RETROSPECTIVE CLINICAL STUDY REPORT 2018-2020 1.1

ACCORDING TO THE STANDARDS
ICH E3 HARMONIZED GUIDELINE AND DIN EN ISO 14155

Retrospective Controlled Study of Clinical Efficacy of the Vivira App "In-Home Therapeutic Training Program" 2018-2020

Type of investigational product
App for smart phones/tablets
CE marked Medical Device

Name of investigational product
Vivira In-Home Therapeutic Training Program App

Indication studied
Pain of the musculoskeletal system

Study type
Retrospective, intraindividual controlled study with N= 2'517 patients suffering from pain of the musculoskeletal system

Name of the sponsor
Vivira Health Lab GmbH

Protocol identification
Vivira-RStudy-ICHReport-APP-1.1-20-F

Development phase of study
Postmarket, retrospective

Study initiation date
Jan 9, 2018

Date of early study termination
n/a

Study completion date
June 15, 2021

Persons responsible for the study report

Professor K. Weise, M.D., Dr. med.
Ärztekammer Certified Medical Specialist in Orthopedic Surgery & Traumatology
Professor of Orthopedic Surgery & Traumatology

Professor H.P. Zenner, M.D., Dr. med.
Ärztekammer Certified Medical Specialist in ORL, Head&Neck Surgery
Professor of Head&Neck Surgery

Clinical Research Organization
Prof. Dr. H.P. Zenner GmbH
Bismarckstr. 66
72072 Tübingen, Germany
Phone: +49 7071 49-645
office@hpzenner.de

Name of company/sponsor signatory

Dr. Philip Heimann, CEO
Vivira Health Lab GmbH
Kurfürstendamm 54/55
10707 Berlin, Germany
Tel.: +49-151-40732078
Email: philip@vivira.com

GCP

The study was performed in compliance with Good Clinical Practices (GCP), including the archiving of essential documents. The GCP essentials 2.6, 2.7, 2.9, do not apply (see Chapter Ethics)

Date of the report

Feb. 02, 2022

Earlier reports

Febr. 13, 2020: Clinical Evaluation Report by Dr. Fritz Ley

June 2, 2020: Medical Report of Vivira User Data by Dr. Fritz Ley

June 15, 2020: Retrospective Study Report 2019-2020 (Vivira-RStudy-ICHReport-APP-1-20-F) by Profs. Drs. med. Weise und Zenner

1. SYNOPSIS

Vivira is the digital application “In-Home Therapeutic Training Program” that is used on smartphones and tablets. Its conformity as MDD risk class I medical product has been declared. Users watch and follow video exercise demonstrations and read instructions and information. They input information regarding their wellbeing and configuration needs. This allows a feedback mechanism providing an individualized therapy program for each patient.

To investigate efficacy of the Vivira app a retrospective, intraindividual controlled study with N=2’517 patients suffering from pain of the musculoskeletal system was performed. A voluntary declaration of informed consent after prior electronic clarification is included in the App and had been obtained electronically. Descriptive statistical methods were used in the analysis of these data. Quantitative variables were described by measures of location (means, medians) and dispersion (standard deviations). Moreover, changes from baseline were calculated and the pertinent measures of location and dispersion were determined, too. Furthermore, 95 % confidence intervals were calculated for the changes from baseline and the Cohen´s effect sizes were additionally determined. For qualitative variables the corresponding absolute and relative frequencies were presented.

The overall analysis of the dataset at hand supports the assumption that Vivira´s digital home exercises can lead to significant improvements in pain scores. Vivira shows a significant improvement of self-reported pain scores, based on a VNRS. Additional post-hoc analyses show significant improvements between the initial and the two week assessment, the initial, the two week and the four week assessment and the initial, the two week, the four week and the eight week assessment.

An exploratory stratification across different pain areas (i.e. upper back, lower back, hip and knee) and different pain durations (i.e. acute, subacute and chronic pain) showed statistically insignificant improvements. We see a tendency towards a relevant improvement in pain scores for lower back (unadjusted p = 0.039), hip (unadjusted p = 0.05) and knee (unadjusted p = 0.088).

With the exception of hip and knee, significant improvements in strength and mobility could be detected between the first and the second assessment of the functional ability. Patients with hip and knee pain, however, show a significant response in increasing their coordination. This indicates a secondary benefit of the examined digital home exercise program. Interestingly, patients with lower back pain improve particularly sustained over an extended period of time (median follow up of 88,5 days, IQR 72-112) in the dimensions of strength and mobility.

Conclusion

This study presents early observational use data on the effectiveness of a digital home exercise program on the overall self-reported pain score reduction and demonstrates significant improvement in its primary analysis. Significant functional improvements, particularly in the domains of strength and mobility, could be demonstrated for upper and lower back pain, but not for hip and knee pain. Yet, coordination improved significantly in patients with hip and knee pain. Interestingly, chronic back pain profited from the extended use and showed significant increases in strength and mobility scores after a median of 88.5 days. Nonetheless, the presented data warrant a careful interpretation and further analyses are required to substantiate the early indicators of a therapeutic benefit of the examined digital home exercise program.

