




Article

Applying AI to Safely and Effectively Scale Care to Address Chronic MSK Conditions

Anabela C. Areias ¹, Dora Janela ¹, Robert G. Moulder ^{1,2}, Maria Molinos ¹, Virgílio Bento ¹, Carolina Moreira ^{1,3}, Vijay Yanamadala ^{1,4,5}, Fernando Dias Correia ^{1,6} and Fabíola Costa ^{1,*}

¹ Sword Health, Inc., Draper, UT 84043, USA; a.areias@swordhealth.com (A.C.A.); d.janela@swordhealth.com (D.J.); robertgm111@gmail.com (R.G.M.); mmolinos@swordhealth.com (M.M.); vbento@swordhealth.com (V.B.); c.moreira@swordhealth.com (C.M.); v.yanamadala@swordhealth.com (V.Y.); fcorreia@swordhealth.com (F.D.C.)

² Institute for Cognitive Science, University of Colorado Boulder, Boulder, CO 80309, USA

³ Instituto de Ciências Biomédicas Abel Salazar, 4050-313 Porto, Portugal

⁴ Department of Surgery, Quinnipiac University Frank H. Netter School of Medicine, Hamden, CT 06473, USA

⁵ Department of Neurosurgery, Hartford Healthcare Medical Group, Westport, CT 06103, USA

⁶ Neurology Department, Centro Hospitalar e Universitário do Porto, 4099-001 Porto, Portugal

* Correspondence: f.costa@swordhealth.com; Tel.: +1-385-308-8034

Abstract: Background/Objectives: The rising prevalence of musculoskeletal (MSK) conditions has not been balanced by a sufficient increase in healthcare providers. Scalability challenges are being addressed through the use of artificial intelligence (AI) in some healthcare sectors, with this showing potential to also improve MSK care. Digital care programs (DCP) generate automatically collected data, thus making them ideal candidates for AI implementation into workflows, with the potential to unlock care scalability. In this study, we aimed to assess the impact of scaling care through AI in patient outcomes, engagement, satisfaction, and adverse events. **Methods:** Post hoc analysis of a prospective, pre-post cohort study assessing the impact on outcomes after a 2.3-fold increase in PT-to-patient ratio, supported by the implementation of a machine learning-based tool to assist physical therapists (PTs) in patient care management. The intervention group (IG) consisted of a DCP supported by an AI tool, while the comparison group (CG) consisted of the DCP alone. The primary outcome concerned the pain response rate (reaching a minimal clinically important change of 30%). Other outcomes included mental health, program engagement, satisfaction, and the adverse event rate. **Results:** Similar improvements in pain response were observed, regardless of the group (response rate: 64% vs. 63%; $p = 0.399$). Equivalent recoveries were also reported in mental health outcomes, specifically in anxiety ($p = 0.928$) and depression ($p = 0.187$). Higher completion rates were observed in the IG (79.9% (N = 19,252) vs. CG 70.1% (N = 8489); $p < 0.001$). Patient engagement remained consistent in both groups, as well as high satisfaction (IG: 8.76/10, SD 1.75 vs. CG: 8.60/10, SD 1.76; $p = 0.021$). Intervention-related adverse events were rare and even across groups (IG: 0.58% and CG 0.69%; $p = 0.231$). **Conclusions:** The study underscores the potential of scaling MSK care that is supported by AI without compromising patient outcomes, despite the increase in PT-to-patient ratios.

Keywords: musculoskeletal pain; physical therapy; telerehabilitation; eHealth; artificial intelligence; workflow; clinical decision support



Citation: Areias, A.C.; Janela, D.; Moulder, R.G.; Molinos, M.; Bento, V.; Moreira, C.; Yanamadala, V.; Correia, F.D.; Costa, F. Applying AI to Safely and Effectively Scale Care to Address Chronic MSK Conditions. *J. Clin. Med.* **2024**, *13*, 4366. <https://doi.org/10.3390/jcm13154366>

Academic Editor: Marco Alessandro Minetto

Received: 29 April 2024

Revised: 15 July 2024

Accepted: 23 July 2024

Published: 26 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Musculoskeletal (MSK) pain affects approximately 1.71 billion people worldwide [1], imposing a substantial societal and economic burden, which is reflected by USD 380.9 billion of yearly medical expenditure in the United States alone [2].

The urge to improve patient access to care has precipitated a transformative shift in MSK management towards digital rehabilitation [3,4] as a way to address geographical

barriers [5,6]. However, beyond the unbalanced geographical distribution of healthcare providers, there is also an insufficient number of trained clinicians to meet the rising demand caused by a continuous increase in the prevalence of MSK conditions [1,7]. This has resulted in persistent access barriers and delayed care, resulting in costly care escalations, as well as a significant prevalence of clinician burnout in recent years [8]. These factors demand a new wave of innovation that can address care access issues in a broad, scalable, and sustainable manner.

In this setting, artificial intelligence (AI) has emerged as a powerful tool in health-care [9]. Examples include the use of AI assistants to significantly reduce the documentation burden and physician burnout [8], and inpatient hospital surveillance systems to assist nurses by providing automatic alarms for patient monitoring [10].

Digital care programs (DCP) benefit from the automatic and centralized collection of data, making them well-suited for the integration of AI. AI systems can help scale MSK care by supporting monitoring and decision-making with data-driven insights, thereby reducing the workload and optimizing outcomes and efficiency, and reaching a higher number of patients in a sustainable way.

Research on AI tools in the context of MSK rehabilitation is still in its early stages. Current studies focus on initial stages of patient treatment, including the triaging [11] and phenotyping of patients [12], and predicting prognosis [13,14]. So far, however, research has not been focused on the impact of the introduction of scalability approaches in clinical outcomes. With this in mind, in this study, we aimed to evaluate whether AI-assisted care scalability compromises care delivery or if patients benefit similarly from high-quality care. To this end, we compared the clinical outcomes, as well as the engagement, satisfaction, and adverse events, of two cohorts with different physical therapists (PT)-to-patient ratios, whereby the cohort with the highest ratio was supported by a machine learning-based tool to enhance workflows and decision-making on patient progression. We hypothesized that patient outcomes would not be significantly different between cohorts, given the potentially optimized workflows obtained by the AI tool's introduction on the increased-ratio cohort.

2. Materials and Methods

2.1. Study Design

This is a post hoc analysis of two prospective, single-arm, IRB-approved studies (New England IRB: 120190313 and Advarra IRB: Pro00063337), which were prospectively registered on ClinicalTrials.gov (NCT04092946, NCT05417685) on 17 September 2019 and 14 June 2022, respectively.

Both cohorts were recruited with the intent to investigate the clinical and engagement-related outcomes of patients with MSK undergoing a DCP. The comparison cohort (PT-to-patient ratio: 1:57 (SD 36), with no assisting AI tool) had patients enrolled between 18 June 2020 and 13 June 2022, whereas the intervention cohort (PT-to-patient ratio: 1:129 (SD 27), with AI tool assistance) enrollment period occurred from 21 November 2022 to 8 July 2023. A washout period was considered, to avoid hybrid situations with pilot testing (between 14 June 2022 and 20 November 2022).

2.2. Population

The beneficiaries of employer health plans who reported chronic musculoskeletal (MSK) pain (defined as persistent or recurrent pain lasting ≥ 12 weeks) in the ankle, elbow, hip, knee, low back, neck, shoulder, wrist, or hand were invited to participate in the study through a dedicated enrollment website.

