalty. All study data were deidentified prior to analysis, and transcripts were reviewed to remove personally identifiable information. Audio recordings and transcripts were stored on secure, password-protected institutional servers accessible only to the study team. Participants received a US $25 Amazon gift card for their participation. The individuals depicted in the figure provided explicit written consent for publication of their images. The individuals shown in Figure 1 provided explicit written consent for their images to be published.

Results

Participant Characteristics

Overall, 11 participants who are DHH enrolled and participated in the pilot study. Due to equipment failure resulting in complete data recording loss with Participant 5, this participant was excluded from the analysis. The 10 remaining participants had an even distribution of genders (Table 1).

Table . Study participant demographics.

ID	Age (years)	Sex	Identity	Hearing loss levels[a]	Wearable technology[b]	Lip reader
P01	61	Female	HoH[c]	Severe	Yes	All of the time
P02	66	Male	HoH	Severe	Yes	Sometimes
P03	66	Male	HoH	Severe	Yes	No
P04	66	Female	HoH	Moderately severe	Yes	Sometimes
P06	43	Female	Deaf	Profound	Yes	All of the time
P07	21	Female	deaf	Moderately severe	Yes	Sometimes
P08	39	Female	HoH	Severe	Yes	All of the time
P09	24	Male	Deaf	Profound	Yes	No
P10	56	Male	deaf	Profound	Yes	Sometimes
P11	20	Male	HoH	Mild	Yes	Sometimes

[a]Hearing loss levels were self-identified, and all participants reported equal hearing loss levels bilaterally.

[b]Wearable technology includes hearing aids and cochlear implants.

[c]HoH: hard of hearing.

The mean age of the participants was 46.2 (SD 19.3) years. Six participants identified as "hard of hearing," 22 as "Deaf," and 2 as "deaf." Seven participants self-reported severe to profound HL, and all participants had bilateral HL. Five participants self-reported congenital HL, 2 reported childhood onsets of HL (<12 y old), and 2 reported HL as adults (>18 y old). Hearing aids were used by 8 participants, and 2 participants used cochlear implants. Seven participants used captioning services in the past. Seven also incorporated smartphone-based hearing assistive technology. Three used "other" tools, including using cupped hands behind ears to assist in hearing. Eight participants reported varying degrees of dependence on lip reading, but 5 participants depended sometimes on lip reading and 5 depended fully on lip reading.

Postscenario SRS Assessments

There were 29 postscenario SRS assessment surveys, 3 survey responses each from 9 participants and 2 survey responses from 1 participant. One survey response from participant P11 was not collected due to a technician error. Overall, participants found the captioning technology not distracting in 86% (25/29) of scenarios (Table 2). In 90% (26/29) of scenarios, participants

trusted the accuracy of generated transcription and felt the captions were easy to watch while conversing with the mock doctor. In 76% (22/29) of scenarios, participants were satisfied with the captioning technology. The technology was least satisfying to participants in the back pain scenarios (70% satisfaction) compared to the high blood pressure (78% satisfaction) and headache (80% satisfaction) scenarios.

Table . Summary of participant assessments regarding live captioning technology compiled from all 3 scenarios and dichotomized.

Questions[a]	Assessments	Values, n (%)
In your discussion with the mock doctor, how distracting were the captions?	Not distracting[b]	25 (86)
How easy or difficult was it to watch the caption while talking with the mock doctor?	Easy[c]	26 (90)
How much did you trust the accuracy of the generated captions?	Trusted[d]	26 (90)
In this scenario, how satisfied were you with the captioning technology?	Satisfied[e]	22 (76)

[a]For all 4 questions, n=29 since 1 of the 10 participants did not participate in 1 of the 3 scenarios.

[b]Not distracting: not at all distracting.

[c]Easy: very easy + somewhat easy.

[d]Trusted: completely trusted + somewhat trusted.

[e]Satisfied: very satisfied + somewhat satisfied.

Participant Experience Surveys

All 10 participants completed structured exit interviews following the captioning scenarios, providing reflections on their overall experience with the SRS (Table 3). Interview responses were analyzed using a predefined framework aligned with domains explored in the postscenario ratings (eg, ease of use, comfort, satisfaction, trust, emotional impact, and support for lip reading). This section summarizes participant perspectives and provides representative quotes to contextualize the quantitative results described above.

Table . Representative participant reflections by theme.

Themes	Relevant quotes[a]	Interpretation
Ease of use	“At first I wasn’t sure what to expect, but after a few lines of text I stopped even thinking about it—it just worked. That made me feel more in control.” (P04)	Participants found the system intuitive and accessible.
Comfort	“I didn’t have to strain or overthink. It just flowed naturally and I didn’t even realize how relaxed I was until the end.” (P08)	Technology reduced cognitive effort and fostered emotional ease.
Satisfaction	“I was happy. I wish all the doctors would have something like this. It made me feel like my experience mattered.” (P03)	Participant expresses satisfaction and a sense of being valued.
Safety and trust	“Because it’s live, it feels very safe. You’re not left guessing, and I felt confident nothing important was missed.” (P01)	Real-time functionality enhanced user confidence and perception of safety.
Emotional response	“I didn’t realize how much stress I usually carry during appointments. This made me feel heard and like I could finally breathe.” (P09)	System reduced communication-related anxiety and supported emotional well-being.
Support or replace lip reading	“With the mask on, it would have been extremely difficult to follow—and with the captioning, it was just leaps better. I wasn’t exhausted from trying to read lips the whole time.” (P07)	The technology was viewed as a vital alternative to lip reading, especially in masked settings.

[a]Relevant quotes from individual participants illustrating each core theme, including insight into perceived usability, comfort, satisfaction, emotional impact, and the role of real-time captions in supporting communication.

Most participants (9/10) described the system as easy to use, frequently using phrases like “very easy” or “easier than usual.” One participant remarked,

> *After a few lines of text I stopped even thinking about it—it just worked. That made me feel more in control.*

Another noted,

> *It was easier than usual because we don't have captioning. It's always nice to have it just in case you miss something.*

Participants also reported high comfort with the system. Descriptions included "very comfortable," "easy to work with," and "high comfortability."

As 1 participant shared,

> *It just flowed naturally, and I didn't even realize how relaxed I was until the end.*

Satisfaction was also high across interviews. While 76% of scenario ratings reflected satisfaction, all participants described themselves as satisfied or very satisfied in exit interviews. One stated,

> *I was happy. I wish all the doctors would have something like this.*"

Another shared,

> *I was pretty satisfied, and the captioning was spot-on.*

When asked about trust in the system, participants frequently described the captions as reliable. One participant reflected,

> *Because it's live, it feels very safe. You're not left guessing.*

A few raised questions about data privacy, with one noting,

> *I would also want to know what happens to the transcript and who has access to it.*

Participants also described emotional benefits from the technology. In total, 9 of 10 participants used words like "reassured," "relaxed," and "comfortable" to describe how the SRS made them feel. One participant shared,

> *This made me feel heard and like I could finally breathe.*

Perceptions of the captions' ability to support or replace lip reading were more varied. Several participants described the system as a helpful supplement or improvement, particularly in masked settings. As one noted,

> *With the mask on, I definitely depended on it more.*

Another stated,

> *I think it's better than lip reading.*

Others expressed that lip reading remained important, with one participant saying,

> *Not going to replace lip reading... captions help, but I still rely on visual cues.*

Beyond these predefined domains, participants spontaneously shared reflections on broader applications of the SRS. Several expressed enthusiasm for expanding its use in real-world clinical settings, with one stating,

> *I would like to see that in many doctor's offices tomorrow.*

Others suggested the system may be particularly helpful for patients who are older, have cognitive impairments, or use interpreters. A few noted that having real-time captions reduced the pressure to maintain constant visual attention, allowing for more natural communication and less fatigue.

Closed Captioning Accuracy

We collected and preprocessed transcripts from 10 mock clinical sessions. Due to varying levels of verbosity among the participants, the total transcript lengths varied substantially, ranging from 1144 to 4704 words.

Overall, participants found the SRS to be sufficiently accurate (Table 4). For instance, P04 noted that the system was *"more accurate than the phone captions"* she typically uses in daily conversations. Similarly, P06 commented on the system's effectiveness compared to human captioners, stating,

> *A lot of the captions I had were court reporters—they caption fast, but sometimes they make mistakes. ... And this one [the SRS], it's more accurate and I see words better.*

Nonetheless, participants expressed concerns about the system's ability to handle more complex or specialized medical vocabulary. For example, P10 questioned *"how it would be with more complex medical terminologies,"* in real clinical settings where more technical jargon and medication names were frequently used.

Table . The word error rate for each scenario, along with the accumulated word error rate for each participant across all 3 scenarios. These word error rate scores specifically reflect the accuracy of the automated speech recognition system in transcribing the mock doctors' speech.

ID	Mock doctor	Scenario 1[a]	Scenario 2[a]	Scenario 3[a]	Accumulated (range: 0.127-0.167)
P01	M1	0.136	0.133	0.125	0.131
P02	M2	0.193	0.141	0.133	0.153
P03	M1	0.129	0.122	0.133	0.127
P04	M2	0.228	0.128	0.151	0.167
P06	M2	0.137	0.152	0.135	0.144
P07	M1	0.127	0.134	0.132	0.133
P08	M2	0.155	0.142	0.152	0.149
P09	M1	0.136	0.131	0.141	0.137
P10	M1	0.127	0.126	0.133	0.129
P11	M2	0.147	0.185	0.134	0.151

[a]The scenario-level word error rates ranged between 0.122 and 0.228.

Joint Display of Integrated Findings

To illustrate convergence between quantitative usability ratings and qualitative interview themes, we constructed a joint display summarizing merged findings and resulting meta-inferences across key domains (Table 5).

Table . Joint display of integrated quantitative and qualitative findings

Domain	Quantitative result	Representative quote	Integrated meta-inference
Ease of use	90% rated captions "easy"	"After a few lines of text I stopped even thinking about it—it just worked."	High usability with minimal cognitive load Captions supported natural conversational flow
Comfort	Not directly measured	"It just flowed naturally, and I didn't realize how relaxed I was."	Technology reduced strain and fostered emotional ease during communication
Satisfaction	76% satisfied	"I wish all the doctors would have something like this."	Satisfaction tied to both functional value and feeling understood and supported
Safety and trust	90% trusted accuracy	"Because it's live, it feels very safe. You're not left guessing."	Real-time display strengthened perceived safety and reliability despite minor errors
Emotional response	Not directly measured	"This made me feel heard and like I could finally breathe."	Captions enhanced psychological safety and reduced anxiety—benefits not captured numerically
Support or replace lip reading	Not directly measured	"With the mask on, I depended on it more… it was leaps better."	Captions supplemented or replaced lip reading, reducing fatigue in masked settings

Discussion

Principal Results

To successfully deploy SRS in clinical settings, it is essential that the system accurately captures and reflects clinicians' speech. Our findings show that although the SRS output was not flawless, its WERs fell between 0.10 and 0.20, a range generally considered acceptable for real-world ASR use [19,26]. Furthermore, participants understood the captions with relative ease, suggesting that transcription quality was sufficient to support comprehension in simulated outpatient scenarios. However, stricter accuracy standards may be required in high-stakes contexts, such as discussions of medications or treatment options, where small errors can have serious consequences.

Although WER is widely used to evaluate ASR performance, it weighs all error types equally, regardless of their impact on comprehension [27]. Prior work has proposed alternative evaluation approaches that aim to capture semantic accuracy or user-centric measures of intelligibility and usefulness [20]. In clinical communication, we support developing evaluation metrics that align more closely with safety-critical requirements. Such metrics would be instrumental in determining when ASR

systems are truly ready for deployment in health care environments. In clinical settings, misrecognition of medical terminology can have consequences far more serious than common transcription errors, especially when involving medication names, diagnoses, or treatment instructions. Because of this, future work should consider safety-critical evaluation frameworks that go beyond traditional WER. Approaches, such as semantic error analysis, comprehension-based scoring, or accuracy, weighting for medically significant terms could better capture the real-world implications of captioning errors in health care communication.

Our participants represented a variety of ages, genders, HL levels, and degrees of dependence on lip reading. However, most participants had previously used captioning technology as an accommodation, so our usability findings may be less generalizable to individuals who are DHH with no prior captioning experience. Also, only 2 participants preferred written communication with hearing people. Therefore, satisfaction with our captioning technology may be higher than our results suggest for people who are DHH and depend more on written communication. Regardless of the scenario, most participants were satisfied with the SRS, trusted its accuracy, found it easy to watch, and were not distracted.

Participants trusted the captioning system despite occasional transcription errors, which embodies the concept of trust-in-automation frameworks, where user reliance is shaped by perceived system reliability and predictability [28]. Exit interviews revealed that beyond meeting technical expectations, the captioning system also meaningfully supported emotional connection, trust, and autonomy during clinical interactions. Participants described the captions as easy to use and grounding. They also noted reduced stress, lower cognitive fatigue, better understanding, and a stronger sense of being heard. Encouragingly, the observed reduction in stress and fatigue is consistent with prior work where assistive technology helped manage cognitive effort during information processing [29,30]. These findings suggest that accessibility tools should be evaluated not only by their accuracy but by their ability to support psychological safety and communication equity [31].