There was a significant number of drop-outs in the course of time. This was expected since the present study is no prospective study. The overall retention rate is 17% after two weeks, 10% after four weeks, 4% after eight weeks and 3% after twelve weeks. A high attrition rate is prone to bias. Thus, these preliminary data warrant a careful interpretation of its clinical implications.

The app's software does not offer the option of an active surveillance of AEs. Vivira's PMS, however, includes the option to record complaints and repairs. These provided no indication of AE relevant hazards and AEs. It was also not necessary to file a notification of hazard or AEs with the supervisory authorities. There was also no recall action. The preclinical risk analysis revealed foreseeable hazards and AEs. However, from the PMS data, authority databases data, data from preclinical risk analysis and expert opinion it can be concluded that the risk of AEs is low.

2. TABLE OF CONTENTS

3. LIST OF ABBREVIATIONS

App *Application*
CE-marked *Certification Mark (European Conformity)*
CIP *Clinical Investigation Plan*
DFT *Dorsiflexion Lunge Test*
DiGA *Digitale Gesundheitsanwendung (digital health application)*
DiGAV *Digitale-Gesundheitsanwndungen-Verordnung (regulation concerning digital health applications)*
EEC *European Economic Community*
EN *European Norm*
FMS *functional movement screen*
GCP *Good Clinical Practices*
GDNÄ *Gesellschaft Deutscher Naturforscher und Ärzte*
GHTF *Global Harmonization Task Force*
i.e. *id est - that is*
IEC *International Electrotechnical Commission*
IMDRF *International Medical Devices Regulators Forum*
ISO *International Organization for Standardization*
ITT *Intention-to-Treat*
LBP *low back pain*
MCID *minimum clinically important diferences*
NRS *Numeric Rating Scale*
ORL *Oto-Rhino-Laryngologie*
PMCF *Post-Market Clinical Follow-up*
PP *Per-Protocol*
QoL *Quality of Life*
SFMA *selective functional movement assessment*
SG *Study Group*
TDD *Test driven development*
TTP *Tübingen Titanium Prostheses*
USA *United States of America*
VAS *Visual Analogue Scale*
VNRS *verbal numerical rating scale*

4. DEFINITION OF TERMS

Adverse event: Side effects that are harmful; any untoward medical occurrence, unintended disease or injury, or any untoward clinical signs (including an abnormal laboratory finding) in subjects, users or other persons whether or not related to the investigational medical device. [EN ISO 14155:2011]

Adverse reaction: Events for which a *causality* link to the tested intervention is well established and strong enough (sensitive and specific)

Android: an open-source operating system used for smartphones and tablet computers.

Android device: an electronic device that runs the operating system Android.

Application: a program or piece of software designed to fulfil a particular purpose.

Backend: the data access layer of the app.

Bias: bias is a systematic deviation of an outcome measure from its true value, leading to either an overestimation or underestimation of a treatment's effect. It can originate from, for example, the way patients are allocated to treatment, the way treatment outcomes are measured and interpreted, and the way data are recorded and reported. [Adapted from GHTF SG5/N2R8:2007]

Broadband internet connection: an internet connection with a high data transfer rate.

Clinical data: the safety and/or performance information that is generated from the clinical use of a device. Clinical data are sourced from:
- clinical investigation(s) of the device concerned; or
- clinical investigation(s) or other studies reported in the scientific literature, of a similar device for which equivalence to the device in question can be demonstrated; or
- published and/or unpublished reports on other clinical experience of either the device in question or a similar device for which equivalence to the device in question can be demonstrated.

Clinical evaluation: a methodologically sound ongoing procedure to collect, appraise and analyse clinical data pertaining to a medical device and to evaluate whether there is sufficient clinical evidence to confirm compliance with relevant essential requirements for safety and performance when using the device according to the manufacturer's Instructions for Use.

Note: In exceptional cases where an instruction for use is not required, the collection, analysis and assessment are conducted taking into account generally recognised modalities of use.

Clinical evidence: the clinical data and the clinical evaluation report pertaining to a medical device. [GHTF SG5/N2R8:2007]

Clinical investigation: systematic investigation in one or more human subjects, undertaken to assess the safety or performance of a medical device.
Note: 'clinical trial' or ' clinical study' are synonymous with ' clinical investigation'.
[EN ISO 14155:2011]

Clinical investigation plan: document that states the rationale, objectives, design and proposed analysis, methodology, monitoring, conduct and record-keeping of the clinical investigation. [EN ISO 14155:2011]

Clinical performance: behaviour of a medical device or response of the subject(s) to that medical device in relation to its intended use, when correctly applied to appropriate subject(s). [EN ISO 14155:2011]

Clinical safety: freedom from unacceptable clinical risks, when using the device according to the manufacturer's Instructions for Use. [MEDDEV 2.7/2 revision 2] Not simply absence of evidence of harm.