Exclusion criteria included: (1) health conditions that were incompatible with at least 20 min of light to moderate exercise; (2) ongoing cancer treatment; and (3) the presence of signs or symptoms indicative of serious pathology (e.g., rapid progressive motor weakness, sensory alterations, or bowel or bladder dysfunction). All participants provided informed consent.

2.3. Intervention

2.3.1. Digital Care Program

Both groups received a DCP (Sword Health Inc., Draper, UT, USA) consisting of exercise, education, and cognitive behavioral therapy (CBT) (as described elsewhere [15–17]), lasting up to 12 weeks according to each patient's condition.

During onboarding, patients chose a PT who would be responsible for prescribing, monitoring, and adjusting the program according to the patient's condition and goals, providing continuous support throughout the program. All PTs included in the study had a Doctor of Physical Therapy degree and a minimum of 3 years of experience.

Patients performed the exercise sessions independently at their convenience, using an FDA-listed class II medical device (Sword Health Inc., Draper, UT, USA) consisting of a dedicated tablet (software version 1.0 in the comparison group and 2.4 in the intervention group), motion tracking, a mobile app, and a cloud-based portal (Supplementary Figure S1). The exercise sessions were displayed on the tablet with real-time video and audio biofeedback on the patient's execution. Data related to sessions, including exercise performance (correct and incorrect movements), range of motion, the number of sets and repetitions completed, and self-reported pain and fatigue levels felt during exercise were automatically collected and stored in the cloud-based portal. This information was accessed by the assigned PT, enabling remote and asynchronous monitoring and adjustments to the prescription whenever needed.

Tailored education and CBT components were delivered through a smartphone app in the form of written articles, videos, audio content, and interactive modules, as described elsewhere [15]. Bidirectional communication was ensured through a built-in secure chat system within the smartphone app and video calls.

2.3.2. Workflow-Related AI Tool

Tasks involving data monitoring to stratify patients based on their need for clinical attention and program adjustments were deemed to be time-consuming by PTs and were, therefore, prioritized for a machine learning (ML)-based system. Subsequently, a multidisciplinary team combining data engineers, computer scientists, and PTs followed a human-in-the-loop framework [18] to develop a ML-based tool focused on streamlining adjustments to care programs (e.g., the introduction of new exercises or increasing exercise dosage). This tool provides recommendations to the PTs, as well as giving the reasons behind the suggestions, through the PT portal, enabling a more efficient revision of the patients' clinical status and supporting the PT's final decision (Figure 1 and Supplementary Table S1). The tool was primarily trained using previous PT decisions and fed with automatically collected data during each session (e.g., exercise performance and self-reported pain and fatigue during sessions; a detailed description of the input data can be found in Supplementary Table S2). Additionally, through a dynamic/iterative learning loop, the algorithm adapted and refined the recommendations, consequently improving its performance.

Algorithm performance was routinely tested against PTs' manual decisions. In instances where the AI tool could not provide a recommendation (insufficient accuracy), a notification prompted the PT to resort directly to a manual decision.

The AI tool was deployed in sequential pilot tests (between 14 June 2022 and 20 November 2022) before broad deployment that reflected the increase in the PT-to-patient ratio.

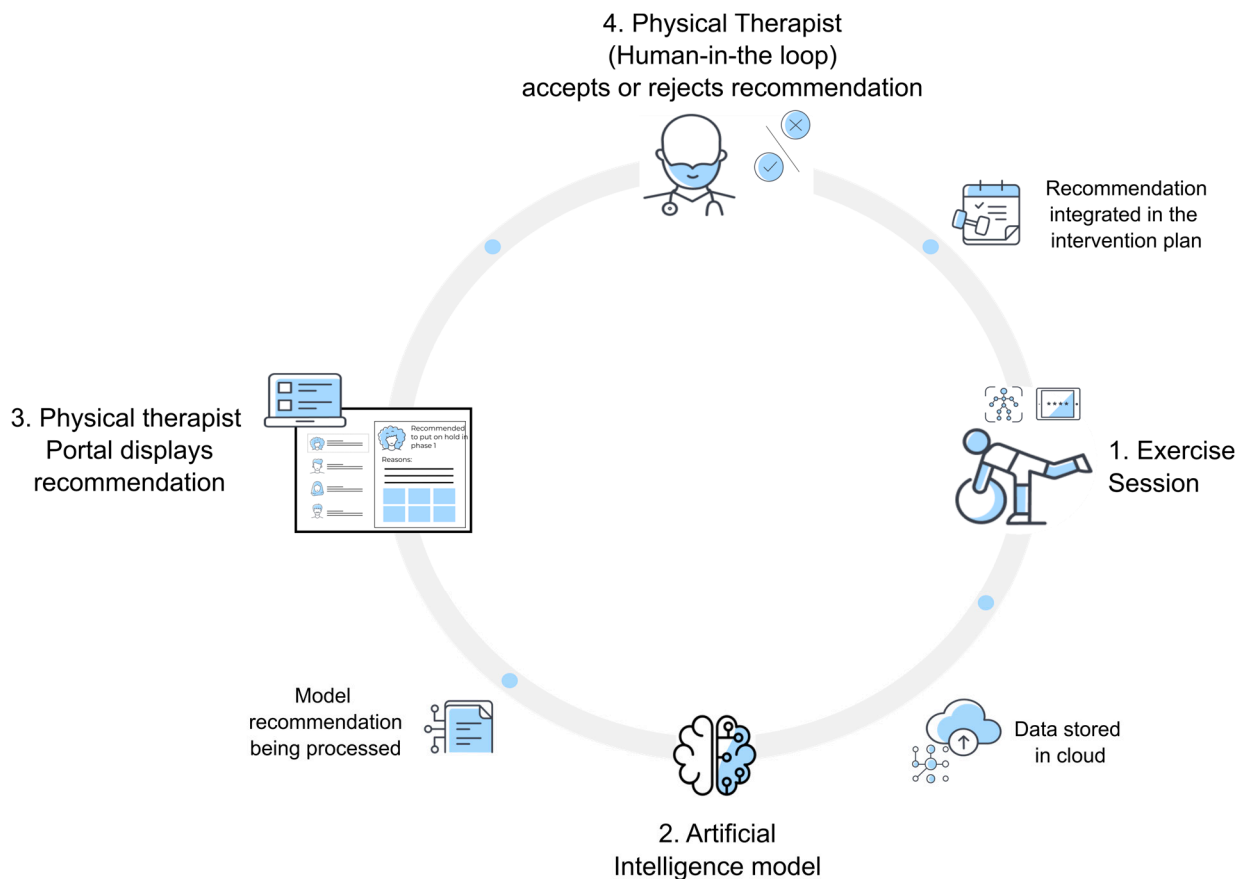


Figure 1. Schematic diagram of the AI tool. (1) Automatically collected data of a patient’s session are stored in the cloud and fed into an artificial intelligence (AI) model. (2) The AI model, primarily trained with previous physical therapist (PT) decisions, generates a recommendation about the possible adjustments to the patient’s care program (example: introduction of new exercises, increasing exercise dosage). (3) The recommendation is then displayed on the PT portal, along with the reasons behind the suggestions, enabling a more efficient revision of the patient’s clinical status to support the PT’s final decision. (4) The PT analyzes the recommendation, accepting or rejecting it, to support the final decision. Finally, a re-tailored intervention plan is sent to the patient’s tablet with a new updated session. This loop framework allows for streamlining adjustments to care programs to facilitate PT workflow.