Additionally, although participants generally trusted the captioning system, a few raised concerns about transcript privacy and data handling. These concerns highlight the ethical need for transparency when implementing automated captioning in health care. This pilot used secure, locally stored recordings without identifiable data, but clinical deployment will require Health Insurance Portability and Accountability Act (HIPAA)-compliant encryption and explicit consent protocols. Adding user controls, such as options to delete transcripts or disable storage, could further strengthen trust among users who are DHH and other vulnerable populations. Nevertheless, participants recommended broader adoption of SRS, particularly for older adults and others facing progressive hearing-related communication barriers, underscoring the system's potential to improve care for a heterogeneous population of DHH patients.

Limitations

While our findings are promising, this study has several limitations. Most participants were experienced caption users and had prior familiarity with assistive communication technologies, which may have positively influenced usability and satisfaction ratings. As a result, these findings may not fully represent the experiences of individuals who are DHH and are less familiar with captioning or other accessibility tools or primary American Sign Language users. Future research should include participants with varying levels of captioning experience and a broader demographic range to better assess generalizability and identify barriers for first-time users. This study was conducted in controlled, simulated settings, which may not fully reflect the complexity and spontaneity of real-world medical encounters. Because these mock scenarios involved medical students rather than practicing clinicians, the communication dynamics may differ from authentic physician–patient interactions. Future work should therefore include real-world clinical deployments to evaluate how captioning systems perform in active care settings and adapt to diverse communication styles and environmental conditions.

Second, although our participant pool included individuals with diverse hearing identities and varying degrees of familiarity with assistive technologies, it does not capture the full range of experiences within the broader community who are DHH. Future work should include longitudinal application in various clinical settings and recruitment of a more diverse participant population to better assess long-term usability and impact.

In addition, our SRS was not specifically optimized for medical vocabulary. This limitation was evident in the system's tendency to misrecognize medical terminology, words that are infrequent in everyday speech yet crucial for accurate clinical communication. Furthermore, while we used WER as a standard quantitative evaluation metric, it does not fully capture how users who are DHH interpret and understand captions, particularly in high-stakes contexts. Future research should explore the development of domain-specific SRS trained on medical speech and adopt evaluation metrics that better reflect comprehension and user experience among individuals who are DHH. Finally, since WER assigns equal weight to all tokens, it does not differentiate between routine transcription errors and those involving safety-critical clinical terms (eg, medications or diagnoses). Therefore, WER may underestimate the potential impact of certain errors in medical contexts.

Future Directions

Improving SRS accuracy for medical terminology remains a key technical priority for clinical use. Strategies may include (1) speech recognition models on deidentified clinical audio to capture the acoustic variability of real-world medical speech [32], (2) embedding domain-specific medical dictionaries and medication name libraries into the language model of the SRS systems to reduce substitution errors [33], (3) leveraging context-aware large language models that can infer meaning from partial or uncertain input [34], and (4) integrating clinician feedback loops for rapid correction of recurring misinterpretations [35]. These enhancements would not only improve accuracy for technical vocabulary but also strengthen user trust and perceived reliability in clinical environments.

Building on these preliminary findings, future work should also explore integration with medical-domain ASR models to

enhance accuracy for specialized terminology and complex clinical dialog. Longitudinal studies will be valuable for assessing maintained usability, user trust, and performance over time. Additionally, testing captioning systems in broader clinical contexts, such as emergency care, geriatrics, and among patients with cognitive impairment, will help determine their adaptability and impact across diverse care settings.

Implications for Clinical Workflow Integration

Our findings demonstrate that real-time captioning is usable and beneficial in clinical settings for patients who are DHH, aligning with prior evidence that captioning improved recall of anesthesia-related consent conversations [36]. Given this demonstrated value, practical integration of captioning tools into clinical workflows will require thoughtful design to minimize disruption while enhancing accessibility. Participants envisioned use cases in which SRS displays could be embedded within existing electronic health record systems or mirrored on clinician tablets to preserve natural eye contact and conversational flow. Integration will also depend on clear institutional protocols for activating captioning on demand, ensuring confidentiality, and providing clinician training on how to engage with patients who are DHH using this technology. Establishing these processes could enable captioning to function as a routine accessibility feature rather than an exception, supporting both efficiency and equitable communication in care delivery.

Conclusions

This pilot study demonstrates that artificial intelligence–enhanced captioning can meaningfully improve communication experiences for individuals who are DHH in clinical settings. Participants found the system intuitive, emotionally supportive, and effective in bridging common communication barriers, especially those worsened by face masks and unfamiliar environments. While traditional captioning tools often fall short in medical contexts, integrating large language models into the speech recognition process offers a promising path toward more coherent, accurate, and human-centered accessibility. By centering on user perspectives, this study highlights the importance of evaluating assistive technologies not only for transcription quality, but for their impact on trust, inclusion, and psychological safety. Future research should build on these early insights to further refine captioning systems, examine their use in real-world clinical care, and ensure that patients who are DHH are active partners in the design of accessible digital health solutions.

Acknowledgments

The authors thank the Deaf/deaf and hard-of-hearing community members who participated in this study for generously sharing their insights and experiences. We are grateful to the Dexter Family Medicine clinic at the University of Michigan for their support in facilitating pilot testing.

Funding

This work was supported by funding from the Blue Cross Blue Shield of Michigan Foundation (2024010111) and by the University of Michigan's e-HAIL (e-Health and Artificial Intelligence Laboratory) program.

Data Availability

All data generated or analyzed during this study are included in this published article. Further inquiries can be directed to the corresponding author.

Authors' Contributions

Methodology (equal), participant recruitment (equal), investigation (equal), data curation (lead), formal analysis (lead), writing – original draft (lead), and writing – review and editing (equal): SEH
Methodology (equal), participant recruitment (equal), technology development (lead), investigation (equal), data curation (equal), formal analysis (equal), writing – original draft (equal), and writing – review and editing (equal): LYW
Methodology (equal), project administration (lead), investigation (equal), data curation (equal), formal analysis (equal), writing – original draft (equal), and writing – review and editing (equal): LJM
Conceptualization (equal), technology development (supporting), methodology (supporting), writing – review and editing (equal), and supervision (equal): DJ
Conceptualization (equal), methodology (equal), writing – original draft (supporting), writing – review and editing (equal), and supervision (lead): MMM

Conflicts of Interest

None declared.

Multimedia Appendix 1
Structured exit interview questions.
[DOCX File, 17 KB - rehab_v13i1e79073_app1.docx]

References

1. Guttman OT, Lazzara EH, Keebler JR, Webster KLW, Gisick LM, Baker AL. Dissecting communication barriers in healthcare: a path to enhancing communication resiliency, reliability, and patient safety. J Patient Saf 2021 Dec 1;17(8):e1465-e1471. [doi: 10.1097/PTS.0000000000000541] [Medline: 30418425]
2. McKee M, James TG, Helm KVT, et al. Reframing our health care system for patients with hearing loss. J Speech Lang Hear Res 2022 Oct 17;65(10):3633-3645. [doi: 10.1044/2022_JSLHR-22-00052] [Medline: 35969852]
3. Quick statistics about hearing, balance, & dizziness. NIH. URL: https://www.nidcd.nih.gov/health/statistics/quick-statistics-hearing [accessed 2025-05-26]
4. Culturally affirmative and linguistically accessible services. National Association of the Deaf. URL: https://www.nad.org/resources/health-care-and-mental-health-services/mental-health-services/culturally-affirmative-and-linguistically-accessible-services/ [accessed 2025-05-10]
5. James TG, Coady KA, Stacciarini JMR, et al. "They're not willing to accommodate deaf patients": communication experiences of deaf american sign language users in the emergency department. Qual Health Res 2022 Jan;32(1):48-63. [doi: 10.1177/10497323211046238] [Medline: 34823402]
6. Marlow NM, Samuels SK, Jo A, Mainous AG. Patient-provider communication quality for persons with disabilities: a cross-sectional analysis of the Health Information National Trends Survey. Disabil Health J 2019 Oct;12(4):732-737. [doi: 10.1016/j.dhjo.2019.03.010] [Medline: 30995967]
7. Diamond L, Izquierdo K, Canfield D, Matsoukas K, Gany F. A systematic review of the impact of patient-physician non-english language concordance on quality of care and outcomes. J Gen Intern Med 2019 Aug;34(8):1591-1606. [doi: 10.1007/s11606-019-04847-5] [Medline: 31147980]
8. Barnett S, McKee M, Smith SR, Pearson TA. Deaf sign language users, health inequities, and public health: opportunity for social justice. Prev Chronic Dis 2011 Mar;8(2):A45. [Medline: 21324259]
9. Barnett SL, Matthews KA, Sutter EJ, et al. Collaboration with deaf communities to conduct accessible health surveillance. Am J Prev Med 2017 Mar;52(3 Suppl 3):S250-S254. [doi: 10.1016/j.amepre.2016.10.011] [Medline: 28215374]
10. Pollard RQ, Barnett S. Health-related vocabulary knowledge among deaf adults. Rehabil Psychol 2009 May;54(2):182-185. [doi: 10.1037/a0015771] [Medline: 19469608]
11. Mitra M, Akobirshoev I, McKee MM, Iezzoni LI. Birth outcomes among U.S. women with hearing loss. Am J Prev Med 2016 Dec;51(6):865-873. [doi: 10.1016/j.amepre.2016.08.001] [Medline: 27687529]
12. Alexander A, Ladd P, Powell S. Deafness might damage your health. Lancet 2012 Mar 17;379(9820):979-981. [doi: 10.1016/S0140-6736(11)61670-X] [Medline: 22423872]
13. ADA.gov. URL: https://www.ada.gov [accessed 2025-05-27]
14. McKee M, Moran C, Zazove P. Overcoming additional barriers to care for deaf and hard of hearing patients during COVID-19. JAMA Otolaryngol Head Neck Surg 2020 Sep 1;146(9):781-782. [doi: 10.1001/jamaoto.2020.1705] [Medline: 32692807]
15. Iezzoni LI, O'Day BL, Killeen M, Harker H. Communicating about health care: observations from persons who are deaf or hard of hearing. Ann Intern Med 2004 Mar 2;140(5):356-362. [doi: 10.7326/0003-4819-140-5-200403020-00011] [Medline: 14996677]
16. Moreland CJ, Ruffin CV, Morris MA, McKee M. Unmasked: how the COVID-19 pandemic exacerbates disparities for people with communication-based disabilities. J Hosp Med 2021 Mar;16(3):185-188. [doi: 10.12788/jhm.3562] [Medline: 33617440]
17. Bernard A, Weiss S, Rahman M, et al. The impact of COVID-19 and pandemic mitigation measures on persons with sensory impairment. Am J Ophthalmol 2022 Feb;234:49-58. [doi: 10.1016/j.ajo.2021.06.019] [Medline: 34197781]
18. Kawas S, Karalis G, Wen T, Ladner RE. Improving real-time captioning experiences for deaf and hard of hearing students. Presented at: ASSETS '16: Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility; Oct 23-26, 2016. [doi: 10.1145/2982142.2982164]
19. Kafle S, Huenerfauth M. Evaluating the usability of automatically generated captions for people who are deaf or hard of hearing. Presented at: ASSETS '17: Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility; Oct 20 to Nov 1, 2017. [doi: 10.1145/3132525.3132542]
20. Favre B, Cheung K, Kazemian S, et al. Automatic human utility evaluation of ASR systems: does WER really predict performance? Presented at: Interspeech 2013; Aug 25-29, 2013 URL: https://www.isca-archive.org/interspeech_2013/favre13_interspeech.pdf [accessed 2025-12-15]
21. Yamamoto K, Suzuki I, Shitara A, Ochiai Y. See-through captions: real-time captioning on transparent display for deaf and hard-of-hearing people. Presented at: ASSETS '21: Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibilit; Oct 18-22, 2021. [doi: 10.1145/3441852.3476551]
22. Suzuki I, Yamamoto K, Shitara A, Hyakuta R, Iijima R, Ochiai Y. See-through captions in a museum guided tour: exploring museum guided tour for deaf and hard-of-hearing people with real-time captioning on transparent display. Presented at: Computers Helping People with Special Needs: 18th International Conference, ICCHP-AAATE 2022; Jul 11-15, 2022. [doi: 10.1007/978-3-031-08648-9_64]