Clinical use: use of a medical device in or on living human subjects. Includes use of a medical device that does not have direct patient contact.

Confluence: is a collaboration software program developed and published by the software company Atlassian.

Coordination: the ability to use different parts of the body together smoothly and efficiently. It is one of the three functional dimensions in the Vivira movement test.

Coordination Score: an efficacy variable used in this document. A score between 0 and 100 is calculated based on the user's self-assessment in the Movement Test specifically relating to a user's abilities in the dimension of coordination.

Device registry: an organised system that uses observational study methods to collect defined clinical data under normal conditions of use relating to one or more devices to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure and that serves predetermined scientific, clinical or policy purpose(s). The term "device registry" is different from the concept of device registration and listing. [MEDDEV 2.12/2 rev2]

Educational content: short content articles provided in the Vivira app giving users information in several categories, e.g. pain, relaxation, etc.

Elixir: a functional, concurrent, general-purpose programming language that runs on the Erlang virtual machine used for the Vivira app.

Equivalent device: a device for which equivalence to the device in question can be demonstrated.

Exercise day: a day where a user has completed at least one exercise. An exercise is completed if the user confirmed that they have performed the exercise and given feedback regarding pain and their completion of reps or holding times.

Exercise difficulty: the integration of two dimensions regarding the exercises assigned to a user. The first dimension determining exercise difficulty is the intensity at which an exercise is to be performed, e.g. the number of repetitions to perform or the length of the holding time. The second dimension of exercise difficulty is defined by the exercise itself. Within each exercise group exercises can be ordered from least to most challenging. A user will progress in exercise difficulty first along all exercise intensities of the easiest exercise of a group before progressing to the next and more challenging exercise.

Exercise history: an overview in the app that depicts on which dates users exercised, which exercises they performed and which feedback they gave.

Exercise program: a specific set and ordering of exercises assigned to each user based on the input they provide at the start of and throughout their use of the Vivira app.

Exercise video: a video displayed in the Vivira app as part of the training program. In the video the correct exercise execution is demonstrated, and additional information is displayed.

Exercise progression: a term defining the Vivira exercise system, which bases exercise difficulty on user feedback. For each user the daily exercise program is compiled based on the exercises that were performed in the past and the feedback the user gave for each exercise. Only if users are able to perform one exercise at a given exercise intensity do they move on (progress) to a higher intensity for the same exercise or a more challenging exercise.

Feasibility study: a clinical investigation that is commonly used to capture preliminary information on a medical device (at an early stage of product design) to adequately plan further steps of device development, including needs for design modifications or parameters for a pivotal study. [MEDDEV 2.7/2 revision 2]

Frontend: the presentation layer of the app.

Google Play Store: a digital distribution platform, developed and maintained by Google Inc., for mobile apps on the Android operating system. The store allows users to browse and download apps developed with the Google software development kit.

GraphQL: an open-source data query and manipulation language for APIs, and a runtime for fulfilling queries with existing data.

Harmonised standards: standards whose references have been published in the Official Journal of the European Communities.

Harms: The totality of possible adverse consequences of an intervention or therapy; direct opposite of benefits, against which they must be compared.

Hazard: potential source of harm. [EN ISO 14971:2012]

Hazard due to substances and technologies: for the purpose of this document, a hazard that is seen with products that share specific characteristics. This includes products that contain the same materials and substances, material combinations, use the same technologies, produce similar abrasion, are used with the same type of surgical approach,

share the same manufacturing procedures or impurities, or share other characteristics.

Household Score: an efficacy variable used in this document. A value between 0 and 11 self-reported by the user as part of the weekly Wellbeing Journal. The score describes to what extent the user felt that they were limited in their home activities.

In-app purchase: refers to the buying of goods and services from inside an application on a mobile device, such as a smartphone or tablet.

Incident: any malfunction or deterioration in the characteristics and/or performance of a device, as well as any inadequacy in the labelling or the instructions for use which, directly or indirectly, might lead to or might have led to the death of a patient, or user or of other persons or to a serious deterioration in their state of health. [MEDDEV 2.12/rev 8]

Information materials supplied by the manufacturer: for the purpose of this document, this refers to the labelling, instructions for use and the manufacturer's promotional materials for the device under evaluation.

Intended purpose: the use for which the device is intended according to the data supplied by the manufacturer on the labelling, in the instructions and/or in promotional materials.