2.3.3. Comparison Group (CG)

This cohort benefited from the standard DCP described above, wherein the information stored in the portal was manually managed and prioritized by the PT, and where the PT-to-patient ratio was 1:57 (SD 36), which is in line with the 1:60 ratio for in-person therapies [19].

2.3.4. Intervention Group: PT Portal Powered by AI Tool

The AI-intervention group (IG) received the same standard DCP as previously described, with a PT-to-patient ratio of 1:129 (SD 27), wherein the PTs were assisted by an AI tool integrated into the web portal.

2.4. Outcomes

Patient assessment was performed at baseline and at the 9th, 18th, and 27th sessions (according to the discharge time point) to analyze longitudinal changes between the baseline and program-end results.

PT-to-patient ratios were calculated month over month by averaging the number of patients being supported by the number of PTs available for a given month, regardless of the patient treatment stage.

Outcome measures comprised:

(1) Pain intensity, as measured by the Numerical Pain Rating Scale (NPRS) that is specific for the symptomatic body region: “Please rate your average pain over the last 7 days, from 0 (no pain at all) to 10 (worst pain imaginable)”. The response rate was calculated considering a minimum clinically important change (MCIC) of 30%. This threshold was defined based on the recommendations from the IMMPACT guidelines for clinical trials assessing chronic pain interventions [20];

(2) Mental health, with anxiety assessed by the 7-item Generalized Anxiety Disorder (GAD-7) scale (range 0–21) [21,22], and depression being assessed by the 9-item Patient Health Questionnaire (PHQ-9) (range 0–27) [22,23]. Higher scores denote worse symptomatology in both scales;

(3) Safety, which is assessed by the adverse event rate;

(4) Engagement, which is assessed by completion of the program (completion rate), average sessions per week, and communication between the PT and the patient;

(5) Satisfaction, which is assessed through the question “On a scale from 0 to 10, how likely is it that you would recommend this intervention to a friend or neighbor?”.

2.5. Safety and Adverse Events

Routine internal quality checks of the AI tool were conducted by comparing the algorithm’s recommendations against the manual decisions made by PTs to continuously assess its performance.

Patients were advised to report any adverse events (e.g., worsening of symptomatology, new symptoms, or other events that could interfere with the patient’s condition or the execution of the program) to the dedicated PT through the available communication channels for further assessment. These reports were documented and addressed using an internal software system, in accordance with internal quality and safety protocols. Additionally, the patient’s pain and fatigue levels during the exercise sessions (assessed by NRS; range 0–10) were collected at the end of each session.

2.6. Sample Size

The sample size estimation was based on the primary outcome—pain level—wherein a MCIC of 30% was selected, based on the psychometric properties of the scale [24]. Considering a power of 80%, a one-sided 0.05 significance-level non-inferiority study, and a 20% dropout rate [15,25], 2702 patients (1351 per group) would be necessary to detect a difference between the two groups.

2.7. Statistical Analysis

Descriptives were used to depict the patients’ demographic and clinical characteristics at baseline, along with engagement metrics. Comparisons between groups were performed using an independent samples *t*-test or the Mann–Whitney U test for continuous variables (with Bonferroni correction), or a chi-squared test for categorical variables. Both cohort reassessments were pivoted to 9, 18, and 27 sessions to ensure standardization of the reassessment timeframe.

Inverse probability weighting (IPW) was applied to control for confounding effects between groups [26] by differentially weighting each participant’s baseline characteristics using demographics that have previously been found to be meaningfully different: age, body mass index (BMI), and employment status. These weights were included in the data analysis to strengthen causal inference and were calculated for the whole cohort or for those with clinically relevant baseline scores (i.e., ≥ 5 , to further compute any changes in the GAD-7 and PHQ-9 outcomes). Multiple-group latent growth curve analysis (LGCA) was used to estimate the overall changes for each outcome, based on trajectories across time, and to per-

form comparisons between groups following an intention-to-treat analysis. LGCA, which applies a structural equation model, recognizes correlations between repeated measures for the same individual, provides model fit measures, and incorporates full information maximum likelihood (FIML) to handle missing data [27]. FIML uses all available data at each time point from all participants to calculate the maximum likelihood estimates, outperforming multiple imputation by chained equations or listwise deletion [28]. The model was adjusted for each individual discharge time point. Cohen’s d effect sizes were calculated for all clinical outcomes by comparing changes between the groups, considering the following thresholds: 0.2—small, 0.5—medium, 0.8—large, and 1.3—very large [29].

The odds of being a responder for pain was calculated using logistic regression analysis, considering an MCIC of 30% [20].

A robust sandwich estimator was used in all models for standard errors. All statistical analyses were conducted using R (version 4.2.2; R Foundation for Statistical Computing) and Python (version 3.11.4, Python Software Foundation, Wilmington, DE, USA), and the level of significance was set at $p < 0.05$ for all tests.

3. Results

From a total of 41,993 participants screened for eligibility, 36,186 patients started the study, of which 12,103 participants were in the CG and 24,083 in the IG cohorts (Figure 2).