23. Olwal A, Balke K, Votintcev D. Wearable subtitles: augmenting spoken communication with lightweight eyewear for all-day captioning. Presented at: UIST '20: Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology; Oct 20-23, 2020 URL: https://dl.acm.org/doi/proceedings/10.1145/3379337 [accessed 2025-12-15] [doi: 10.1145/3379337.3415817]
24. Speech-to-Text. Google Cloud. URL: https://cloud.google.com/speech-to-text [accessed 2025-05-27]
25. Morris AC, Maier V, Green P. From WER and RIL to MER and WIL: improved evaluation measures for connected speech recognition. Presented at: Interspeech 2004; Oct 4-8, 2004 URL: https://www.isca-archive.org/interspeech_2004/morris04_interspeech.pdf [accessed 2025-12-15] [doi: 10.21437/Interspeech.2004-668]
26. jitsi/jiwer. GitHub. URL: https://github.com/jitsi/jiwer [accessed 2025-05-27]
27. Ye-Yi W, Acero A, Chelba C. Is word error rate a good indicator for spoken language understanding accuracy. Presented at: 2003 IEEE Workshop on Automatic Speech Recognition and Understanding (IEEE Cat No03EX721); Nov 30 to Dec 4, 2003. [doi: 10.1109/ASRU.2003.1318504]
28. Lee JD, See KA. Trust in automation: designing for appropriate reliance. Hum Factors 2004;46(1):50-80. [doi: 10.1518/hfes.46.1.50_30392] [Medline: 15151155]
29. Sweller J. Cognitive load during problem solving: effects on learning. Cogn Sci 1988 Jun;12(2):257-285. [doi: 10.1016/0364-0213(88)90023-7]
30. Paas F, Renkl A, Sweller J. Cognitive load theory and instructional design: recent developments. Educ Psychol 2003 Jan 1;38(1):1-4. [doi: 10.1207/S15326985EP3801_1]
31. Bickmore TW, Paasche-Orlow MK. The role of information technology in health literacy research. J Health Commun 2012;17 Suppl 3(sup3):23-29. [doi: 10.1080/10810730.2012.712626] [Medline: 23030559]
32. Afonja T, Olatunji T, Ogun S, Etori NA, Owodunni A, Yekini M. Performant ASR models for medical entities in accented speech. Presented at: Interspeech 2024; Sep 1-5, 2024 URL: https://www.isca-archive.org/interspeech_2024/afonja24_interspeech.pdf [accessed 2025-12-15]
33. Michalopoulos G, Wang Y, Kaka H, Chen H, Wong A. UmlsBERT: clinical domain knowledge augmentation of contextual embeddings using the unified medical language system metathesaurus. Presented at: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics; Jun 6-11, 2021 URL: https://aclanthology.org/2021.naacl-main [accessed 2025-12-15] [doi: 10.18653/v1/2021.naacl-main.139]
34. Hsie C, Moreira C, Nobre IB, et al. DALL-m: context-aware clinical data augmentation with llms. arXiv. Preprint posted online on May 1, 2025. [doi: 10.48550/ARXIV.2407.08227]
35. Wu LY, Jain D. EvolveCaptions: empowering DHH users through real-time collaborative captioning. arXiv. Preprint posted online on Oct 2, 2025. [doi: 10.48550/ARXIV.2510.02181]
36. Spehar B, Tye-Murray N, Myerson J, Murray DJ. Real- time captioning for improving informed consent: patient and physician benefits. Reg Anesth Pain Med 2016;41(1):65-68. [doi: 10.1097/AAP.0000000000000347] [Medline: 26650432]

Abbreviations

ASR: automatic speech recognition
DHH: Deaf/deaf or hard of hearing
HIPAA: Health Insurance Portability and Accountability Act
HL: hearing loss
SRS: speech recognition system
WER: word error rate

Edited by S Munce; submitted 14.Jun.2025; peer-reviewed by A Roundtree, SD Ubur; revised version received 21.Nov.2025; accepted 03.Dec.2025; published 15.Jan.2026.

Please cite as:
Hughes SE, Wu LY, Ma LJ, Jain D, McKee MM
Assessing the Role of Medical Caption Technology to Support Physician-Patient Communication for Patients With Hearing Loss: Mixed Methods Pilot Study
JMIR Rehabil Assist Technol 2026;13:e79073
URL: https://rehab.jmir.org/2026/1/e79073
doi:10.2196/79073

Original Paper

Voice-Assisted Technology for People With Parkinson's Disease Experiencing Speech and Voice Difficulties: Co-Designing Solutions Using Design Thinking

Jodie Mills[1], BSc, MSc; George Kernohan[2], BSc, PhD; Katy Pedlow[1], BSc, PhD; Orla Duffy[1], BSc, PhD

[1]School of Health Sciences, Faculty of Life and Health Sciences, University of Ulster, Belfast, Northern Ireland, United Kingdom

[2]School of Nursing, Faculty of Life and Health Sciences, University of Ulster, Belfast, United Kingdom

Corresponding Author:
Jodie Mills, BSc, MSc
School of Health Sciences
Faculty of Life and Health Sciences
University of Ulster
2-24 York Street
Belfast, Northern Ireland, BT15 1AP
United Kingdom
Phone: 44 2895365308
Email: mills-j12@ulster.ac.uk

Abstract

Background: While smart speakers are emerging as a novel health care technology, people with Parkinson's Disease (PwPD) and speech and language therapists (SaLTs) have reported difficulties using smart speakers with speech and voice impairments in research. To date, PwPD have identified frustration with having to repeat themselves to be understood, devices timing out before they had finished speaking, and being unable to have a conversation with smart speakers. SaLTs have reported technical and practical challenges in implementing voice-assisted technology tools. Both PwPD and SaLTs indicated a lack of knowledge about what smart speakers could do, as well as concerns about privacy and the listening nature of the devices.

Objective: This study aims to co-design solutions that support the use of smart speakers for speech and voice difficulties experienced by PwPD.

Methods: Based on the Design Thinking framework, a multistage design process was conducted, involving a lay steering group and 2 online co-design workshops. Twenty participants, including PwPD, carers, SaLTs, design and technology experts, and third-sector staff, collaborated during the co-design workshops. The ideate phase included brainstorming and ranking, and conventional content analysis was used to specify prototypes.

Results: Two main prototypes were created: (1) education and guidance, including privacy and therapeutic usage guides for PwPD and SaLTs to address troubleshooting and delivery considerations; and (2) new speech and language therapy (SLT)–specific features for smart speakers. Participants provided feedback on their experiences of co-design, highlighting feeling valued, the balance of perspectives, and making improvement suggestions. Feedback aligned with the UK standards for public involvement.

Conclusions: Smart speakers could enhance accessibility, therapy engagement, and long-term speech outcomes, offering scalable, cost-effective solutions to support SLT services, patient independence, and reduced service demand. Smart speaker solutions with a SLT focus enable PwPD to self-manage speech and voice difficulties at home and reinforce therapy gains between clinic visits. Co-designed with users, these prototypes are intended to address health disparities and relieve pressure on SLT services, offering a scalable and sustainable solution that enhances efficiency and supports ongoing rehabilitation within health care systems.

(JMIR Rehabil Assist Technol 2026;13:e84364) doi:10.2196/84364

KEYWORDS

voice-assisted technology; speech and language therapy; co-design; participatory methods; dysarthria; speech and voice; Parkinson's Disease

Introduction

Background

Voice-assisted technology (VAT) is defined as a device that uses natural language processing or automatic speech recognition (ASR) to interpret spoken language and translate it into actionable requests.

Smart speakers are commercially available VAT devices that are controlled using voice commands and are usually connected to the internet (current examples include Amazon Alexa and Google Assistant). They can feature built-in control systems for tasks on demand, including smart home automation, providing general information (not limited to weather, recipes, or health information), person-to-person calls, sending and receiving messages, and playing music. New models with screens can also support audio and video streaming. Smart speakers are readily available for purchase and use by the general public [1].

It has been reported that VAT prompts some participants with speech difficulties to modify their speech to enable interaction with VAT [2-6]. People with Parkinson's disease (PwPD) have reported adapting their speech by speaking more slowly, loudly, and clearly when interacting with a smart speaker [4]. Considering that 90% of PwPD present with reduced speech intelligibility and limited vocal loudness [7], VAT may hold potential as a therapeutic adjunct in speech and language therapy (SLT). This prior evidence indicates that VAT may enhance access to therapy [8].

Some therapists have reported using VAT to promote improved volume, clarity, and intelligibility of speech [9]. In addition to offering biofeedback on speech clarity, these tools have provided structured opportunities for home-based practice, fostering self-awareness and supporting the self-management of dysarthria and other speech difficulties [9,10]. PwPD have reported increased clarity of speech and volume when using VAT, and have used VAT as a communication partner to practice their speech and rebuild confidence in using their voice [11]. Both speech and language therapists (SaLTs) and PwPD agree that the objective nature of VAT is key to promoting interaction and providing feedback on speech. Textbox 1 presents a hypothetical vignette illustrating how a person with speech or voice difficulties may interact with smart speakers. This vignette is informed by the understanding and findings of previous research [11].

Textbox 1. Case study

1. Case

John is 65 and has had Parkinson's disease for 5 years. His phonation is impacted by poor breath support, resulting in a breathy, hoarse voice with low volume. His articulation is reduced, resulting in imprecise speech production, which reduces speech clarity and intelligibility. His speech is also hypernasal, and nasal emissions are noted. He has hypokinetic dysarthria. At home, his family can understand him, but he is frequently told that they "can't hear him" and that he "needs to speak up." This is frustrating for John, as he reports that "he feels like he is shouting," which suggests impaired self-awareness of his speech.

2. Use

John uses his smart speaker daily. When he speaks to the smart speaker, it replies with "Sorry, I didn't get that" approximately 50% of the time. As a result, John raises his volume and repeats his request. Often, he uses a loud voice, overarticulates his words, slows down, and speaks as soon as he takes a breath. The smart speaker responds when he uses these strategies. This demonstrates that smart speakers provide feedback on volume and clarity of speech in the form of an external cue: "Sorry, I didn't get that." This can encourage increased self-awareness of speech volume and intelligibility, and result in the use of LOUD, clear speech strategies. As John's smart speaker can time out before he has finished speaking, he uses *adaptive listening mode* (available on Amazon devices), which is found in the accessibility settings and gives him longer to speak.

John also plays the game "Word Tennis" on his smart speaker. He has to think of words within a semantic category quickly and remember to use a LOUD, clear voice when answering. This task focuses on a word-level activity within the speech hierarchy and adds a cognitive load to increase difficulty, which aligns with Lee Silverman Voice Treatment (LSVT) LOUD principles [12]. He also enjoys sport and cooking and uses his smart speaker to search for recipes. Common functional requests include "Add cheese and potatoes to my shopping list," "Show me my cooking library," and "What was the Man United score today?" Sometimes, he even uses his smart speaker like a diary: "Leave a sticky note for...", where he records a voice note on his smart speaker to remind someone to feed their dog.

3. Summary

Overall, interacting with his smart speaker allows John to practice a LOUD, clear voice at home, with external feedback on speech volume and clarity that may help improve his self-awareness.

Despite the facilitators discussed in Textbox 1, several barriers to the effective use of VAT among PwPD and SaLTs remain [9,11]. For example, PwPD have reported feeling frustrated by needing to repeat themselves to be understood, by devices timing out before they had finished speaking, and by being unable to have a conversation with their smart speaker [11]. SaLTs also indicated that they faced technical and practical challenges in implementing VAT tools [9]. Both PwPD and SaLTs reported a lack of knowledge about smart speaker capabilities and concerns surrounding privacy and data security.

Addressing these challenges is essential to enable the integration of VAT into SLT practice. Design Thinking is a user-centered innovation framework used to guide the development of new health care technologies, often utilizing co-design approaches [13-15]. It offers a structured approach to identifying problems and generating solutions through empathy, collaboration, and iterative prototyping and testing. This study is informed by the define, ideate, and prototyping phases of the Design Thinking process. Figure 1 outlines the Design Thinking process, and Table 1 shows the connections between the phases of the Design

Thinking framework, the specific research questions to be addressed, and the methods used.

Co-design has been used to foster collaboration that stimulates new ideas, clarifies concepts, and creates solutions that prioritize the needs and lived experiences of end users [16]. Co-design workshops have been used in SLT, health technology research, and with older adult populations [17-19], with improved outcomes for technology adoption compared with noncollaborative design processes [17,19]. Co-design is critical when developing technologies for SLT [18] and has value in engaging people with communication difficulties [20-22]. We set out to follow the co-design cycle and principles, meeting the criteria for true co-design under the ladder of co-production [23-25], through the identification and development of recommendations from participants with communication difficulties [26,27].

Figure 1. The Design Thinking stages [13] from empathize to prototyping, which can be used during the development of new health care technologies [15].