Investigator: individual member of the investigation site team designated and supervised by the principal investigator at an investigation site to perform critical clinical-investigation-related procedures or to make important clinical investigation-related decisions. [EN ISO 14155:2011]

iOS: an operating system used for mobile devices manufactured by Apple Inc.

iOS App Store: a digital distribution platform, developed and maintained by Apple Inc., for mobile apps on its iOS operating system. The store allows users to browse and download apps developed with Apple's iOS software development kit.

iOS device: an electronic device that runs the operating system iOS.

Java: a general-purpose programming language used for the Android operating system.

Jira: a proprietary issue tracking product developed by the software company Atlassian that allows bug tracking and agile project management, among other things.

Kotlin: a cross-platform, statically typed, general-purpose programming language with type inference used for the Android operating system.

Leisure Score: an efficacy variable used in this document. A value between 0 and 11 self-reported by the user as part of the weekly Wellbeing Journal. The score describes to what extent the user felt that they were limited in their leisure activities.

Limitation in quality of life: an efficacy variable used in this document. The limitation in quality of life is represented by a number the user selects on a numerical verbal rating scale. 0 is the lowest and 10 the highest limitation of quality of life.

Mobility Score: an efficacy variable used in this document. A score between 0 and 100 is calculated based on the user's self-assessment in the Movement Test specifically relating to a user's abilities in the dimension of mobility.

Movement Test: a self-assessment in the Vivira app, which is performed every four weeks. Based on the user's ability to perform a set of defined movements a score is compiled for each dimension (Stability, Mobility, Coordination) and a total score is calculated.

MVP: a derivation of the model–view–controller (MVC) architectural pattern which mostly used for building user interfaces for the Android operating system.

Onboarding: the configuration phase of the app, where users answer questions that determine the assignment of a therapy program. Onboarding is completed after users have answered all questions and created a Vivira account.

Pain score: the number a user selects on a numerical rating-scale that represents the intensity of their perceived pain.

Paywall: an arrangement whereby access is restricted to users who have paid to use the app.

Phoenix: a web development framework written in the functional programming language Elixir.

PMCF plan: the documented, proactive, organised methods and procedures set up by the manufacturer to collect clinical data based on the use of a CE-marked device corresponding to a particular design dossier or on the use of a group of medical devices belonging to the same subcategory or generic device group as defined in Directive 93/42/EEC. The objective is to confirm clinical performance and safety throughout the expected lifetime of the medical device, the acceptability of identified risks and to detect emerging risks on the basis of factual evidence. [MEDDEV 2.12/2 rev.2]

PMCF study: a study carried out following the CE marking of a device and intended to answer specific questions relating to clinical safety or performance (i.e. residual risks) of a device when used in accordance with its approved labelling. [MEDDEV 2.12/2 rev.2]

Premium user: a person who uses or operates Vivira and is charged based on the usage duration they selected.

Progression system: the system defining how user feedback after each exercise determines the next assigned set of exercises. The progression system ensures that users are assigned exercises that are in line with their individual abilities.

Progress visualization: a section of the app that provides graphic representations of the self-reported input that users have made in the Wellbeing Journal and the Movement Test.

Risk: combination of the probability of occurrence of harm and the severity of that harm. [EN ISO 14971:2012]

Risk management: systematic application of management policies, procedures and practices to the tasks of analysing, evaluating, controlling and monitoring risk. [EN ISO 14971:2012]

RxJava: a programming library used for the Android operating system.

RxKotlin: a programming library used for the Android operating system.

Safety: Substantive evidence of an absence of harm. Not simply absence of evidence of harm.

Serious adverse event:
Serious adverse events[1] *during clinical investigations:* in their most severe forms, threaten life or function. If suspected to be product-related (adverse reactions), might be significant enough to lead to important changes in the way the medicinal product is developed (e.g., change in dose, population, needed monitoring consent forms).

Adverse event that
a) led to death,
b) led to serious deterioration in the health of the subject, that either resulted in
1) a life-threatening illness or injury, or
2) a permanent impairment of a body structure or a body function, or
3) in-patient or prolonged hospitalization, or
4) medical or surgical intervention to prevent life-threatening illness or injury or permanent impairment to a body structure or a body function,
c) led to foetal distress, foetal death or a congenital abnormality or birth defect.
NOTE: Planned hospitalization for a pre-existing condition, or a procedure required by the CIP [Clinical Investigation Plan], without serious deterioration in health, is not considered a serious adverse event. [EN ISO 14155:2011]

Side effects: Unintended effects, not necessarily imply harmful

[1] As defined by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, document E2A (available at www.ich.org)

Smartphone: a mobile phone that performs many of the functions of a computer, typically having a touchscreen interface, internet access, and an operating system capable of running downloaded apps.

Stability: Stability is the resistance of a muscle or group of muscles to control joint position and balance. It is one of the three functional dimensions in the Vivira movement test.