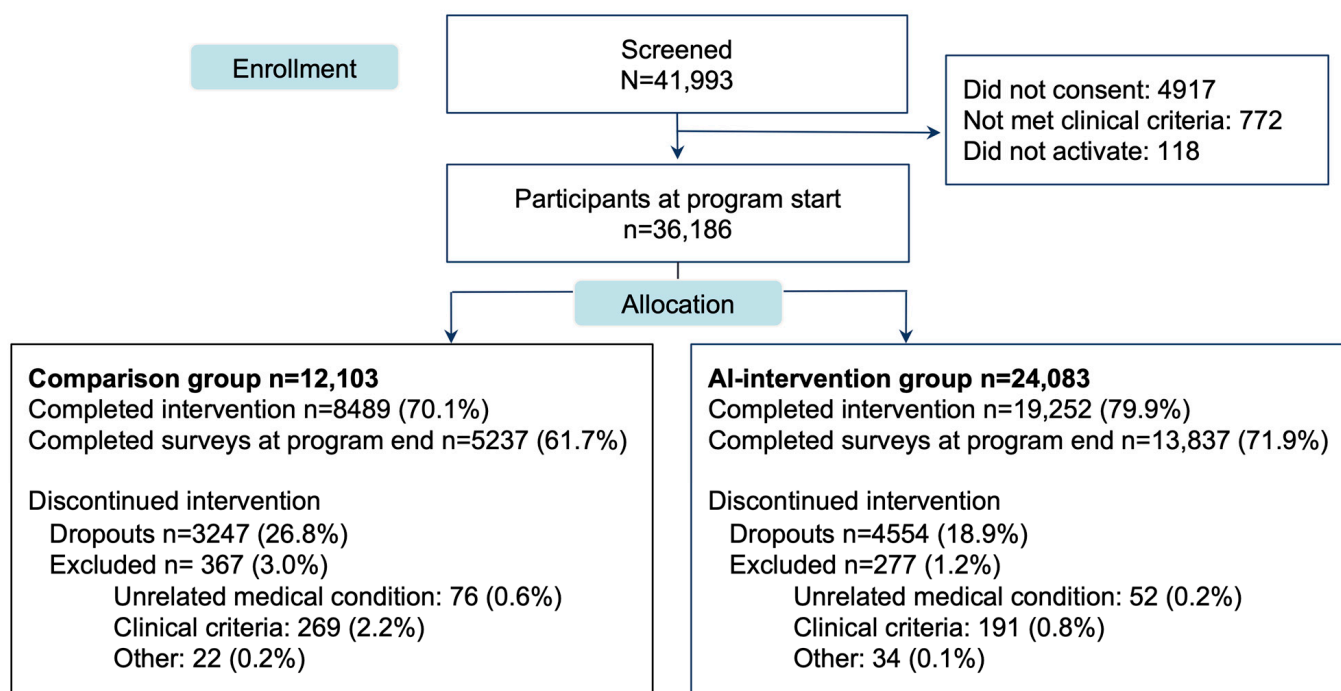


Figure 2. Study flow chart showing the number of participants who were screened and allocated to the comparison group (CG) and AI-intervention group (IG). Exclusions by clinical criteria consist of: (1) new symptoms that demand clearance by another healthcare provider who is not a physical therapist; (2) insufficient improvement or worsening of the condition, suggestive of the need for care escalation.

The defined cohorts were assisted by PTs with a difference in workload (PT-to-patient ratio: IG: 1:129 SD 27 vs. CG 1:57, SD 36) of 2.3× times.

3.1. Baseline Characteristics

Both groups exhibited a similar proportion of women (58%, $p = 0.414$) (Table 1). Statistically significant differences between groups were observed in BMI, albeit to a small extent (IG: 39.6%, CG: 37.9%, and $p < 0.001$). On average, the IG cohort was younger,

having a lower proportion of patients aged 60 years or older than the CG cohort (16.7% vs. 21.6%, and $p < 0.001$). Racial diversity was observed regardless of the group, with the highest proportion of patients identifying as non-Hispanic whites in both groups (IG: 67.9% and CG: 66.3%). Both groups were composed predominantly of individuals with higher education (considering a bachelor's or graduate degree: IG: 64.1%; CG: 66.2%). Participants predominantly resided in urban areas, with no significant differences between groups ($p = 0.077$). The IG cohort reported a higher percentage of full-time employment (87.5% vs. 75.7%), while the CG cohort had a greater proportion of part-time workers (16.1% vs. 4.2% in the IG; $p < 0.001$).

Table 1. Baseline characteristics for the AI-intervention group (IG) and comparison group (CG), following an intention-to-treat analysis. Filtered cases correspond to participants who presented with clinically relevant baseline scores (≥ 5) in terms of mental health variables.

| | IG (N = 24,083) | CG (N = 12,103) | p-Value |
|---|-----------------|-----------------|---------|
| Age (years), mean (SD) | 48.5 (11.6) | 50.0 (11.7) | <0.001 |
| Age categories (years), N (%): | | | |
| <25 | 342 (1.4) | 72 (0.6) | |
| 25–40 | 6099 (25.3) | 2795 (23.1) | |
| 41–60 | 13,615 (56.5) | 6623 (54.7) | <0.001 |
| >60 | 4027 (16.7) | 2613 (21.6) | |
| Gender, N (%) ^a : | | | |
| Woman | 14,126 (58.7) | 6984 (57.9) | |
| Man | 9849 (41.0) | 5044 (41.8) | |
| Non-binary | 67 (0.3) | 35 (0.3) | 0.414 |
| Other | 4 (0.0) | 1 (0.0) | |
| BMI (kg/m ²), mean (SD) ^b | 29.7 (7.0) | 29.3 (6.7) | <0.001 |
| BMI categories (kg/m ²), N (%) ^b : | | | |
| Underweight (<18.5) | 199 (0.8) | 102 (0.8) | |
| Normal (18.5–25) | 6223 (25.9) | 3349 (27.7) | |
| Overweight (≥ 25 –30) | 8100 (33.7) | 4051 (33.5) | 0.001 |
| Obese (≥ 30) | 9528 (39.6) | 4584 (37.9) | |
| Race/ethnicity, N (%) ^c : | | | |
| Asian | 2311 (9.8) | 928 (10.2) | |
| Black | 2297 (9.8) | 982 (10.8) | |
| Hispanic | 2268 (9.6) | 890 (9.8) | 0.017 |
| Non-Hispanic White | 15,991(67.9) | 6013 (66.3) | |
| Other | 690 (2.9) | 250 (2.8) | |
| Education level, N (%) ^d : | | | |
| Less than high school diploma | 219 (0.9) | 74 (0.7) | |
| High school diploma | 2172 (9.1) | 823 (8.1) | |
| Some college | 6133 (25.8) | 2528 (25.0) | 0.001 |
| Bachelor's degree | 9397 (39.5) | 4162 (41.2) | |
| Graduate degree | 5846 (24.6) | 2523 (25.0) | |
| Geographic location, N (%) ^e : | | | |
| Urban | 21,321 (88.8) | 10,767 (89.4) | |
| Rural | 2693 (11.2) | 1276 (10.6) | 0.077 |
| Employment status, N (%) ^f : | | | |
| Full-time job | 20,807 (87.5) | 8987 (75.7) | |
| Part-time job | 997 (4.2) | 1908 (16.1) | |
| Retired | 991 (4.2) | 608 (5.1) | <0.001 |
| Not employed | 990 (4.2) | 376 (3.2) | |

Table 1. Cont.

| | IG (N = 24,083) | CG (N = 12,103) | p-Value |
|---|-----------------|-----------------|---------|
| Clinical data, mean (SD) | | | |
| Analgesic intake, N (%) | 5580 (23.2) | 2850 (23.6) | 0.423 |
| Symptomatic anatomical area: | | | |
| Ankle | 1395 (5.8) | 482 (4.0) | |
| Elbow | 523 (2.2) | 262 (2.2) | |
| Hip | 2592 (10.8) | 1267 (10.5) | |
| Knee | 3633 (15.1) | 1832 (15.1) | |
| Low back | 8589 (35.7) | 4590 (37.9) | <0.001 |
| Neck | 2601 (10.8) | 1304 (10.8) | |
| Shoulder | 3730 (15.5) | 1928 (15.9) | |
| Wrist | 1020 (4.2) | 438 (3.6) | |
| Pain intensity ^c , mean (SD) | 4.73 (1.9) | 4.84 (1.9) | <0.001 |
| Mental health ≥ 5 , mean (SD): | | | |
| GAD-7 ^g | 8.72 (3.94) | 8.81 (4.06) | 0.083 |
| PHQ-5 ^h | 9.52 (4.20) | 9.23 (4.27) | 0.688 |

Data represent the mean \pm standard deviation or the number of patients and the percentage of the total where listed. Missing values or prefer not to answer: ^a 76, ^b 50, ^c 3566, ^d 2309, ^e 129, ^f 522. ^g GAD-7 ≥ 5 : IG = 3645 (30.2%) and CG = 8347 (34.7%); ^h PHQ-9 ≥ 5 : IG = 2803 (23.2%) and CG = 6192 (25.7%). Abbreviations: BMI, body mass index; GAD-7 generalized anxiety disorder 7-item scale; PHQ-9 patient health 9-item questionnaire.