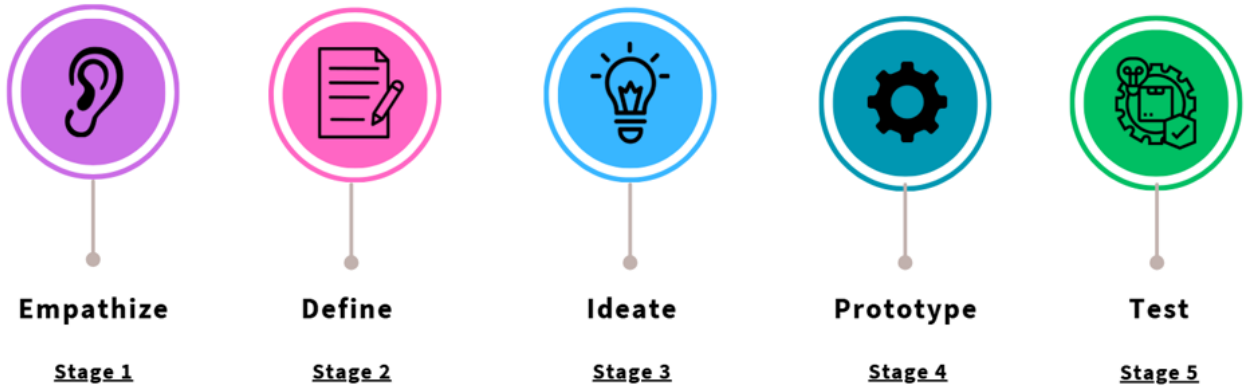


Table 1. Ways in which existing barriers to the therapeutic use of VAT[a] by people with Parkinson's can be solved.

Objective	Objective 1: To consider and select problem statements from the perspectives of experts by experience.	Objective 2: To create solutions to problem statements associated with VAT usage, alongside people with Parkinson's, carers, SaLTs[b], technology and design experts, and third-sector representatives.	Objective 3: To prioritize co-designed solutions to inform prototype VAT interventions.	Objective 3: To prioritize co-designed solutions to inform prototype VAT interventions.
Design Thinking stage	Stage 2: Define	Stage 3: Ideate	Stage 3: Ideate	Stage 4: Prototype
Method	Patient, public involvement workshop	Co-design workshop A	Co-design workshop B	Inductive content analysis

[a]VAT: voice-assisted technology.

[b]SaLT: speech and language therapist.

This research is intended to address previously noted barriers to VAT use [9,11] and aims to co-design solutions to previously identified challenges with VAT by working with PwPD, carers, SaLTs, charity representatives, and technology and design experts. This approach was taken to ensure that new technologies can be used in ways that meet end user needs. We sought to create solutions by using commercial technology, without coding or modifying VAT devices. This ensures that solutions are low cost and accessible, enhancing the potential for wider adoption of VAT in SLT contexts.

Aim

We set out to co-design solutions to support the use of smart speakers in SLT to improve volume and intelligibility for PwPD, using a Design Thinking framework. The research addressed the following question: "How can we facilitate the therapeutic use of VAT by people with Parkinson's Disease?"

Our study objectives (mapped onto Design Thinking stages) are as follows:

- To consider and select problem statements from the perspectives of experts by experience (Define).
- To co-create solutions to problem statements associated with VAT usage, alongside PwPD, carers, SaLTs, technology and design experts, and third-sector representatives (Ideate).
- To prioritize co-designed solutions to inform prototype VAT interventions (Prototype).

Methods

Participatory Co-Design Approach

Participatory methods such as co-design allow interventions to be designed around end user needs. This study co-designed solutions to previously identified barriers regarding the use of VAT when speech and voice difficulties were present. This results in technology that more readily meets user needs [28] and helps to avoid digital exclusion [29]. Workshops were held online, removing geographical and physical barriers and

enabling SaLTs from throughout the United Kingdom to share their experiences.

Ethical Considerations

Ethical approval was granted by the Ulster University Research Ethics Committee in January 2025 (approval number FCNUR-24-078-A). This study is part of a larger PhD project using Design Thinking. Previous phases of work aligned with the empathize stage, and the current co-design phase aligns with the define, ideation, and prototyping stages. All participants provided informed consent before the workshops. Participant outputs were anonymous, and ground rules were agreed upon to maintain confidentiality. Participants did not receive payment or financial incentives.

Patient and Public Involvement

A patient and public involvement (PPI) steering group was established to provide a voice for key stakeholders and ensure their active role in shaping the research. This group included a SaLT with firsthand experience using VAT in clinical settings, a person living with Parkinson's, and a caregiver. These 3 experts by experience coassessed the barriers to VAT usage identified in previous research [9,11] and co-decided the top 5 problems that reflected their experiences, in keeping with the cycle of coproduction [24].

Study Recruitment

PwPD and carers were recruited via a third-sector organization (Parkinson's UK) using advertisements on the Parkinson's UK research portal, Research Support Network monthly emails, and flyers at local Parkinson's support groups in Northern Ireland. SaLTs were recruited through the Royal College of Speech and Language Therapists, including the Parkinson's Clinical Excellence Network. In addition, Parkinson's UK staff and technology or design experts were recruited through the lead author's (JM) professional network. Some participants had established a relationship with the lead researcher through work with the local branch of Parkinson's UK in Northern Ireland.

Previous research was used to determine the number of participant collaborators invited to share their experiences during the workshops (n=20) [30,31]. Participants were asked to contact the research team to express interest in the study and were screened according to predefined criteria (Table 2). Potential participants were sent study information and consent forms by email or post, depending on preference, and were asked to indicate their availability. Once consent was obtained, participants completed a demographic survey and received links for the online workshops. This enabled interaction between diverse experiences. Participants were placed into smaller, experience-diverse groups of 4-5 participants to encourage idea generation in a safe and supportive environment.

Table 2. Inclusion and exclusion criteria for people with Parkinson's, carers, speech and language therapists, third-sector representatives, and technology or design experts.

Participant group	Inclusion	Exclusion
People with Parkinson's	• Adults over 18 years old • Mild to moderate dysarthria/voice difficulties (to include users of augmentative, alternative communication) • Diagnosis of Parkinson's disease • Current or previous use of VAT[a] • Have access to a laptop, with a camera, that facilitates videoconferencing software	• Moderate or severe cognitive impairment • History of other neurological disorders
Carers	• Adults over 18 years old • Live with or care for a PwPD[c] or both • Experience of facilitating the use of VAT with a PwPD • Have access to a laptop, with a camera, that facilitates videoconferencing software	N/A[b]
Speech and language therapists	• Adults over 18 years old • Who currently have/have had a clinical caseload of PwPD in the past 5 years • Who have used VAT in practice and have basic knowledge of the devices • Have a laptop, with a camera, that facilitates videoconferencing software	N/A
Third-sector staff	• Adults over 18 years old • Currently working in a third-sector organization for PwPD • Involvement and relationships with the local Parkinson's community • Basic knowledge of speech and voice difficulties in Parkinson's disease • Have a laptop, with a camera, that facilitates videoconferencing software	N/A
Technology/design experts	• Adults over 18 years old • Experience of VAT and detailed knowledge of its capabilities or relevant experience in designing or developing health care technologies	N/A

[a]VAT: voice-assisted technology.

[b]N/A: not applicable.

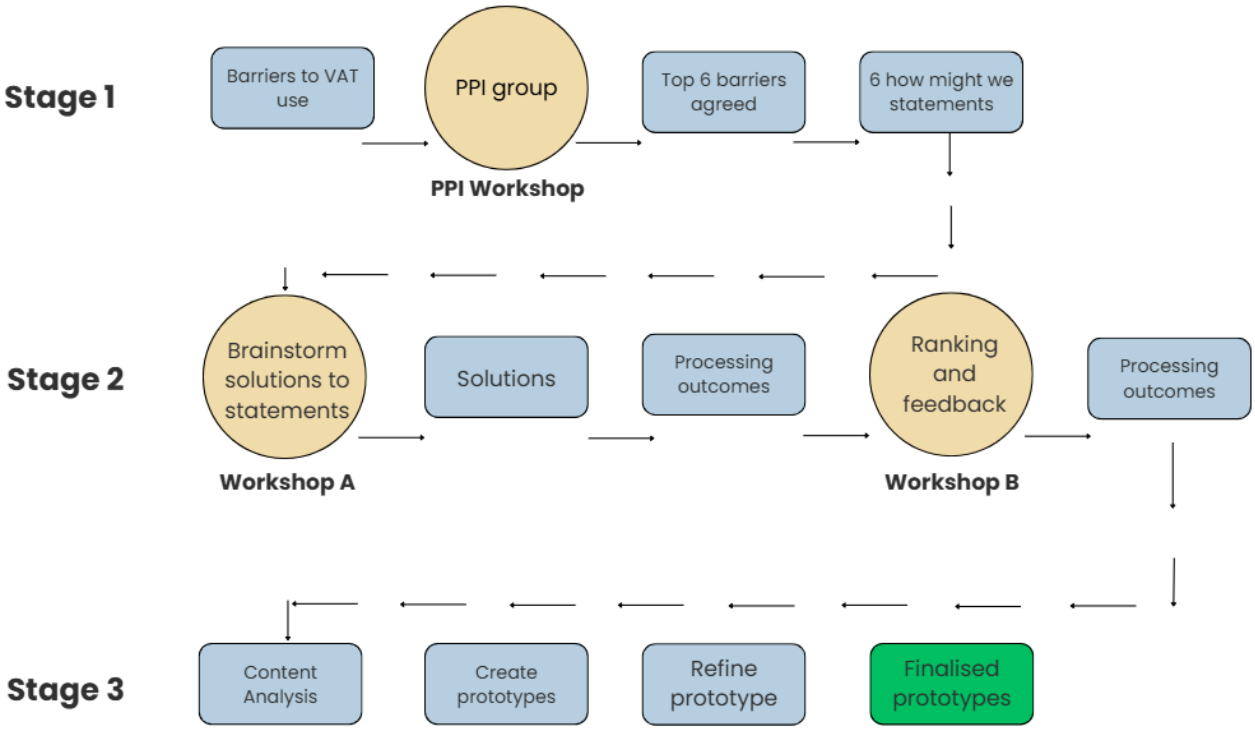
[c]PwPD: people with Parkinson's disease.

Procedure

Overview

Using principles underpinned by Design Thinking and Participatory methodology, a series of 2 co-design workshops were undertaken, informed by insights gained from previous research [30-32]. The 3-stage co-design procedure is outlined below (see Figure 2) and aligns with the define, ideate, and prototype phases of the Design Thinking framework, as presented in Figure 1.

Figure 2. The co-design process from the development of problem statements, through workshop completion, analysis, prototype creation, and refinement. PPI: patient and public involvement; VAT: voice-assisted technology.



Stage 1: Define Phase

Barriers identified by PwPD, their carers, and SaLTs during an earlier stage of the research [9,11] were reviewed by the PPI group, ensuring that the research was shaped by lived experience. Researchers JM and OD were present during the workshop. The group identified and agreed on the top 6 problems to be brought forward to workshop A, and, as per Design Thinking guidance, these were reframed into "How Might We" statements. These "How Might We" statements are used in Design Thinking to help people reframe a problem-focused perspective into solution-focused thinking [33], and previous research highlights the value of shaping research based on lived experience, as this ensures that the problems being solved are meaningful to end users [34]. The final 6 "How Might We" statements are presented in Table 3.

Table 3. Procedures and outcomes for stages 1, 2, and 3 of the co-design process.

Stage	Design Thinking framework stage	Procedure	Outcomes
1	Define	**PPI[a] workshop** • The PPI group was presented with all barriers from previous research [9,11]. • JM shared her screen via videoconferencing, using Canva PowerPoint software to present the barriers identified in previous research and to assist with live decision-making. This acted as contemporaneous notes. OD also took field notes. • PPI group identified 6 top problems—felt to be a reasonable number of barriers to brainstorm solutions to within an hour. • 6 problems reframed into "How Might We" statements.	• The following "How Might We" statements were used in workshop A: • How might we help people understand smart speaker privacy and reduce their fears? • How might we help people when smart speakers do not work? • How might we help people to have a conversation with a smart speaker? • How might we help people to know what smart speakers can do? • How could we deliver this information? • How could smart speaker technology be adapted?
2a	Ideate	**Co-design workshop A** • A document with problem statements and examples of the problems was emailed to the participants before the workshop. • Welcome and introduction PowerPoint, which presented an overview of previous research, problems experienced by people using smart speakers, and an overview of the co-design process. • Two example brainstorming activities completed. • Smaller groups discussed 6 problem statements, rotating after 10 minutes on each problem statement. All groups completed different problems at the same time. • Facilitators shaped discussions and noted contributions on a live Word document. Each problem statement had a separate Word document. • Participants were invited to share solutions to the problem statements around using VAT[b] with a speech or voice difficulty. • Participants were thanked for their time and the workshop ended.	• Solutions to each problem statement were reviewed. Similar solutions for each problem statement were grouped together and combined. No content was removed, while ensuring there was a feasible number of ideas to rank. • Ideas are presented fully in Multimedia Appendix 1, as they appeared during the workshop. • A reduced number of solutions were placed into tables for each group to rank in workshop B.
2b	Ideate	**Co-design workshop B** • Welcome and recap of solutions generated in workshop A. Participants received a document of the solutions ahead of the workshop to aid their preparation. • Each facilitator worked through 3 solutions documents with their group and ranked their top 5 priorities for each solution. • A similar procedure to workshop A was followed: each solution was captured in a separate Word document, and groups rotated after 10 minutes on each problem statement, inputting to the same solution document. • Each participant was asked a short series of questions about their experience of the co-design workshops. Questions were based on the UK standards for public involvement. • Participants were thanked for time, next steps were explained, and the workshop ended.	• Lead researcher (JM) collated solutions and rankings and removed any solutions that were not ranked. • Solutions ranked and their rankings taken forward to stage 3 (prototyping).