Stability Score: an efficacy variable used in this document. A score between 0 and 100 is calculated based on the user's self-assessment in the Movement Test specifically relating to a user's abilities in the dimension of stability.

Sufficient clinical evidence: an amount and quality of clinical evidence to guarantee the scientific validity of the conclusions.

Surveillance of harms, passive: participants spontaneously report

Surveillance of harms, active: participants are asked (in structured questionnaires or interviews or predefined laboratory or other diagnostic tests at prespecified time intervals)

Swift: a general-purpose, multi-paradigm, compiled programming language developed by Apple Inc. for iOS.

Tablet PC: a mobile device, typically with a mobile operating system and touchscreen display processing circuitry, and a rechargeable battery in a single, thin and flat package.

Total Functional Score: an efficacy variable used in this document. The total score (between 0 and 100) is calculated based on the user's self-assessment in the Movement Test.

Toxicity: Describes product-related harms. The term may be most appropriate for laboratory-determined measurements, although it is also used in relation to clinical events.

Abnormal laboratory values may be described as laboratory-determined toxicity. The disadvantage of the term “toxicity” is that it implies causality. If authors cannot prove causality, the terms “abnormal laboratory measurements” or “laboratory abnormalities” are more appropriate to use.

Test-driven development (TDD): a process of modifying the code in order to pass a test designed previously. In Software Engineering, It is sometimes known as "Test First Development."

Test user: a person who uses or operates Vivira free of charge for a limited time.

User: a person who uses or operates an application.

User interface: the means by which the user and the Vivira app interact.

Visualization: Human perception and cognition of an image.

Vivira account: Users of the Vivira app create an account, which enables them to access the app using their email address and a password. Each Vivira user and vivira account is assigned a unique user identification (User ID).

Wellbeing Journal: a set of questions users answer on a weekly basis. Questions cover pain intensity in the last week, limitations in quality of life, limitations at home, limitations at work and limitations in leisure activities.

Work Score: an efficacy variable used in this document. A value between 0 and 11 self-reported by the user as part of the weekly Wellbeing Journal. The score describes to what extent the user felt that they were limited in their work.

5. ETHICS

Independent Ethics Committee

The study concept was advised by the *Ethik-Kommission der Landesärztekammer Baden-Württemberg* under the appraisal number F-2020-075.

Clinical Trial Registry

The study is registered under DRKS00021785 in the German Clinical Trials Registry DRKS.

Ethical Conduct of the Study

This is a retrospective study of efficacy data of a CE-marked class I medical device. After online clarification a voluntarily submitted declaration of informed consent for data use has been obtained by the process owner. Access to person related data was restricted to the data owners. The data access of the evaluators/authors was restricted to anonymous metadata. The study was conducted in accordance with the ethical principles that have their origins in the Declaration of Helsinki and in compliance with Good Clinical Practices (GCP), including the archiving of essential documents. For the subsequent reasons, however, the GCP essentials 2.6, 2.7, 2.9, do not apply:

- 2.6: This is a retrospective study, thus a prestudy Ethics Committee appraisal is not applicable
- 2.7: The medical device is a CE-marked product that has been used by lay persons without prescription by a doctor. Thus, medical decisions by doctors do not apply.
- 2.9: A voluntarily submitted declaration of informed consent for data use after prior clarification has been obtained. However, the oral clarification by a doctor was not applicable.

Patient Information and Consent

A voluntarily submitted declaration of informed consent after prior electronic clarification is included in the App and has been obtained electronically. The oral clarification by a doctor was not applicable. Along this line processing purposes according to § 4 paragraph 2 sentence 1 number 1 to 3 DiGAV are listed and summarised in a consent. This is done by means of a consent form issued by the manufacturer as unilaterally consent form, which complies with the requirements of the DSGVO with regard to being informed, being submitted voluntarily

and explicitly and corresponding to revocability. Consent to data processing within the scope of a study to prove positive effects of care can be obtained from users. The user can request optional consents as such and can individually refuse or revoke them. The display of user questionnaires via DiGA for the collection and subsequent processing of feedback on the user experience or on possible technical problems is also subject to obtaining consent. This also applies to data processing for further development of the App. A comprehensive tracking of user activity is not possible.

6. INVESTIGATORS AND STUDY ADMINISTRATIVE STRUCTURE[2]

Sponsor
Vivira Health Lab GmbH, Berlin (Dr. Philip Heimann, CEO)

CRO
Clinical Research Organization
H.P. Zenner Clinical GmbH&CoKG, Tübingen

Principal investigators/study report
Professor Dr. med. K. Weise, M.D.
Professor Dr. med. H.P. Zenner, M.D.