Pain level differences between groups were not clinically meaningful [24] (IG: 4.84, SD 1.9 and CG: 4.73, SD 1.9; $p < 0.001$), and no significant differences were observed in terms of anxiety (IG: 8.72, SD 3.94 and CG 8.81, SD 4.06; $p = 0.083$) and depression (IG: 9.52, SD 4.20 and CG: 9.23, SD 4.27; $p = 0.688$).

3.2. Outcomes

3.2.1. Clinical Outcomes

Outcome model estimates for the IG and CG groups, adjusted with inverse probability weighting to balance the groups, and respective model fitness are presented in Supplementary Tables S3 and S4, respectively.

Significant improvements in pain were observed in both groups (Supplementary Table S3), as reflected by similar response rates [20]: 64% (95%CI 62–65%) for the IG cohort and 63% (95%CI 61–64%) for the CG cohort (difference between groups: $p = 0.399$; effect size 0.01).

Additionally, significant improvements in the mental health domain were observed similarly in both groups. The anxiety mean change was -4.0 (95%CI -4.2 ; -3.7) in the IG cohort and -4.1 (95%CI -4.4 ; -3.9) in the CG cohort, with no significant differences between groups (0.002 (95%CI -0.03 ; 0.03), $p = 0.928$; effect size: -0.04). Similarly, the depression mean change was -4.0 (95%CI -4.3 ; -3.7) in the IG cohort and -4.7 (95%CI -5.1 ; -4.3), in the CG cohort (both $p < 0.001$), regardless of the group (a very small effect size of -0.30 ; differences between groups: -0.03 (95%CI -0.07 ; 0.01), $p = 0.187$).

3.2.2. Engagement and Satisfaction

A higher completion rate was observed in the IG cohort (79.9% (N = 19,252)) than in the CG cohort (70.1% (N = 8489), $p < 0.001$). No meaningful differences in the average sessions performed per week were observed between groups, albeit being statistically significant (IG: 2.09, SD 1.03 vs. CG: 2.06, SD 1.16, $p < 0.001$).

There were significantly more reach-outs to patients in the IG cohort (25.11, SD 18.54 vs. 17.33, SD 21.88, in the CG cohort; $p < 0.001$). However, the number of messages initiated by patients remained similar, with the IG cohort sending, on average, 14.6 messages (SD 17.9) compared to 13.9 messages (SD 25.87) in the CG cohort (despite a significance of $p = 0.018$).

High and similar satisfaction with the program was observed in both groups (IG: 8.76 out of 10, SD 1.75, vs. CG: 8.60, SD 1.76, $p = 0.021$), despite the statistical difference.

3.2.3. Adverse Events

Intervention-related adverse events were rare and were evenly distributed across the groups, with the IG cohort reporting 140 (0.58%) and the CG cohort 83 (0.69%) ($p = 0.231$), detailed description in Supplementary Table S5. Condition-related (e.g., surgery) and condition/intervention-unrelated adverse events (e.g., respiratory infections, allergic reactions, falls outside intervention, cancer onset, etc.) accounted for the majority of the total adverse events (89.8%; $N = 1963/2186$).

4. Discussion

4.1. Main Findings

In this study, patient engagement, satisfaction, clinical outcomes, and adverse events were evaluated before and after a 2.3-fold increase in PT-to-patient ratio, supported by an AI tool integrated in the PT clinical portal. Similar clinical outcomes (pain, anxiety, and depression) were observed across both groups (denoted by very small effect sizes: -0.30 – 0.01 , $p = 0.187$ – 0.928).

Engagement was not negatively affected, with both groups presenting a similar frequency of sessions per week (IG: 2.09, SD 1.03 vs. CG: 2.06, SD 1.16, $p < 0.001$) and with the IG cohort reporting higher completion rates (79.9% ($N = 19,252$) vs. 70.1% ($N = 8489$), $p < 0.001$). Reach-outs initiated by the PT were more frequent in the IG (25.1 SD 18.5 vs. 17.3 SD 21.9, $p < 0.001$), while communication initiated by the members remained stable between groups (IG: 14.6, SD 17.9 vs. CG: 13.9, SD 25.9, $p = 0.018$). Satisfaction with the program was consistently high for both groups (IG: 8.76 out of 10, SD 1.75 vs. CG: 8.60, SD 1.76; $p = 0.021$). Crucially, the rate of adverse events remained consistently low across both groups.

All the above seem to suggest that care scalability is possible through the integration of AI tools in the workflow, without compromising clinical outcomes or program safety.

4.2. Comparison with Previous Studies

System sustainability promotion [30] and improvement in care delivery efficiency [8,10] have been fostered by AI transformational force in several healthcare sectors [31–35]. However, the application of AI in MSK care is still in its early stages.

Safety, interpretability, and transparency are key factors when developing AI tools. Many guidelines and research are now available to guide AI tool design following responsible AI principles [18,36,37]. Herein, we used a human-in-the-loop approach [38] where the final decision always rests with the healthcare provider. Clinicians can then validate AI-generated recommendations, correct errors, and provide additional insights based on their experience and expertise. Additionally, interpretability and transparency can be fostered by a more complete interface where the rationale supporting a given recommendation is provided [36,39], as was the case with the described AI tool.

To our knowledge, this is the first study assessing the outcomes after scaling an MSK management program leveraging the implementation of an AI-based system to streamline PT workflow. An encouraging finding of the present study was that neither engagement nor patient satisfaction was negatively impacted following scaling. This contrasts with the concerns expressed by both clinicians and patients that remote interventions, especially those operating at scale, might lead to reduced patient accountability, motivation, and adherence [40,41]. Completion rate was increased in the IG, reinforcing the lack of negative outcome from the scale context. The higher number of PT reach-outs in the IG may have contributed to increased engagement. This could be associated with streamlined workflows and the reduced burden of administrative tasks, diverting more time to patient contact. Additionally, a higher number of dropouts and unrelated medical exclusions were reported in the CG, which was possibly associated with the particular timeframe of the cohort in question (including the COVID-19 pandemic), which may have also contributed to a lower completion rate.

Importantly, clinical outcomes remained fairly stable between cohorts, as testified by the similar response rate for pain (64% vs. 63%, $p = 0.399$), and also in the mental health domain (anxiety change difference: 0.002 (95%CI -0.03 ; 0.03), $p = 0.928$, effect size: -0.04 , and depression change difference: -0.03 (95%CI -0.07 ; 0.01), $p = 0.187$, effect size: -0.30). The feedback loop between pain intensity and mental health issues is well known [42], as is the impact of a proper therapeutic alliance to address these intricacies and to effectively support condition management [43]. At the same time, and despite the increase in the PT-to-patient ratio, the patients' ability to communicate with their assigned PT was not affected, as evidenced by the similar number of communications initiated by the patients themselves in both groups (IG: 14.6 SD 17.9 vs. CG: 13.9 SD 25.9). As explored in other digital interventions, the ability to reach the clinician when needed has been reported as pivotal in initiating and maintaining a trust relationship that fosters better outcomes [40,44,45].

The stability of clinical outcomes between groups seems to advocate for care scalability and the use of technology to support PTs' clinical work, which follows the trend reported previously for AI tools developed for other healthcare professionals [10,46]. Crucially, the adverse event rate remained low and consistent despite the increase in the PT-to-patient ratio, underscoring the priority given to safety throughout the design, development, and implementation of this AI tool into the PTs' workflow. The adverse event rate herein reported was lower than that described in in-person interventions [47,48]. Specifically, in real-world digital settings, adverse events are scarcely reported, with a large retrospective study also reporting a low adverse event rate following the use of a self-supporting app [49]. This scarcity could be associated with the lack of the systematic collection and reporting of adverse events, particularly when considering that most digital interventions are self-managed without the support of a healthcare professional [50].