Stage	Design Thinking framework stage	Procedure	Outcomes
3	Prototyping	**5 stages of inductive content analysis conducted for solutions and rankings from workshop B** • Initial and final codes were discussed for refinement and agreement with the research team and broadly grouped into themes. • Themes were mapped into 2 prototypes.	• Two prototypes were created, namely, prototype 1 (education and guidance) and prototype 2 (developing new SLTc-specific features for smart speakers). • An overview of the link between workshop 2 outputs and final prototypes is available in Multimedia Appendix 2. • Participant feedback was collated.

[a]PPI: patient and public involvement.

[b]VAT: voice-assisted technology.

[c]SLT: speech and language therapy.

Stage 2: Ideate Phase

Co-design workshops were facilitated by 2 qualified SaLTs (JM and OD), 2 health care professionals (KP and RB), and an academic (GK), with participants working in smaller groups during the workshops. Facilitator guidance and training before the workshops ensured methodological consistency when participants worked within the groups (Multimedia Appendix 3). Co-design principles of valuing lived experience, sharing power, and respect were presented at the beginning of each workshop, aiming to reduce power imbalances between researchers and participants. Both workshops lasted approximately 1 hour and were conducted via videoconferencing to enable data collection across a wider geographical area [35], avoid travel, and facilitate workshops in the evenings.

Data collection was recorded as notes in live Word (Microsoft Corporation) documents. By recording content-only contributions, participant anonymity was ensured from the outset, as no identifiers were associated with the contributions. Although participants may have been known to other group members, they were asked to respect everyone's right to confidentiality by not sharing contributions outside the group setting. Additionally, workshops were not audio- or video-recorded, in keeping with co-design principles. Although this may have contributed to some data loss, the live recording of workshop contributions helped to mitigate this risk.

In workshop A, participants brainstormed solutions to the "How Might We" statements shown above. The process for this workshop is shown in Table 3. Following workshop A, these solutions were refined by combining similar ideas and removing duplicates (Multimedia Appendix 1) in preparation for workshop B. In workshop B, participants reviewed the solutions to the 6 problem statements created during workshop A and were asked to rank the top solutions for each problem statement from 1 (top priority) to 5 (lower priority). Ranking is regarded as a way to prioritize and reach an agreement [36]. To ensure priorities accurately reflected participants' lived experience, each problem statement was ranked by 2 groups; however, this meant that not every group ranked every problem statement. The workshop procedure is outlined in Table 3, and the solutions are presented in the "Results" section. At the conclusion of the workshop, participants were asked by facilitators to provide feedback on their experiences of the co-design process. Following workshop B, the lead researcher (JM) collated the rankings and removed any solutions that were not ranked.

Stage 3: Prototyping

Outcomes from the workshops were analyzed using content analysis and used to create prototypes, in line with the Design Thinking process. Conventional inductive content analysis, following 5 stages, was used to allow categories to emerge directly from the workshop outputs and to reduce the volume of information [37].

The lead researcher (JM) read through the workshop outputs and any associated field notes several times to become immersed in the data. This supported note-making on initial ideas in a reflexive journal, allowing consideration of connections, similarities, and differences within the data. This process also highlighted that participants generally lacked knowledge about smart speakers, were fearful of hackers, and wanted speech-accessible smart speakers. These insights challenged the lead researcher's confirmation bias, encouraging empathy with the experiences of PwPD and allowing the research to be shaped by user needs. This highlighted the importance of creating new features for smart speakers that better meet users' needs, as well as utilizing existing features. Reflexivity also allowed the lead researcher to reflect on her multiple roles as a SaLT, facilitator, and analyst, and the potential for these roles to introduce interpretation bias toward a clinical perspective. As a result, an audit trail was developed to demonstrate the analysis process and enhance trust in decision-making during analysis [38].

Following data immersion, the lead author created a mind map to inductively group ideas and develop initial codes for analysis. Initial codes were shared with OD, KP, and GK for discussion and refinement, enhancing credibility through investigator triangulation and peer debriefing. The final codes were both descriptive (eg, "privacy concern") and interpretative (eg, "need to increase motivation for speech practice"), and the meaning of each code was documented to ensure reliability during coding.

Subsequently, the workshop outputs were coded. This was conducted by hand, using graph paper and colored pens to assign meaning to each output.

Coding was conducted 3 times on separate days by 1 coder (JM) and was presented to the research team for discussion, redrafting, and agreement. It is acknowledged that coding by a single researcher may introduce bias, and, upon reflection, the involvement of 2 coders may have enabled data triangulation and enhanced data credibility. Despite this, peer debriefing helped to minimize potential impacts on the analysis of results. Similar codes were grouped into broader themes to capture meaning across the outputs. This process was also completed by hand, using colored pens to illustrate relationships between codes. Reflexive notes were recorded, discussed with coauthors (OD, KP, and GK), and refined accordingly. Finally, themes were conceptually mapped into 2 prototypes, in keeping with the Design Thinking Framework.

The content analysis process described above, from ranked ideas to the creation of prototypes, is available in Multimedia Appendix 2 as an audit trail, enhancing the credibility of the outputs [39]. All solutions that were ranked in workshop B were included in the prototypes to ensure that the prototypes reflected the wants and needs of participants. Furthermore, direct quotations, where available, from participant feedback are presented to provide a direct voice and to link outputs with interpretations [32]. Findings were sent to all participants for member checking to ensure that the written findings reflected their lived experiences and to enhance the rigor of the research.

Results

Study Participants

A total of 20 participants were recruited; 19 participated in co-design workshop A, and 16 in co-design workshop B. Overall, 15 participants took part in both workshops (Tables 4 and 5).

Table 4. Makeup of breakout rooms in workshop A.

Group number	Number of participants, n	Participants				
		People with Parkinson's disease	Carer	Speech and language therapist	Third sector	Technology/design
1	4	✓	✓	✓	N/A[a]	✓
2	3	✓	✓	✓	N/A	N/A
3	4	✓	✓	✓	✓	N/A
4	4	✓	✓	✓	N/A	✓
5	4	✓	N/A	✓	N/A	✓

[a]N/A: not applicable.

Table 5. Makeup of breakout rooms in workshop B.

Group number	Facilitator	Number of participants	Participants				
			People with Parkinson's disease	Carer	Speech and language therapist	Third sector	Technology/design
1	GK	4	✓	✓	✓	N/A[a]	✓
2	OD	4	✓	✓	✓	N/A	N/A
3	JM	4	✓	✓	✓	✓	N/A
4	KP	4	✓	N/A	✓	✓	✓

[a]N/A: not applicable.

The ranking of solutions aligned with the 6 problem statements is outlined in Multimedia Appendix 4. For problem statements 1, 2, 4, and 6, 3 solutions were ranked by both groups, and 4 solutions were ranked by 1 group. For problem statement 3, 2 solutions were ranked by both groups, and 5 solutions were ranked by 1 group. For problem statement 5, all 4 solutions were ranked by both groups, as only 4 options were presented. Solutions that were not given a rank by any group were removed from the results presented below. The full list of ideas available for ranking during workshop B is provided in Multimedia Appendix 1.

Stage 1: Prototyping Results

Rankings were collated into 2 main prototypes by the primary researcher and agreed upon by the team: (1) educational guidance on the therapeutic use of smart speakers, and (2) developing new SLT-specific features for smart speakers (Multimedia Appendices 5 and 6). These prototypes present the results outlined above, emphasizing cross-cutting themes. Participants' experiences of the co-design process are also presented. The process from solutions to prototypes is fully detailed in Multimedia Appendix 2.

Prototype 1: Education and Guidance

Guides for PwPD and SaLTs, detailing how to use smart speakers to improve volume, intelligibility, and clarity of speech, were unanimously agreed upon by participants. The contents of these guides are described in Multimedia Appendix 5. Participants highlighted a gap in knowledge between the traditionally available features of smart speakers and an understanding of how these features could be repurposed to benefit speech and voice in Parkinson's disease. The suggested skills catalog for therapy would create a repository of standard smart speaker features and skills that could be utilized with therapeutic intent by SaLTs and PwPD. Suggestions included integrating prompts and positive reinforcement by building routines, for example: "Could you speak louder?" or "Well done, great practice today." PwPD felt that verbal prompts to speak louder or clearer, along with positive reinforcement from smart speakers, would replicate cuing provided by SaLTs during direct therapy and motivate home practice.

Participants indicated that routines could be used to practice scripted conversations, and that these should be personalized, include prompts to help sustain conversations, and contain only personal information that users felt comfortable sharing with their smart speaker.

It was evident that not all participants were aware of these accessibility features, which are designed to maximize engagement with smart speakers, and that education in this area may help to encourage more natural conversational reciprocity. For example, the conversation mode available on Amazon Alexa devices.

Participants also indicated that education about privacy relating to smart speaker use was required. It was reported that education for both PwPD and SaLTs would help to alleviate fears regarding personal data storage and General Data Protection Regulation (GDPR) concerns.

Prototype 2: Developing New SLT-Specific Features for Smart Speakers

Participants indicated that an Alexa skill could be created to support speech therapy, as shown in Multimedia Appendix 6. Suggestions included delivering LSVT through a smart speaker or developing a speech therapy game to support speech and voice practice. Participants suggested that this could include increased feedback, such as visual cues on a screen for volume and speech clarity and live transcription of speech that repeats back what was heard, to support self-awareness in PwPD. It is acknowledged that newer Amazon Alexa models, such as the Echo Show 10, already offer subtitling features within the settings, which provide live captioning of speech or video calls.

Additionally, participants were excited about the potential for artificial intelligence (AI) integration within smart speakers and suggested that this could be used to enable more intelligent conversations with the device. Many participants indicated that current smart speakers lacked this capability. Although intelligent conversation has not yet been integrated as a core functionality across Amazon Alexa devices, skills such as ChatGPT were perceived to facilitate live, functional conversation. Furthermore, Alexa Plus, a paid feature for Amazon devices, uses generative AI to remember previous interactions and continue conversations over time. It also offers 5 personalities, which may help users feel as though they are conversing with a person rather than a device. However, Alexa Plus is not yet available in Northern Ireland, where this research was conducted. Additionally, the *follow-up* mode within Alexa accessibility settings prevents users from having to repeat the device wake word, which Amazon suggests supports a more conversational interaction with smart speakers.

Participants also indicated that extended listening time for smart speakers would prevent mid-sentence interruptions. It is acknowledged that Amazon Alexa devices currently offer an *adaptive listening* feature in the accessibility settings, which extends input time and accommodates speech differences. Although a few participants were aware of this feature, they did not indicate its impact on their smart speaker interactions.

Furthermore, enhanced privacy features were suggested. Again, it is understood that, under Alexa privacy settings, voice commands can be enabled to clear Alexa voice history; for example, "Alexa, delete everything I've ever said." Additionally, although Alexa cannot be trained to respond only to certain voices, there is an option to set up a voice profile to receive more personalized content and prevent unauthorized voice purchases.

Although participants acknowledged that adapting smart speakers to better recognize dysarthric speech could hamper their therapeutic value, they felt that this would improve accessibility for the devices more generally. They sought devices that could gradually learn their speech patterns over time, as well as deal effectively with regional accents. Notably, there is currently no research exploring the impact of improved speech recognition in smart speakers on therapy outcomes in SLT.

In addition to ranking solutions, participants were asked about their experience of co-design using questions based on the UK standards for public involvement. Overall, participants valued the online workshop format, which facilitated engagement for those with limited mobility. They felt the workshops were informal yet professional and found the tasks interesting, positively challenging them to think of solutions. Small groups were reportedly the right size for supported discussions, and participants felt this was an effective way to gather substantial information. Participants discussed their expectations and involvement in co-design, describing feeling included and respected:

> *I had some experience of delivering co-design, so I had an idea of how it should be done...I felt valued, and felt everyone has been really equally valued, no matter how you're coming at it; person with Parkinson's, speech therapist, whatever. We've all been treated equally, with respect.* [Person from Parkinson's UK]

> *I felt heard and respected throughout and you did a good job of facilitating conversations for us to feel heard.* [SaLT]

> *The carer and patient are heard. So often in NHS setting they are the last ones to be heard y'know, what*

> *would they know. But here, they were put front and centre.* [Carer]

Participants also felt that the right people were involved in the co-design process and that there was a good balance between perspectives:

> *It's involved so many stakeholders that come from that same place of making improvements for people living with Parkinson's. It was great to see various individuals are spoken to and included.* [PwPD]

> *There was a really good balance of people from different backgrounds...It absolutely worked and its so important to get everyone's view; it's mostly important to hear people with Parkinson's, carers you work alongside. You get a really holistic picture of what is the most important thing from different perspectives.* [SaLT]

> *It was useful to be able to discuss together in a group and helpful to consider all views: SaLTs, patients and tech experts.* [PwPD]

Participants provided feedback on engagement challenges and future improvements. Some PwPD or carers felt that a bridging workshop between creating and ranking solutions would be helpful. This could have included a session to discuss all brainstormed solutions and integrate them with real-world examples. Although elements of this were included in the workshops, they felt that a third workshop would have given them time to digest the large number of solutions and some more complex ideas before ranking them. One clinician who was unable to attend the first brainstorming workshop felt that this would have helped orient her more fully before the ranking task. Others suggested that color grouping or collapsing solutions for each problem statement by themes may have made it easier to rank statements. Participants also indicated that more prompts were required to remind them to think creatively and that "anything was possible."