Steering, data collection and monitoring
Automatic and integrated property of the App

Statistics and statistical report
Dr. med. Leo Benning, MPH

[2] In the appendix may be found a list of the investigators with their affiliations, their role in the study and their qualifications

7. INTRODUCTION

Musculoskeletal conditions are among the most important contributors to the global burden of disease[3]. As the most prevalent conditions among working populations, they do not only contribute greatly to direct, but also to indirect health care costs[4]. At the same time, the access to and the availability of adequate therapeutic means for the spectrum of musculoskeletal conditions remain challenging[5]. Yet, it has repeatedly been shown that physical activity and structured exercise programs are effective in addressing certain kinds of musculoskeletal conditions. This applies in particular to unspecific and degenerative musculoskeletal pain[6].

In the recent past, health apps have been developed to address these and other allocation problems of a sufficient spread of therapy methods. In 2017, the number of health apps released from iTunes and Google Play exceeded 300 000, with nearly 25% dealing with disease self-management. One-third of adults in the US with smartphones or tablets use health apps to achieve health behaviour goals and help with medical decision making. International standards exist with regard to software engineering, privacy, security and usability of mobile apps in general (e.g. International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) standards). For health apps an important standard was set by the German DiGAV 2020 for provision within the German social security system GKV.

To address the outlined challenge, this study presents preliminary use data of a digital home exercise program for unspecific and degenerative pain in the back, the hip and the knee and demonstrates early data on self-reported pain score reductions and functional improvements.

The present study report presents the results of a retrospective observational study on the effectiveness and safety of the Vivira In-Home Therapeutic Training Program App (Vivira app). Furthermore, data on the adherence to the digital home exercise program are presented. The app was used for the treatment of patients suffering from pain of the musculoskeletal system including non-specific pain in hip, knee and back, but also in osteoarthrosis. The app comprises a comprehensive multidisciplinary feedback controlled treatment concept, including patient education and video-supported physiotherapy. The content of the app is in accordance with current German guidelines including the National Care Guideline "Non-Specific Back Pain" for the management of LBP[7]. As primary efficacy variable using a verbal numerical rating scale (VNRS) the study compared pretherapeutic pain levels with the relevant data at the end of a period of up to 3 months after intention to treat. Thus, efficacy was controlled by an

[3] , Lim SS, Abbafati C, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2020;396(10258):1204-1222.

[4] Bevan S. Economic impact of musculoskeletal disorders (MSDs) on work in Europe. Best Pract Res Clin Rheumatol. 2015;29(3):356-373.

[5] McCallum CA. Access to physical therapy services among medically underserved adults: a mixed-method study. Phys Ther. 2010;90(5):735-747

[6] van Gool CH, Penninx BW, Kempen GI, et al. Effects of exercise adherence on physical function among overweight older adults with knee osteoarthritis. Arthritis Rheum. 2005;53(1):24-32.

[7] AWMF 2017. Nationale Versorgungsleitlinie Kreuzschmerzen

intraindividual control. Descriptive variables included functional scores on mobility, stability and coordination.

8. STUDY OBJECTIVES

Study objective is the retrospective controlled determination of postmarket clinical efficacy of the Vivira App “In-Home Therapeutic Training Program”.

9. STUDY PRODUCT

Vivira is a CE-marked home exercise program and consists of a series of specific exercises that include a two-dimensional progression module. In brief, patients are guided through a pain and functional assessment at baseline and are prompted to provide close feedback to whether they can complete individual exercises presented and to whether these exercises cause any complaints. If a complaint, primarily a pain, is reported, the progression module pauses and reassesses the intensity of the exercise program. Exercises are selected to address functional deficits. Additional, overall pain scores are collected every week and a follow-up functional assessment is prompted every month. As functional assessments are not part of the safety assessment, these prompts could be skipped.

Product design

Vivira is a Security Class B software.

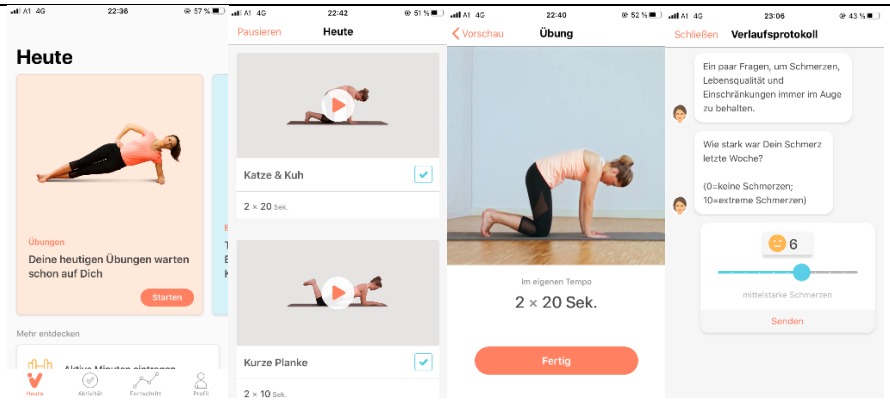
Product variants

During the study period, Vivira could be be tested by users free of charge for 7 days. After 7 days of free usage, the user may purchase a Vivira Premium package for 3, 6, or 12 months of usage.