Overall, the findings of this study support the feasibility and safety of scaling remote care, utilizing AI-based tools to expand the outreach of healthcare professionals. These results align with a recent study that surveyed the perspectives of physical therapists regarding AI in healthcare, wherein 78.2% of participants agreed or strongly agreed that implementing AI in healthcare could increase work productivity [51,52].

4.3. Strengths and Limitations

The primary strength of the current study lies in its novelty, as it directly evaluates the scalability of MSK care that is supported by an AI-based system, using a large and diverse sample from a real-world context. Importantly, with the rise of AI tool development and adoption, concerns and awareness have been raised about the risks of poor AI conceptualization/design [53]. Therefore, the WHO guidelines [18] stress the importance of preserving the epistemic authority of healthcare providers in medical decision-making over the influence of automated (and sometimes biased and/or black-boxed) AI systems. Following WHO best practice, the AI tool herein described adopted a human-in-the-loop framework, not only to develop the tool itself but also to guarantee that all final decisions were made after critical reasoning and assessment by PTs. Moreover, the continuous internal quality check of the AI tool's performance against PT manual decisions safeguards against the possibility of poor recommendations.

However, there are limitations to consider. This non-randomized, controlled before-and-after study lacks randomization between groups, which was mitigated through the application of inverse probability weighting during statistical analysis. Additionally, the absence of an additional control group undergoing PT-to-patient ratio scaling without the implementation of an AI tool could have helped mitigate the influence of unaccounted confounding variables. However, ethical considerations prevented the pursuit of this strategy as safety and effectiveness concerns outweighed the potential additional insights that it could have provided. Consequently, there remains the possibility of other confounding variables that are not addressed in the study influencing outcomes, such as the cohort's timeframe or the potential use of complementary treatments. While the study addresses

a large sample size, the study population included the beneficiaries of employer health plans; therefore, further investigation in other cohorts should be carried out. Finally, future research should consider an evaluation of workflow time savings and qualitative analysis of the providers' perceptions of the tool for further workflow sustainability and increased workforce wellbeing.

5. Conclusions

This is the first study demonstrating that similar clinical outcomes (pain, anxiety, and depression) can be attained after an increase in the PT-to-patient ratio that is supported by the implementation of an AI-based system to streamline PT workflow. Moreover, patient engagement was not impacted, as evidenced by higher completion rates and a consistent level of satisfaction with the program. The stable adverse event rate across groups further supports the safety of scaling care through AI. The study advances current knowledge by underscoring the potential of AI technology in MSK management, particularly to enhance PT workflows. Future research should aim to clarify and quantify the gains in workflow efficiency and the associated cost-effectiveness of such approaches, as well as the perceptions of providers regarding the impact of such tools on their work.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jcm13154366/s1>, Figure S1: Participants performing exercise sessions of their digital care pro-program for (a) wrist/hand, and (b) low back conditions; Table S1. Description of the exercise prescription; Table S2. Input variables feeding the AI tool and example of the AI model output to the physical therapist; Table S3: Model estimates of clinical outcomes for each group, following an intent-to-treat analysis; Table S4: Latent growth curve model fit; Table S5: Adverse events related to the intervention registered for both I-intervention group (IG) and control group (CG).

Author Contributions: All authors made a significant contribution to the work, reported as follows: Conceptualization: F.C.; data curation: M.M. and A.C.A.; formal analysis: R.G.M.; interpretation of data: A.C.A., D.J. and F.C.; writing—original draft: A.C.A., D.J. and F.C.; writing—reviewing and editing: A.C.A., D.J., M.M., R.G.M., C.M., V.Y., F.D.C. and F.C.; funding: V.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Sword Health Inc. and was also developed within the scope of project n° 62—“Responsible AI”, financed by European Funds, namely, the Recovery and Resilience Plan—“Componente 5: Agendas Mobilizadoras para a Inovação Empresarial”—included in the NextGenerationEU funding program.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of New England (protocol code 120190313 on 17 September 2019) and Advarra (protocol code Pro00063337 on 14 June 2022).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

Acknowledgments: The authors would like to acknowledge the team of physical therapists responsible for managing the participants; Clara Matos and Daniel Rodrigues for the availability, insights, and discussions of AI methodologies; Francois P Manil for support in the calculation of PT-to-member ratios (all employees or previous employees of Sword Health, Inc.).

Conflicts of Interest: The authors declare the following competing financial interest: A.C.A., D.J., M.M., R.G.M., C.M., V.Y., F.D.C. and F.C. are employees of Sword Health, the sponsor of this study. F.D.C., V.Y. and V.B. also hold equity in Sword Health.