Participants also discussed the project's focus on smart speakers, as well as their advantages and disadvantages. A few participants felt that it would be easier to create an app, as many are available for smartphones, and most people use these devices. However, most participants felt that the voice interaction of smart speakers offered advantages over smartphones, particularly for people with a tremor. Additionally, participants felt that smart speakers could remind and motivate users to practice, whereas with an app, users often have to self-motivate or remind themselves.

Discussion

Principal Findings

This study aimed to co-produce solutions to support smart speaker use for speech and voice difficulties and to inform a future intervention. PwPD, carers, SaLTs, Parkinson's UK staff, and technology and design experts collaborated during 2 online co-design workshops to brainstorm and prioritize solutions to problems identified in prior research [9,11]. Two prototypes were developed: (1) education and guidance on the therapeutic use of smart speakers and (2) the development of new speech therapy–specific features for smart speakers. By incorporating collaborators' priorities and needs, the study offers a foundation for a future smart speaker–based intervention for speech and voice therapy in PwPD.

Impact of Co-Design

This project recognizes the need to involve end users early and meaningfully when designing health care interventions [40,41], contributing to the quality and relevance of co-designed outcomes [42]. This aligns with the Design Thinking framework, specifically the ideate and prototyping phases. While co-production with people with aphasia is increasing, there is limited evidence on co-design in SLT, especially for motor speech disorders [21,32,43]. Therefore, this research continues to contribute to and develop the evidence base regarding co-design in SLT, particularly for people with dysarthria. This study is unique, as it is believed to be the first co-design study with PwPD who have speech and voice difficulties that co-designs solutions to problems experienced when using commercial VAT technology.

Participants described personal benefits of co-design, including gaining knowledge, social interaction, and feeling heard and validated, echoing previous co-production findings [32,44] and aligning with public involvement standards [45]. These benefits are particularly relevant for PwPD, who often experience reduced participation due to speech and voice issues [46], highlighting how co-design can empower participants. Power sharing and partnership can enhance engagement and lead to more patient-centered outcomes [47], and involving SaLTs may also improve future implementation of such tools into clinical practice [48]. In wider co-design research in SLT, participants with communication difficulties report improved confidence, motivation, and sense of well-being [27], and their involvement can lead to more and better-quality outcomes [43]. Overall, this demonstrates how co-production can allow participants, such as PwPD, to feel in control, empowered, and validated. For PwPD, this co-design study both physically and metaphorically provided them with a voice, building on current evidence. Despite this, wider research also acknowledges that relinquishing power in research can be challenging for researchers, requiring an active effort to make the co-design process truly collaborative [18].

Additionally, collaborator feedback highlighted the importance of skilled facilitation in enabling communication during workshops. Although evidence on co-design facilitation strategies for people with speech and voice difficulties is limited, facilitators used clinical experience and evidence-based strategies [32] to support PwPD. These included allowing preparation time before workshops, building rapport, giving extra time to speak, screen-sharing key points, regularly checking understanding, and summarizing discussions [43,49,50]. Such approaches are crucial for inclusive and accessible co-design. Some collaborators suggested improvements, such as offering more workshops and using multimedia formats to make tasks easier, which extends the evidence base on co-design with PwPD who have speech and voice difficulties. This balance of positive experiences and suggested improvements reflects the range of participants, lived experiences, and heterogeneous needs. Advantages and

disadvantages of co-design methods should be evaluated from a range of perspectives to achieve a balance between the needs of a diverse group of PwPD.

Participants indicated that training for SaLTs in the therapeutic use of smart speakers for speech and voice difficulties was a priority. Wider research supports this finding, showing that education and guidance are required to support therapeutic adoption by SaLTs and PwPD [4,9-11], and that digital health interventions for older adults should include education in effective device use, digital literacy skills, and technical support throughout [51,52]. Tailored education and guidance may contribute to PwPD and SaLTs successfully adopting and using smart speakers to support speech and voice difficulties. As such, this study begins to advance understanding of how to support VAT adoption into clinical SLT practice. While smart speaker features make them valuable tools for chronic health management among older adults [53,54], older people in particular can struggle to comprehend the full range of smart speaker functions [11,54].

Guidance should clearly link device features to SLT goals to promote understanding and demonstrate how devices can help people achieve their SLT practice and related goals [55,56]. This may positively impact digital literacy for PwPD, supporting device adoption and regular use [56], again contributing to advances in knowledge regarding the clinical adoption of VAT. Similarly, SaLTs in our earlier research made several content suggestions for guidance to empower them to use VAT [9], including sample therapy plans, scripts, goal-setting frameworks, and evidence-based practice. However, this is the first study to collate these elements into an education and guidance prototype for SaLTs. Simplified guidance is particularly important, as clinicians often discontinue technologies they perceive as overly complex for clients [57]. Similar requirements for implementation guidance have been reported in SLT research using commercial technologies, such as virtual reality (VR) [19,58,59]. These studies highlight that therapeutic usage guides should promote the ease of use and usefulness of commercial technologies to support clinical adoption and provide opportunities to trial the devices. However, it is important to note that although commercial VR technology was used, the VR program itself was specifically created by researchers. This suggests that guidance must explain how smart speakers' out-of-the-box "Alexa skills" are relevant to SLT, given that the commercial use of the technology is not intended to be therapeutic. Unlike custom VR programs, smart speakers are off-the-shelf products not originally designed for health care. Therefore, guidance must explicitly link commercial features to therapeutic aims and support clinicians in adapting features to individual client needs, ultimately contributing to the adoption of VAT into clinical practice.

Furthermore, privacy and data protection are significant barriers to the adoption of smart speakers [60]. Common concerns include the recording of conversations and data misuse, which can deter both clinicians and clients [51,61-63]. To address this, usage guides for PwPD should include clear, accessible privacy information, support informed consent, and clearly explain how devices handle user data [64]. Given SaLTs' responsibility for safeguarding client data, guidance should map VAT's GDPR compliance and potential risks to SLT governance policies, such as Data Protection Impact Assessments. This study, therefore, begins to answer questions posed by previous research [10] regarding how VAT may be implemented in accordance with clinical governance requirements. Previous findings highlight that many SaLTs lack clarity on which technologies meet governance and GDPR standards [57]. Reassuring both clinicians and PwPD about privacy may improve confidence and facilitate adoption [65]. Future evaluation of guidance acceptability and usability could apply frameworks such as the Technology Acceptance Model or Unified Theory of Acceptance and Use of Technology 2.

Delivery

Participants suggested delivering training through Royal College of Speech and Language Therapy–led webinars, live demonstrations, and group sessions led by trained SaLTs. While previous research has not identified optimal delivery formats [4,10], this study provides new insights into practical implementation and contributes to the evidence base regarding the therapeutic use of VAT in SLT clinical practice. The literature indicates that older adults often prefer hands-on, task-based learning supported by written instructions [66,67]. A training program using VAT as a tool for activities of daily living with adults with cognitive communication disorders indicated a need for written, easy-to-follow instructions, with hands-on support to overcome low technological literacy [67]. Group-based workshops can offer a supportive, low-pressure environment for exploration and skill-building with in-person support [52]. This is particularly important for users with limited experience or confidence in digital tools.

Findings indicate that SaLTs are central to introducing and supporting smart speaker use in therapy. When clinicians demonstrate relevance and ease of use, PwPD may be more likely to adopt the technology [67]. By increasing perceived usefulness and reducing concerns, training can enhance performance expectancy and digital engagement. Additionally, previous research on integrating commercial technologies in SLT has highlighted the importance of multifaceted training approaches, including device trials, workshops, clinical manuals, and information technology support [58,68]. Additional methods, such as guided observation and co-delivered interventions, may be necessary to bridge the gap between knowledge and practice [58,69]. As such, this study begins to address gaps in knowledge regarding the implementation of VAT as a therapeutic tool for speech and voice difficulties associated with Parkinson's disease.

Participants highlighted the need to develop SLT-specific features for smart speakers, designed for therapeutic use. For example, Cassano et al [70] described a SaLT building a custom skill. At the time of publication, at least three speech therapy Alexa skills existed: Speech Therapy Practice, Speech Device Practice, and Let's Talk. Additionally, 2 further speech therapy skills were identified but are no longer publicly available on the Amazon Skills store: Speech Doctor, as discussed by Makin et al [71], and Speech Therapy by Cathal Killeen. Notably, Speech Therapy Practice is a live Alexa skill developed by a SaLT that enables people with aphasia to practice very basic

words and phrases, such as colors, opposites, who/what questions, and yes/no questions. While this may potentially act as a starting point for SaLTs, the skill lacks applicability to practicing phrases and sentences and, in its current state, is unlikely to meet the speech practice needs of PwPD. To date, there are no specific Alexa skills for adults with Parkinson's or targeting dysarthria, and our research highlights the potential for future development. Future research may seek to work with developers to create an Alexa skill for this population that can be used to support home practice of speech therapy exercises. Features may include prompts for loud, clear speech; increased feedback on volume and intelligibility with suggestions for improvement; the ability to monitor progress; visual displays and biofeedback; reminders to complete therapy tasks; and LSVT-style exercises with gamification [9,11]. Such features align with wider studies integrating technology into SLT and related areas, including apps using Google Glass [72], smart speaker–based physical activity interventions [73,74], and social engagement tools for people with disabilities [75]. Development platforms like Alexa Skills Kit and Alexa Blueprint may offer scalable, cost-effective options, enabling a focus on increasing motivation, engagement, and potential adherence to intervention programs. A curated hub of Alexa skills that can be used for SLT goals may also support clinical implementation. For example, Esquivel et al [76] developed a repository of Alexa skills and recommendations for people with disabilities, by people with disabilities. Future research may explore the acceptability of a speech therapy–specific Alexa skill and its implementation within clinical practice.

Given the commercial nature of smart speakers, it may be beneficial to first assess their current therapeutic value before creating bespoke skills. As our study focused on co-design processes and did not include a formal evaluation of intervention usability, effectiveness, or acceptability, future research may consider testing the current prototypes to determine real-world clinical impact and user outcomes. This study establishes the rationale for a future feasibility study to examine the effectiveness of VAT as a therapeutic tool for speech and voice difficulties in Parkinson's disease. At the time of writing, no studies have been conducted in this area using commercial VAT. Emerging SLT research shows benefits for speech clarity in populations with intellectual disabilities and speech sound disorders [5,71], citing immediate rewards, spaced practice, enhanced autonomy, intrinsic motivation, and reduced social barriers as mechanisms of change in speech. However, these interventions do not follow established SLT intervention protocols [5,71]. Therefore, future studies should evaluate the effectiveness and usability of smart speakers for PwPD using principles of neuroplasticity and motor learning from SLT protocols, such as LSVT LOUD or Speak Out!

Despite this, the challenges surrounding the therapeutic use of smart speakers cannot be ignored. Smart speakers rely on evolving ASR models, a type of AI, which are continually being improved. ASR models can change without warning, presenting a risk to the reliability of baseline measurements and the measurement of therapy goals [77]. Furthermore, ASR errors are often higher than expected for dysarthric speech, speakers of minority languages, and those with regional accents [78-80]. Without clear and specific feedback on device or speech errors, both PwPD and SaLTs are left without information about where the "error" lies, whether it is speech- or device-related. These risks may reinforce maladaptive speech behaviors if speech practice is based on inconsistent or misleading responses from the device. It may also damage client motivation and confidence, with PwPD blaming themselves for technological errors. Research demonstrates that speakers can attribute ASR errors to themselves and link this to their sense of identity, including racial, regional, and locational identity [77]. To mitigate this lack of transparency, it is essential that SaLTs educate potential VAT users on strategies for adapting speech, managing frustration, and correctly interpreting VAT errors, as well as raising awareness of the limited ASR training on dysarthric speech and some minority or foreign languages [9,11]. This highlights the importance of a therapeutic usage manual for smart speakers for people with speech and voice difficulties and for SaLTs, as indicated in the current findings.