Product properties

Vivira is a digital application that is used on smartphones and tablets. Users tap on the device’s screen to control the application. They watch the video exercise demonstrations and read instructions and information. They input information regarding their wellbeing and configuration needs. The Vivira App is designed to allow handling according to today's established design standards. A more detailed description may be found in the Description of Product.

Use requires broadband internet connection and a smartphone or tablet.

Category	Product
Product Name	Vivira
Screen shots of user interface	
Project Name	n/a
Product Models	app for iOS devices, app for Android devices
Device group, subgroup	Software app
CE Status	CE-marked 2017
Market	Having received the CE Mark Class I in March 2017, the product was placed on the market in EU countries and in Switzerland, with short user tests in the U.S. and in certain Asian countries. Currently, the Product is on the market in Germany, Austria, Ireland, the UK, and Switzerland.
Classification	Class I according to MDD appendix VII, class IIa according to MDR
Accessories	None
Manufacturer(s)	Vivira Health Lab GmbH, Kurfürstendamm 54/55, 10707 Berlin
User group	Patients, lay persons
Medical indication	• Gonarthrosis • Nonspecific knee pain • Coxarthrosis • Nonspecific hip pain • Nonspecific back pain • Spine osteochondrosis
Age group	18 yrs. and older
Gender	m/f
Type and severity of the medical condition	Acute and chronic moderate to severe
Range of time	Transient application Number of repeat exposures: daily for 3 months up to unlimited period of time
Intended use	Vivira helps users to reduce physiotherapeutically treatable pain of the musculoskeletal system

Nature and duration of contact to the human body	Vivira has no contact with the human body
Material in contact to the human body	None
Organs in contact with material	None
Tissue in contact with material	None
Fluids in contact with material	None

Table 1: Vivira App Product

Use

Patients go through an

- **introduction** to configure their training program. The patient confirms that from a medical point of view, he is able to train on his own. If he does not confirm this, he is requested to consult a doctor before using Vivira. Afterwards they will receive
- **daily specific exercises** tailored to their complaint region, their indication (if any) and their progression ability. Patients usually receive four exercises a day, each of which is accompanied by video instructions. According to the instructions, they are accompanied by another video in which the correct execution is demonstrated again and again. The exercises of one day form a training session, and for each exercise patients give
- **feedback** on pain and their ability to perform the exercise. This feedback determines the
- **progression of the exercises** in further course. Patients record their daily physical activity and weekly their progress in pain, quality of life and limitations. They perform a monthly movement test for stability, mobility and neuromuscular control of their locomotor system. The training and activity level, the training progress, as well as the course curves are visualized in
- **diagrams.**
- **Key data** about the program and the progress achieved can be exported at any time and shared with a doctor/therapist. Patients can manage their Vivira account, their training programs and their data directly in the app at any time.

Interventions

The Vivira training system was developed by the Vivira Medical Board in cooperation with the Vivira physiotherapy team. Chairman of the Vivira Medical Board is Dr. med. Markus

Klingenberg, who is an Ärztekammer certified specialist for Orthopedics and Orthopedic Surgery with Beta Klinik, Bonn, Germany (for qualification descriptions see CV at the end of this document). The following interventions are part of the program:

Independent practicing. Vivira enables independent practice in a home setting: independent of place and time patients receive daily 4 exercises, their correct execution being explained with video and text.

Personalized Progression. Patients start after stating their area of complaint, their indication (provided available) and their limitations, a correspondingly specific training program. The personalization of the exercise contents is patient feedback controlled from the first exercise performed. Via the feedback after each exercise, exercises are adapted to the patient's abilities on a daily basis. The progression algorithm allows patients to constantly practice in their range of abilities, so that they are neither under- nor overchallenged. At the beginning of use, the algorithm approaches the capability range "from below", so that no excessive demands are made. This minimizes the risk of AEs.

Functional training. The effectiveness of the exercises is provided by the functional exercise approach. Musculoskeletal functionality as a whole is trained according to the functional approach and may contribute to pain reduction. Focus is on mobility
(agility), stability (strength) and neuromuscular control (coordination).

Continuous cycle (exercises, progress recording, tests). Patients experience a continuous cycle of exercises, progress recording and movement tests. The cycle continues over the entire application time, over levels of difficulty, creates awareness for one's own change process, motivates and thus contributes to the adherence to therapy.

Education. Within the framework of education, the patient learns an adequate disease model. Thus, he is thought to be motivated to perform the upcoming exercises correctly and with a large compliance and adherence.