References

1. Cieza, A.; Causey, K.; Kamenov, K.; Hanson, S.W.; Chatterji, S.; Vos, T. Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: A systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **2020**, *396*, 2006–2017. [CrossRef]
2. Dieleman, J.L.; Cao, J.; Chapin, A.; Chen, C.; Li, Z.; Liu, A.; Horst, C.; Kaldjian, A.; Matyas, T.; Scott, K.W.; et al. US Health Care Spending by Payer and Health Condition, 1996–2016. *JAMA* **2020**, *323*, 863–884. [CrossRef]
3. Koonin, L.M.; Hoots, B.; Tsang, C.A.; Leroy, Z.; Farris, K.; Jolly, T.; Antall, P.; McCabe, B.; Zelis, C.B.R.; Tong, I.; et al. Trends in the Use of Telehealth during the Emergence of the COVID-19 Pandemic—United States, January–March 2020. *MMWR Morb. Mortal. Wkly. Rep.* **2020**, *69*, 1595–1599. [CrossRef] [PubMed]
4. Mitchai, P.M.; Mapinduzi, J.; Verbrugghe, J.; Michiels, S.; Janssens, L.; Kossi, O.; Bonnechère, B.; Timmermans, A. Mobile technologies for rehabilitation in non-specific spinal disorders: A systematic review of the efficacy and potential for implementation in low- and middle-income countries. *Eur. Spine J.* **2023**, *32*, 4077–4100. [CrossRef] [PubMed]
5. Argent, R.; Daly, A.; Caulfield, B. Patient Involvement With Home-Based Exercise Programs: Can Connected Health Interventions Influence Adherence? *JMIR Mhealth Uhealth* **2018**, *6*, e47. [CrossRef] [PubMed]
6. Fiani, B.; Siddiqi, I.; Lee, S.C.; Dhillon, L. Telerehabilitation: Development, Application, and Need for Increased Usage in the COVID-19 Era for Patients with Spinal Pathology. *Cureus* **2020**, *12*, e10563. [CrossRef] [PubMed]
7. Landry, M.D.; Hack, L.M.; Coulson, E.; Freburger, J.; Johnson, M.P.; Katz, R.; Kerwin, J.; Smith, M.H.; Wessman, H.C.; Venskus, D.G.; et al. Workforce Projections 2010–2020: Annual Supply and Demand Forecasting Models for Physical Therapists Across the United States. *Phys. Ther.* **2016**, *96*, 71–80. [CrossRef] [PubMed]
8. American Academy of Family Physicians Innovation Labs. *Using and AI Assistant to Reduce Documentation Burden in Family Medicine*; AAFP Innovation Labs: Leawood, KS, USA, 2021. Available online: <https://resources.suki.ai/i/1437132-using-an-ai-assistant-to-reduce-documentation-burden-in-family-medicine/3?> (accessed on 15 April 2024).
9. Lanotte, F.; O'Brien, M.K.; Jayaraman, A. AI in Rehabilitation Medicine: Opportunities and Challenges. *Ann. Rehabil. Med.* **2023**, *47*, 444–458. [CrossRef] [PubMed]
10. McGrath, S.P.; Perreard, I.M.; Garland, M.D.; Converse, K.A.; Mackenzie, T.A. Improving Patient Safety and Clinician Workflow in the General Care Setting with Enhanced Surveillance Monitoring. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 857–866. [CrossRef]
11. Annika, G.; Christoff, Z.; Kerstin, L. Evaluation of the accuracy of an artificial intelligence in identifying contraindications to exercise therapy—Comparison with and interrater reliability of physical therapists judgments. *Health Technol.* **2024**, *14*, 513–522. [CrossRef]
12. Tagliaferri, S.D.; Wilkin, T.; Angelova, M.; Fitzgibbon, B.M.; Owen, P.J.; Miller, C.T.; Belavy, D.L. Chronic back pain sub-grouped via psychosocial, brain and physical factors using machine learning. *Sci. Rep.* **2022**, *12*, 15194. [CrossRef]
13. Tschuggnall, M.; Grote, V.; Pirchl, M.; Holzner, B.; Rumpold, G.; Fischer, M.J. Machine learning approaches to predict rehabilitation success based on clinical and patient-reported outcome measures. *Inform. Med. Unlocked* **2021**, *24*, 100598. [CrossRef]
14. Zmudzki, F.; Smeets, R. Machine learning clinical decision support for interdisciplinary multimodal chronic musculoskeletal pain treatment. *Front. Pain Res.* **2023**, *4*, 1177070. [CrossRef]
15. Areias, A.C.; Molinos, M.; Moulder, R.G.; Janela, D.; Scheer, J.K.; Bento, V.; Yanamadala, V.; Cohen, S.P.; Correia, F.D.; Costa, F. The potential of a multimodal digital care program in addressing healthcare inequities in musculoskeletal pain management. *NPJ Digit. Med.* **2023**, *6*, 188. [CrossRef] [PubMed]
16. Cui, D.; Janela, D.; Costa, F.; Molinos, M.; Areias, A.C.; Moulder, R.G.; Scheer, J.K.; Bento, V.; Cohen, S.P.; Yanamadala, V.; et al. Randomized-controlled trial assessing a digital care program versus conventional physiotherapy for chronic low back pain. *NPJ Digit. Med.* **2023**, *6*, 121. [CrossRef]
17. Costa, F.; Janela, D.; Molinos, M.; Moulder, R.G.; Lains, J.; Francisco, G.E.; Bento, V.; Yanamadala, V.; Cohen, S.P.; Correia, F.D. Digital rehabilitation for hand and wrist pain: A single-arm prospective longitudinal cohort study. *Pain Rep.* **2022**, *7*, e1026. [CrossRef] [PubMed]
18. World Health Organization (WHO). *Ethics and Governance of Artificial Intelligence for Health. Guidance on Large Multi-Modal Models*; WHO: Geneva, Switzerland, 2024. Available online: <https://www.who.int/publications/i/item/9789240029200> (accessed on 26 March 2024).
19. Chorzewski, R. *AAOE Physical Therapy Survey: White Paper with Summary Results*; American Association of Orthopedic Executives (AAOE): Indianapolis, IN, USA, 2016. Available online: <https://www.multibriefs.com/briefs/aaoe/whitepaper011216.pdf> (accessed on 15 July 2024).
20. Dworkin, R.H.; Turk, D.C.; Wyrwich, K.W.; Beaton, D.; Cleeland, C.S.; Farrar, J.T.; Haythornthwaite, J.A.; Jensen, M.P.; Kerns, R.D.; Ader, D.N.; et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J. Pain* **2008**, *9*, 105–121. [CrossRef] [PubMed]
21. Spitzer, R.L.; Kroenke, K.; Williams, J.B.; Löwe, B. A brief measure for assessing generalized anxiety disorder: The GAD-7. *Arch. Intern. Med.* **2006**, *166*, 1092–1097. [CrossRef]
22. Bijker, L.; Sleijsers-Koehorst, M.L.S.; Coppieters, M.W.; Cuijpers, P.; Scholten-Peeters, G.G.M. Preferred Self-Administered Questionnaires to Assess Depression, Anxiety and Somatization in People With Musculoskeletal Pain—A Modified Delphi Study. *J. Pain* **2020**, *21*, 409–417. [CrossRef] [PubMed]