However, it should be noted that projects such as Voiceitt, Google Euphonia, and Project Relate aim to improve ASR accuracy in recognizing dysarthric speech, and Accessible Voice Interaction Technology for Aphasia (AVITA) aims to improve the accessibility of smart speakers [81], which may have the unintended consequence of limiting certain therapeutic applications of smart speakers in SLT. When smart speaker recognition is improved, speech difficulties no longer affect recognition, meaning all speech is easily recognized. This can be problematic, as speech that may not be intelligible in real life is recognized by devices. Consequently, this hampers therapeutic applications, because positive biofeedback provided by smart speakers does not reflect the speaker's intelligibility to unfamiliar listeners in everyday contexts. Indeed, participants in this research indicated that future adaptations of smart speakers, outside of therapeutic contexts, should aim to better recognize dysarthric speech and regional accents. Future smart speaker designs may bridge the gap between standard out-of-the-box devices and fully customized skills. For example, smart speakers could allow users to set recognition thresholds, enabling both increased accessibility for users and therapeutic usage for clinicians. Given the rapid pace of innovation, continued review of emerging literature and technologies is recommended throughout the development and implementation stages.

Limitations

This co-design study offered valuable insights into developing VAT tools for PwPD with speech and voice difficulties; however, limitations are evident.

Participants suggested an additional workshop between the ideation and prioritization phases, that is, between workshops A and B. A bridging session could have allowed more reflection and improved understanding, potentially leading to rankings that more accurately reflected lived experience. Furthermore, although recruitment was successful, there was some participant dropout between workshops A and B. This necessitated merging groups in workshop B, which may have influenced group dynamics and limited continuity of discussion.

Despite efforts to recruit a diverse group, the sample was small (n=20), and certain perspectives, such as those of people with advanced Parkinson's or severe dysarthria, were underrepresented. This may limit the generalizability of the findings.

Finally, given the rapidly evolving technology landscape in AI and ASR, some recommendations may become outdated by the time of implementation. This includes changes in smart speaker capabilities, privacy policies, and integration with large language models (eg, AI conversational agents).

Conclusions

This study highlights the value of co-designing smart speaker interventions with PwPD, carers, SaLTs, third sector representatives and technology and design experts to address challenges in using VAT for speech therapy. Using a participatory Design Thinking approach, user-centered solutions were generated to improve the accessibility, usability, and therapeutic potential of smart speakers.

Two prototypes were developed: (1) education and guidance for PwPD and SaLTs, and (2) speech therapy–specific smart speaker features.

The outputs balance commercial technology with clinical needs, focusing on privacy, troubleshooting, and feedback for home use, while reinforcing co-design as a powerful method for developing digital health tools. Co-design also ensured that interventions reflected lived experience and clinical insight, enhancing the likelihood of adoption and sustained use. This research strengthens the evidence for co-design in SLT and supports smart speakers as tools to enhance therapy access, promote self-management, and reduce pressure on SLT services. Future work should develop and evaluate these prototypes to assess their real-world impact and scalability.

Acknowledgments

We thank Rory Bradley for his contributions to the co-design workshops in this research. The authors would also like to thank Donna McGuckin, Emmet Leyden, and Dr Marc Parker for their contribution to this research as members of the patient and public involvement (PPI) group.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Results condensed from workshop A.
[DOCX File , 22 KB - rehab_v13i1e84364_app1.docx]

Multimedia Appendix 2
Ranking to prototype creation.
[DOCX File , 28 KB - rehab_v13i1e84364_app2.docx]

Multimedia Appendix 3
Facilitator guidance.
[DOCX File , 21 KB - rehab_v13i1e84364_app3.docx]

Multimedia Appendix 4
Ranking results.
[DOCX File , 21 KB - rehab_v13i1e84364_app4.docx]

Multimedia Appendix 5
Prototype 1: education and guidance on the therapeutic use of smart speakers.
[DOCX File , 17 KB - rehab_v13i1e84364_app5.docx]

Multimedia Appendix 6
Prototype 2: developing new speech therapy–specific features for smart speakers.
[DOCX File , 16 KB - rehab_v13i1e84364_app6.docx]

References

1. The Infinite Dial UK. Edison Research. 2025. URL: https://www.edisonresearch.com/wp-content/uploads/2025/05/Infinite-Dial-UK-2025-webinar-presentation-for-download.pdf [accessed 2025-09-01]
2. Pradhan A, Lazar A, Findlater L. Use of intelligent voice assistants by older adults with low technology use. ACM Trans Comput-Hum Interact 2020 Sep 16;27(4):1-27. [doi: 10.1145/3373759]

3. Cave R, Bloch S. The use of speech recognition technology by people living with amyotrophic lateral sclerosis: a scoping review. Disabil Rehabil Assist Technol 2023 Oct;18(7):1043-1055 [FREE Full text] [doi: 10.1080/17483107.2021.1974961] [Medline: 34511007]
4. Duffy O, Synnott J, McNaney R, Brito Zambrano P, Kernohan WG. Attitudes toward the use of voice-assisted technologies among people with Parkinson disease: findings from a web-based survey. JMIR Rehabil Assist Technol 2021 Mar 11;8(1):e23006 [FREE Full text] [doi: 10.2196/23006] [Medline: 33704072]
5. Smith E, Sumner P, Hedge C, Powell G. Smart speaker devices can improve speech intelligibility in adults with intellectual disability. Int J Lang Commun Disord 2021 May;56(3):583-593. [doi: 10.1111/1460-6984.12615] [Medline: 33772998]
6. Bleakley A, Rough D, Roper A, Lindsay S, Porcheron M, Lee M, et al. Exploring smart speaker user experience for people who stammer. 2022 Presented at: ASSETS '22: The 24th International ACM SIGACCESS Conference on Computers and Accessibility; October 23-26, 2022; Athens, Greece p. 23-26. [doi: 10.1145/3517428.3544823]
7. Miller N. Communication changes in Parkinson's disease. Pract Neurol 2017 Aug;17(4):266-274. [doi: 10.1136/practneurol-2017-001635] [Medline: 28687681]
8. Miller N, Deane KHO, Jones D, Noble E, Gibb C. National survey of speech and language therapy provision for people with Parkinson's disease in the United Kingdom: therapists' practices. Int J Lang Commun Disord 2011;46(2):189-201. [doi: 10.3109/13682822.2010.484849] [Medline: 21401817]
9. Mills J, Duffy O, Pedlow K, Kernohan G. Exploring speech and language therapists' perspectives of voice-assisted technology as a tool for dysarthria: qualitative study. JMIR Rehabil Assist Technol 2025 Sep 02;12:e75044 [FREE Full text] [doi: 10.2196/75044] [Medline: 40896823]
10. Kulkarni P, Duffy O, Synnott J, Kernohan WG, McNaney R. Speech and language practitioners' experiences of commercially available voice-assisted technology: web-based survey study. JMIR Rehabil Assist Technol 2022 Jan 05;9(1):e29249 [FREE Full text] [doi: 10.2196/29249] [Medline: 34989694]
11. Mills J, Duffy O, Pedlow K, Kernohan G. Exploring the perceptions of voice-assisted technology as a tool for speech and voice difficulties: focus group study among people with Parkinson disease and their carers. JMIR Rehabil Assist Technol 2025 Jul 16;12:e75316 [FREE Full text] [doi: 10.2196/75316] [Medline: 40669073]
12. Ramig L, Halpern A, Spielman J, Fox C, Freeman K. Speech treatment in Parkinson's disease: randomized controlled trial (RCT). Mov Disord 2018 Nov;33(11):1777-1791 [FREE Full text] [doi: 10.1002/mds.27460] [Medline: 30264896]
13. Altman M, Huang TT, Breland JY. Design thinking in health care. Prev Chronic Dis 2018 Sep 27;15:E117 [FREE Full text] [doi: 10.5888/pcd15.180128] [Medline: 30264690]
14. Sumner J, Chong LS, Bundele A, Wei Lim Y. Co-designing technology for aging in place: a systematic review. Gerontologist 2021 Sep 13;61(7):e395-e409 [FREE Full text] [doi: 10.1093/geront/gnaa064] [Medline: 32506136]
15. Brown T. Design and thinking. Harvard Educational Review 2013 Apr 01;86(1):84-92. [doi: 10.17763/1943-5045-83.1.243]
16. Vargas C, Whelan J, Brimblecombe J, Allender S. Co-creation, co-design, co-production for public health - a perspective on definition and distinctions. Public Health Res Pract 2022 Jun 15;32(2):3222211 [FREE Full text] [doi: 10.17061/phrp3222211] [Medline: 35702744]
17. Benjamin-Thomas TE, Corrado AM, McGrath C, Rudman DL, Hand C. Working towards the promise of participatory action research: learning from ageing research exemplars. International Journal of Qualitative Methods 2018 Dec 26;17(1):e-1609406918817953. [doi: 10.1177/1609406918817953]
18. Roper A, Skeat J. Innovation through participatory design: collaborative qualitative methods in the development of speech-language pathology technology. Int J Speech Lang Pathol 2022 Oct;24(5):527-532. [doi: 10.1080/17549507.2022.2050943] [Medline: 35506478]
19. Brassel S, Brunner M, Power E, Campbell A, Togher L. Speech-language pathologists' views of using virtual reality for managing cognitive-communication disorders following traumatic brain injury. Am J Speech Lang Pathol 2023 Mar 23;32(2S):907-923. [doi: 10.1044/2022_ajslp-22-00077]
20. Wilson S, Roper A, Marshall J, Galliers J, Devane N, Booth T, et al. Codesign for people with aphasia through tangible design languages. CoDesign 2015 Jan 09;11(1):21-34. [doi: 10.1080/15710882.2014.997744]
21. Cruice M, Aujla S, Bannister J, Botting N, Boyle M, Charles N, et al. Creating a novel approach to discourse treatment through coproduction with people with aphasia and speech and language therapists. Aphasiology 2021 Jul 03;36(10):1159-1181. [doi: 10.1080/02687038.2021.1942775]
22. Shiggins C, Coe D, Gilbert L, Research Collaboration A, Mares K. Development of an “aphasia-accessible participant in research experience survey” through co-production. Aphasiology 2022 Jan 31;39(12):1659-1692. [doi: 10.1080/02687038.2021.1996532]
23. Arnstein S. A Ladder Of Citizen Participation. Journal of the American Institute of Planners 1969 Jul;35(4):216-224 [FREE Full text] [doi: 10.1080/01944366908977225]
24. Elwyn G, Nelson E, Hager A, Price A. Coproduction: when users define quality. BMJ Qual Saf 2020 Sep;29(9):711-716 [FREE Full text] [doi: 10.1136/bmjqs-2019-009830] [Medline: 31488570]
25. Penny J, Slay J. Commissioning for outcomes and co-production: a practical guide for local authorities. New Economics Foundation. London, UK: New Economics Foundation; 2014. URL: https://neweconomics.org/2014/06/commissioning-outcomes-co-production [accessed 2026-01-25]

26. Connery A, Salsberg J. Exploring participatory health research and its application to speech and language therapy research practices. Int J Lang Commun Disord 2024;59(4):1257-1268. [doi: 10.1111/1460-6984.12994] [Medline: 38130139]
27. McMenamin R, Griffin M, Grzybowska B, Pound C. Working together: experiences of people with aphasia as co-researchers in participatory health research studies. Aphasiology 2021 Jun 25;39(12):1601-1622. [doi: 10.1080/02687038.2021.1923948]
28. Cook T, Boote J, Buckley N, Vougioukalou S, Wright M. Accessing participatory research impact and legacy: developing the evidence base for participatory approaches in health research. Educational Action Research 2017 Jul 31;25(4):473-488. [doi: 10.1080/09650792.2017.1326964]
29. Roper A, Wilson S, Neate T, Marshall J. Speech and language. In: Yesilada Y, Harper S, editors. A Foundation for Research. London, UK: Springer; 2019:121-131.
30. McNaney R, Tsekleves E, Synnott J. Future opportunities for IoT to support people with Parkinson's. 2020 Presented at: CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, HI p. 1-15 URL: https://dl.acm.org/doi/10.1145/3313831.3376871 [doi: 10.1145/3313831.3376871]
31. Mc Menamin R, Isaksen J, Manning M, Tierney E. Distinctions and blurred boundaries between qualitative approaches and public and patient involvement (PPI) in research. Int J Speech Lang Pathol 2022 Oct 28;24(5):515-526 [FREE Full text] [doi: 10.1080/17549507.2022.2075465] [Medline: 35762365]
32. Clay P, Walton T, Malin E, Hutchinson M, Levitt K, Williams C, et al. Better conversations with Parkinson's: co-production of a novel speech and language therapy intervention with people living with Parkinson's. Research for All 2024;8(1):1-17. [doi: 10.14324/RFA.08.1.07]
33. Beckman SL. To frame or reframe: where might design thinking research go next? California Management Review 2020 Mar 02;62(2):144-162. [doi: 10.1177/0008125620906620]
34. Mah H, Dobson R, Thomson A. The importance of lived experience: a scoping review on the value of patient and public involvement in health research. Health Expect 2025 Apr 26;28(2):e70205-e70205 [FREE Full text] [doi: 10.1111/hex.70205] [Medline: 40135803]
35. Archibald MM, Ambagtsheer RC, Casey MG, Lawless M. Using Zoom videoconferencing for qualitative data collection: perceptions and experiences of researchers and participants. International Journal of Qualitative Methods 2019 Sep 11;18(1):1-8. [doi: 10.1177/1609406919874596]
36. Larkman CS, Lanyon L, Rose ML. Co-designing solutions to the challenges speech pathologists and interpreters encounter when they collaborate to deliver aphasia therapy. Disabil Rehabil 2025 Dec;47(24):6468-6482 [FREE Full text] [doi: 10.1080/09638288.2025.2486459] [Medline: 40188384]
37. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
38. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
39. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
40. O'Cathain A, Croot L, Duncan E, Rousseau N, Sworn K, Turner KM, et al. Guidance on how to develop complex interventions to improve health and healthcare. BMJ Open 2019 Aug 15;9(8):e029954 [FREE Full text] [doi: 10.1136/bmjopen-2019-029954] [Medline: 31420394]
41. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
42. National Institute for Health and Care Research (NIHR). Briefing notes for researchers in public involvement in NHS, health and social care research. NIHR. 2021. URL: https://www.nihr.ac.uk/documents/briefing-notes-for [accessed 2025-05-26]
43. Jayes M, Moulam L, Meredith S, Whittle H, Lynch Y, Goldbart J, et al. Making public involvement in research more inclusive of people with complex speech and motor disorders: the I-ASC project. Qual Health Res 2021 Jun 28;31(7):1260-1274 [FREE Full text] [doi: 10.1177/1049732321994791] [Medline: 33645331]
44. Bird M, Ouellette C, Whitmore C, Li L, Nair K, McGillion MH, et al. Preparing for patient partnership: a scoping review of patient partner engagement and evaluation in research. Health Expect 2020 Jun 10;23(3):523-539 [FREE Full text] [doi: 10.1111/hex.13040] [Medline: 32157777]
45. UK standards for public involvement. UK Public Involvement Standards Development Partnership/National Institute for Health and Care Research (NIHR). URL: https://sites.google.com/nihr.ac.uk/pi-standards/standards [accessed 2025-05-26]
46. Gillivan-Murphy P, Miller N, Carding P. Voice treatment in Parkinson's disease: patient perspectives. JPRLS 2019 Jul;Volume 9(1):29-42. [doi: 10.2147/jprls.s180183]
47. Carman KL, Dardess P, Maurer M, Sofaer S, Adams K, Bechtel C, et al. Patient and family engagement: a framework for understanding the elements and developing interventions and policies. Health Aff (Millwood) 2013 Feb;32(2):223-231. [doi: 10.1377/hlthaff.2012.1133] [Medline: 23381514]