23. Kroenke, K.; Spitzer, R.L.; Williams, J.B. The PHQ-9: Validity of a brief depression severity measure. *J. Gen. Intern. Med.* **2001**, *16*, 606–613. [CrossRef]
24. Smith, S.M.; Dworkin, R.H.; Turk, D.C.; McDermott, M.P.; Eccleston, C.; Farrar, J.T.; Rowbotham, M.C.; Bhagwagar, Z.; Burke, L.B.; Cowan, P.; et al. Interpretation of chronic pain clinical trial outcomes: IMMPACT recommended considerations. *Pain* **2020**, *161*, 2446–2461. [CrossRef] [PubMed]
25. Scheer, J.; Costa, F.; Molinos, M.; Areias, A.; Janela, D.; Moulder, R.G.; Lains, J.; Bento, V.; Yanamadala, V.; Cohen, S.P.; et al. Racial and Ethnic Differences in Outcomes of a 12-Week Digital Rehabilitation Program for Musculoskeletal Pain: Prospective Longitudinal Cohort Study. *J. Med. Internet Res.* **2022**, *24*, e41306. [CrossRef] [PubMed]
26. Sterne, J.A.; Hernán, M.A.; Reeves, B.C.; Savović, J.; Berkman, N.D.; Viswanathan, M.; Henry, D.; Altman, D.G.; Ansari, M.T.; Boutron, I.; et al. ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* **2016**, *355*, i4919. [CrossRef]
27. McNeish, D.; Matta, T. Differentiating between mixed-effects and latent-curve approaches to growth modeling. *Behav. Res. Methods* **2018**, *50*, 1398–1414. [CrossRef] [PubMed]
28. Xiao, J.; Bulut, O. Evaluating the Performances of Missing Data Handling Methods in Ability Estimation From Sparse Data. *Educ. Psychol. Meas.* **2020**, *80*, 932–954. [CrossRef]
29. Sullivan, G.M.; Feinn, R. Using Effect Size-or Why the P Value Is Not Enough. *J. Grad. Med. Educ.* **2012**, *4*, 279–282. [CrossRef]
30. Åhlin, P.; Almström, P.; Wänström, C. Solutions for improved hospital-wide patient flows—A qualitative interview study of leading healthcare providers. *BMC Health Serv. Res.* **2023**, *23*, 17. [CrossRef] [PubMed]
31. Mohsen, F.; Ali, H.; El Hajj, N.; Shah, Z. Artificial intelligence-based methods for fusion of electronic health records and imaging data. *Sci. Rep.* **2022**, *12*, 17981. [CrossRef]
32. Huang, J.; Neill, L.; Wittbrodt, M.; Melnick, D.; Klug, M.; Thompson, M.; Bailitz, J.; Loftus, T.; Malik, S.; Phull, A.; et al. Generative Artificial Intelligence for Chest Radiograph Interpretation in the Emergency Department. *JAMA Netw. Open* **2023**, *6*, e2336100. [CrossRef]
33. Oren, O.; Gersh, B.J.; Bhatt, D.L. Artificial intelligence in medical imaging: Switching from radiographic pathological data to clinically meaningful endpoints. *Lancet Digit. Health* **2020**, *2*, e486–e488. [CrossRef]
34. Hyland, S.L.; Faltys, M.; Hüser, M.; Lyu, X.; Gumbsch, T.; Esteban, C.; Bock, C.; Horn, M.; Moor, M.; Rieck, B.; et al. Early prediction of circulatory failure in the intensive care unit using machine learning. *Nat. Med.* **2020**, *26*, 364–373. [CrossRef] [PubMed]
35. Niederer, D.; Schiller, J.; Groneberg, D.A.; Behringer, M.; Wolfarth, B.; Gabrys, L. Machine learning-based identification of determinants for rehabilitation success and future healthcare use prevention in patients with high-grade, chronic, nonspecific low back pain: An individual data 7-year follow-up analysis on 154,167 individuals. *Pain* **2023**, *165*, 772–784. [CrossRef] [PubMed]
36. World Health Organization (WHO). *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*; WHO: Geneva, Switzerland, 2021. Available online: <https://iris.who.int/bitstream/handle/10665/341996/9789240029200-eng.pdf?sequence=1> (accessed on 27 June 2024).
37. U.S. Food & Drug Administration; HealthCanada and Medicines and Healthcare products Regulatory Agency (MHRA). *Good Machine Learning Practice for Medical Device Development: Guiding Principles*. Available online: <https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles> (accessed on 27 June 2024).
38. Chen, R.J.; Wang, J.J.; Williamson, D.F.K.; Chen, T.Y.; Lipkova, J.; Lu, M.Y.; Sahai, S.; Mahmood, F. Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nat. Biomed. Eng.* **2023**, *7*, 719–742. [CrossRef] [PubMed]
39. Cutillo, C.M.; Sharma, K.R.; Foschini, L.; Kundu, S.; Mackintosh, M.; Mandl, K.D.; Beck, T.; Collier, E.; Colvis, C.; Gersing, K.; et al. Machine intelligence in healthcare—Perspectives on trustworthiness, explainability, usability, and transparency. *NPJ Digit. Med.* **2020**, *3*, 47. [CrossRef] [PubMed]
40. Fernandes, L.G.; Devan, H.; Fioratti, I.; Kamper, S.J.; Williams, C.M.; Saragiotto, B.T. At my own pace, space, and place: A systematic review of qualitative studies of enablers and barriers to telehealth interventions for people with chronic pain. *Pain* **2022**, *163*, e165–e181. [CrossRef]
41. Fernandes, L.G.; Oliveira, R.F.F.; Barros, P.M.; Fagundes, F.R.C.; Soares, R.J.; Saragiotto, B.T. Physical therapists and public perceptions of telerehabilitation: An online open survey on acceptability, preferences, and needs. *Braz. J. Phys. Ther.* **2022**, *26*, 100464. [CrossRef] [PubMed]
42. Crofford, L.J. Psychological aspects of chronic musculoskeletal pain. *Best. Pract. Res. Clin. Rheumatol.* **2015**, *29*, 147–155. [CrossRef] [PubMed]
43. Taccolini Manzoni, A.C.; Bastos de Oliveira, N.T.; Nunes Cabral, C.M.; Aquaroni Ricci, N. The role of the therapeutic alliance on pain relief in musculoskeletal rehabilitation: A systematic review. *Physiother. Theory Pract.* **2018**, *34*, 901–915. [CrossRef]
44. Elliott, T.; Tong, I.; Sheridan, A.; Lown, B.A. Beyond Convenience: Patients’ Perceptions of Physician Interactional Skills and Compassion via Telemedicine. *Mayo Clin. Proc. Innov. Qual. Outcomes* **2020**, *4*, 305–314. [CrossRef]
45. Licciardone, J.C.; Tran, Y.; Ngo, K.; Toledo, D.; Peddireddy, N.; Aryal, S. Physician Empathy and Chronic Pain Outcomes. *JAMA Netw. Open* **2024**, *7*, e246026. [CrossRef]
46. Zayas-Cabán, T.; Okubo, T.H.; Posnack, S. Priorities to accelerate workflow automation in health care. *J. Am. Med. Inform. Assoc.* **2022**, *30*, 195–201. [CrossRef] [PubMed]

47. Bagg, M.K.; Wand, B.M.; Cashin, A.G.; Lee, H.; Hübscher, M.; Stanton, T.R.; O'Connell, N.E.; O'Hagan, E.T.; Rizzo, R.R.N.; Wewege, M.A.; et al. Effect of Graded Sensorimotor Retraining on Pain Intensity in Patients with Chronic Low Back Pain: A Randomized Clinical Trial. *JAMA* **2022**, *328*, 430–439. [[CrossRef](#)] [[PubMed](#)]
48. Ingwersen, K.G.; Jensen, S.L.; Sørensen, L.; Jørgensen, H.R.; Christensen, R.; Søgaard, K.; Juul-Kristensen, B. Three Months of Progressive High-Load Versus Traditional Low-Load Strength Training Among Patients With Rotator Cuff Tendinopathy: Primary Results From the Double-Blind Randomized Controlled RoCTEx Trial. *Orthop. J. Sports Med.* **2017**, *5*, 2325967117723292. [[CrossRef](#)] [[PubMed](#)]
49. Jain, D.; Norman, K.; Werner, Z.; Makovoz, B.; Baker, T.; Huber, S. Using Postmarket Surveillance to Assess Safety-Related Events in a Digital Rehabilitation App (Kaia App): Observational Study. *JMIR Hum. Factors* **2021**, *8*, e25453. [[CrossRef](#)] [[PubMed](#)]
50. Zhao, P.; Yoo, I.; Lancey, R.; Varghese, E. Mobile applications for pain management: An app analysis for clinical usage. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, 106. [[CrossRef](#)] [[PubMed](#)]
51. Alsobhi, M.; Khan, F.; Chevidikunnan, M.F.; Basuodan, R.; Shawli, L.; Neamatallah, Z. Physical Therapists' Knowledge and Attitudes Regarding Artificial Intelligence Applications in Health Care and Rehabilitation: Cross-sectional Study. *J. Med. Internet Res.* **2022**, *24*, e39565. [[CrossRef](#)]
52. Khullar, D.; Casalino, L.P.; Qian, Y.; Lu, Y.; Krumholz, H.M.; Aneja, S. Perspectives of Patients About Artificial Intelligence in Health Care. *JAMA Netw. Open* **2022**, *5*, e2210309. [[CrossRef](#)]
53. Alowais, S.A.; Alghamdi, S.S.; Alsuhebany, N.; Alqahtani, T.; Alshaya, A.I.; Almohareb, S.N.; Aldairem, A.; Alrashed, M.; Bin Saleh, K.; Badreldin, H.A.; et al. Revolutionizing healthcare: The role of artificial intelligence in clinical practice. *BMC Med. Educ.* **2023**, *23*, 689. [[CrossRef](#)]

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