48. Singer I, Klatte IS, de Vries R, van der Lugt R, Gerrits E. Using co-design to develop a tool for shared goal-setting with parents in speech and language therapy. Int J Lang Commun Disord 2022 Nov 20;57(6):1281-1303 [FREE Full text] [doi: 10.1111/1460-6984.12753] [Medline: 35859264]
49. Burton A, Ogden M, Cooper C. Planning and enabling meaningful patient and public involvement in dementia research. Curr Opin Psychiatry 2019 Nov;32(6):557-562. [doi: 10.1097/YCO.0000000000000548] [Medline: 31306247]
50. Swinburn K. Working alongside people with stroke and aphasia. In: Volkmer A, Broomfield K, editors. Seldom Heard Voices in Service User Involvement: The How and Why of Meaningful Collaboration. Havant, UK: J&R Press; 2022:129-150.
51. McCloud R, Perez C, Bekalu MA, Viswanath K. Using smart speaker technology for health and well-being in an older adult population: pre-post feasibility study. JMIR Aging 2022 May 09;5(2):e33498 [FREE Full text] [doi: 10.2196/33498] [Medline: 35532979]
52. Masters K, Correia R, Nemethy K, Benjamin J, Carver T, MacNeill H. Online learning in health professions education. Part 2: Tools and practical application: AMEE Guide No. 163. Medical Teacher 2023 Sep 23;46(1):18-33. [doi: 10.1080/0142159x.2023.2259069]
53. Kim S, Choudhury A. Exploring older adults' perception and use of smart speaker-based voice assistants: a longitudinal study. Computers in Human Behavior 2021 Nov 18;124(3):106914-106924. [doi: 10.1016/j.chb.2021.106914]
54. Chang F, Sheng L, Gu Z. Investigating the integration and the long-term use of smart speakers in older adults' daily practices: qualitative study. JMIR Mhealth Uhealth 2024 Feb 12;12:e47472 [FREE Full text] [doi: 10.2196/47472] [Medline: 38345844]
55. Arnold A, Kolody S, Comeau A, Miguel Cruz A. What does the literature say about the use of personal voice assistants in older adults? A scoping review. Disabil Rehabil Assist Technol 2024 Jan;19(1):100-111. [doi: 10.1080/17483107.2022.2065369] [Medline: 35459429]
56. Portz JD, Bayliss EA, Bull S, Boxer RS, Bekelman DB, Gleason K, et al. Using the technology acceptance model to explore user experience, intent to use, and use behavior of a patient portal among older adults with multiple chronic conditions: descriptive qualitative study. J Med Internet Res 2019 Apr 08;21(4):e11604 [FREE Full text] [doi: 10.2196/11604] [Medline: 30958272]
57. Tazi F, Nandakumar A, Dykstra J, Rajivan P, Das S. Privacy, security, and usability tradeoffs of telehealth from practitioners' perspectives. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2023 Oct 25;67(1):1862-1867. [doi: 10.1177/21695067231199689]
58. Vaezipour A, Aldridge D, Koenig S, Theodoros D, Russell T. A mixed-method investigation of clinician acceptance, barriers and enablers of virtual reality technology in communication rehabilitation. Disabil Rehabil 2022 Jul 14;44(15):3946-3958. [doi: 10.1080/09638288.2021.1895333] [Medline: 33715566]
59. Mills J, Duffy O. Speech and language therapists' perspectives of virtual reality as a clinical tool for autism: cross-sectional survey. JMIR Rehabil Assist Technol 2025 Feb 27;12:e63235-e63235 [FREE Full text] [doi: 10.2196/63235] [Medline: 40014826]
60. Maccario G, Naldi M. Alexa, is my data safe? The (ir)relevance of privacy in smart speakers reviews. International Journal of Human–Computer Interaction 2022 Apr 20;39(6):1244-1256. [doi: 10.1080/10447318.2022.2058780]
61. Lustgarten SD, Garrison YL, Sinnard MT, Flynn AW. Digital privacy in mental healthcare: current issues and recommendations for technology use. Curr Opin Psychol 2020 Dec;36:25-31 [FREE Full text] [doi: 10.1016/j.copsyc.2020.03.012] [Medline: 32361651]
62. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220 [FREE Full text] [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
63. Quinn K, Leiser Ransom S, O'Connell C, Muramatsu N, Marquez DX, Chin J. Assessing the feasibility and acceptability of smart speakers in behavioral intervention research with older adults: mixed methods study. J Med Internet Res 2024 Aug 30;26:e54800 [FREE Full text] [doi: 10.2196/54800] [Medline: 39213034]
64. Zhang R, Li H, Chen A, Liu Z, Lee Y. AI privacy in context: a comparative study of public and institutional discourse on conversational AI privacy in the US and Chinese social media. Social Media + Society 2024 Oct 28;10(4):1-19. [doi: 10.1177/20563051241290845]
65. Saripalle R, Patel R. From command to care: a scoping review on utilization of smart speakers by patients and providers. Mayo Clin Proc Digit Health 2024 Jun;2(2):207-220 [FREE Full text] [doi: 10.1016/j.mcpdig.2024.03.002] [Medline: 40207172]
66. Mitzner TL, Fausset CB, Boron JB, Adams AE, Dijkstra K, Lee CC, et al. Older adults' training preferences for learning to use technology. Proc Hum Factors Ergon Soc Annu Meet 2008 Sep;52(26):2047-2051 [FREE Full text] [doi: 10.1177/154193120805202603] [Medline: 25309139]
67. Shi Z, Du X, Li J, Hou R, Sun J, Marohabutr T. Factors influencing digital health literacy among older adults: a scoping review. Front Public Health 2024;12:1447747 [FREE Full text] [doi: 10.3389/fpubh.2024.1447747] [Medline: 39555039]
68. Glegg SMN, Levac DE. Barriers, facilitators and interventions to support virtual reality implementation in rehabilitation: a scoping review. PM R 2018 Nov;10(11):1237-1251.e1 [FREE Full text] [doi: 10.1016/j.pmrj.2018.07.004] [Medline: 30503231]

69. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
70. Cassano F, Pagano A, Piccinno A. Supporting speech therapies at (smart) home through voice assistance. 2021 Presented at: International Symposium on Ambient Intelligence; October 6-8, 2021; Salamanca, Spain. [doi: 10.1007/978-3-031-06894-2_10]
71. Makin L, Smith E, Hedge C, Sumner P, Powell G. Smart speakers are an acceptable and feasible speech practice tool for children with speech difficulties. Disabil Rehabil Assist Technol 2025 Oct 17;20(7):2510-2521 [FREE Full text] [doi: 10.1080/17483107.2025.2491636] [Medline: 40244154]
72. McNaney R, Vines J, Roggen D, Balaam M, Zhang P, Poliakov I, et al. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Apr 26 Presented at: CHI '14: The SIGCHI Conference on Human Factors in Computing Systems; May 1, 2014; Toronto, Canada p. 2551-2554. [doi: 10.1145/2556288.2557092]
73. Jansons P, Fyfe J, Via JD, Daly RM, Gvozdenko E, Scott D. Barriers and enablers for older adults participating in a home-based pragmatic exercise program delivered and monitored by Amazon Alexa: a qualitative study. BMC Geriatr 2022 Mar 25;22(1):248-258 [FREE Full text] [doi: 10.1186/s12877-022-02963-2] [Medline: 35337284]
74. Carlin A, Logue C, Flynn J, Murphy MH, Gallagher AM. Development and feasibility of a family-based health behavior intervention using intelligent personal assistants: randomized controlled trial. JMIR Form Res 2021 Jan 28;5(1):e17501 [FREE Full text] [doi: 10.2196/17501] [Medline: 33507155]
75. Greuter S, Balandin S, Watson J. Social games are fun: exploring social interactions on smart speaker platforms for people with disabilities. In: CHI PLAY '19 Extended Abstracts: Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts. 2019 Presented at: CHI PLAY '19: The Annual Symposium on Computer-Human Interaction in Play; October 22-25, 2019; Barcelona, Spain p. 429-435. [doi: 10.1145/3341215.3356308]
76. Esquivel P, Gill K, Chung C, Ding D, Faieta J. Smart speakers and skill use: what do we know? Disabil Rehabil Assist Technol 2025 Feb 08;20(2):488-497. [doi: 10.1080/17483107.2024.2387801] [Medline: 39115270]
77. Cave R. How people living with amyotrophic lateral sclerosis use personalized automatic speech recognition technology to support communication. J Speech Lang Hear Res 2024 Nov 07;67(11):4186-4202. [doi: 10.1044/2024_jslhr-24-00097]
78. Mengesha Z, Heldreth C, Lahav M, Sublewski J, Tuennerman E. "I don't think these devices are very culturally sensitive."—Impact of automated speech recognition errors on African Americans. Front Artif Intell 2021;4:725911 [FREE Full text] [doi: 10.3389/frai.2021.725911] [Medline: 34901836]
79. Green J, MacDonald RL, Jiang PP, Cattiau J, Heywood R, Cave R. Automatic speech recognition of disordered speech: personalized models outperforming human listeners on short phrases. 2021 Presented at: The 22nd Annual Conference of the International Speech Communication Association; August 30 to September 3, 2021; Brno, Czechia p. 4778-4782 URL: https://www.isca-archive.org/interspeech_2021/green21_interspeech.pdf [doi: 10.21437/interspeech.2021-1384]
80. Tobin J, Nelson P, MacDonald B, Heywood R, Cave R, Seaver K, et al. Automatic speech recognition of conversational speech in individuals with disordered speech. J Speech Lang Hear Res 2024 Nov 07;67(11):4176-4185. [doi: 10.1044/2024_jslhr-24-00045]
81. Improving access to Alexa for people with aphasia. City of St George's, University of London. 2025. URL: https://tinyurl.com/45fcp8k6 [accessed 2025-09-10]

Abbreviations

AI: artificial intelligence
ASR: automatic speech recognition
GDPR: General Data Protection Regulation
LSVT: Lee Silverman Voice Treatment
PPI: patient and public involvement
PwPD: people with Parkinson’s disease
SaLT: speech and language therapist
SLT: speech and language therapy
VAT: voice-assisted technology
VR: virtual reality

Edited by S Munce; submitted 22.Sep.2025; peer-reviewed by D Singhal, Y Lin; comments to author 04.Nov.2025; revised version received 18.Dec.2025; accepted 13.Jan.2026; published 04.Feb.2026.

Please cite as:
Mills J, Kernohan G, Pedlow K, Duffy O
Voice-Assisted Technology for People With Parkinson's Disease Experiencing Speech and Voice Difficulties: Co-Designing Solutions Using Design Thinking
JMIR Rehabil Assist Technol 2026;13:e84364
URL: https://rehab.jmir.org/2026/1/e84364
doi:10.2196/84364
PMID:

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org