10. INVESTIGATIONAL PLAN

Overall Study Design and Plan-Description

The study presents observational data on the primary outcome of overall pain score reduction and the secondary outcomes of reporting interval-specific and stratum-specific pain score reductions, functional improvement and adherence to the home exercise program. Effectiveness is assessed with self-reported pain scores, assessed with a verbal-numerical

rating scale (VNRS), which has been established to be a reliable[8] and valid instrument[9] to capture pain score intensity as a patient-reported outcome measure (PROM). The primary hypothesis test for the assessment of pain score changes is a non-parametric, two-sided Skillings-Mack test, which has been outlined elsewhere in detail[10]. In brief, it allows the analysis of unbalanced and incomplete block designs with relevant missing data by design or by random. The functional assessment is developed on the basis of established orthopedic functional tests and employs the principles of functional regional interdependence[11]. To enable a patient-directed self-assessment, these tests are presented with audiovisual guidance and results are entered on a binary scale (i.e. test could be completed or test could not be completed). Through expert consensus of a panel of orthopedic surgeons and physical therapists, the weighted transformation of the functional tests was performed, so that discrete functional scores could be computed. For secondary analyses of pain scores and functional scores, a Wilcoxon signed-rank test, a Kruskall-Willis test and a one-way ANOVA were employed. Distributions were assessed using Bartlett's test. Corrections for familywise errors were performed using the Bonferroni procedure. Adherence was assessed on the basis of whether patients were actively using the home exercise program at pre-defined thresholds (two weeks, four weeks, eight weeks, twelve weeks), given they met the inclusion criteria, as outlined below. Patients were enrolled through a self-selection process.

Discussion of Study Design, including the Choice of Control

The data collected using the Vivira data base can naturally only be evaluated retrospectively. Data allowed an intraindivual control. Data values, before the start of treatment and values after the end of treatment are available, so that an intra-individual control is possible.

Selection of Study Population

Inclusion criteria

- Age >= 18 years
- Report of any applicable pain area (i.e. upper back, lower back, hip, or knee)
- Initial pain score (VNRS) > 0/10
- Completion of at least one exercise during the study period

Exclusion criteria

[8] Alghadir AH, Anwer S, Iqbal A, Iqbal ZA. Test-retest reliability, validity, and minimum detectable change of visual analog, numerical rating, and verbal rating scales for measurement of osteoarthritic knee pain. J Pain Res. 2018;11:851-856.

[9] Hjermstad MJ, Fayers PM, Haugen DF, et al. Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review. J Pain Symptom Manage. 2011;41(6):1073-1093.

[10] Chatfield M, Mander A. The Skillings-Mack test (Friedman test when there are missing data). Stata Journal. 2009;9(2):299-305.

[11] Schellenberg KL, Lang JM, Chan KM, Burnham RS. A clinical tool for office assessment of lumbar spine stabilization endurance: prone and supine bridge maneuvers. Am J Phys Med Rehabil. 2007;86(5):380-386.

- Patients who were not able to follow the exercise protocol

Removal of patients from therapy or assessment
This was done by decision by the patient. The use of the app was voluntary at all times during the exercises. Patients could stop using the app and thus the execution of the exercises at any time without giving reasons.

Treatments

Treatments administered
The level of the exercises depended on the feedback of the patient. The feedback thus leads to an extensive individualization of the therapy within the framework of the available interventions.

Identity of investigational product
Vivira App

Method of assigning patients to treatment groups
Patients have installed the product on a voluntary basis and have voluntarily decided to use it. If the patient has given his informed consent to the use of his data within the scope of the use, these data were stored in the database and were later used for the retrospective evaluation study.

Selection of level in the study
The app's software offers a feedback mechanism, which allows the level of the exercises to be adjusted individually to the patient.

Selection of timing for each patient
The app's software allows an individual timing by the patient.

Blinding
Patients were naturally not blinded by the product. The physicians evaluating the study had no relation to the patients and received the statistically evaluated data in fully anonymized form for evaluation.

Prior and concomitant therapy
Other accompanying therapy was not influenced by the therapy with the app.

Treatment compliance
Checking of compliance with treatment motion sequences was not possible.

Treatment adherence
The adherence could be tracked through the software of the app by means of the actual usage.

11. Efficacy Variables

Overview

Specific efficacy variables assessed

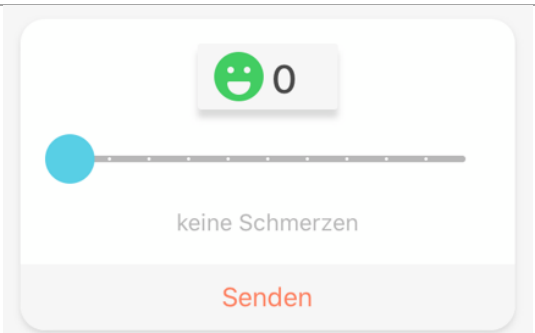
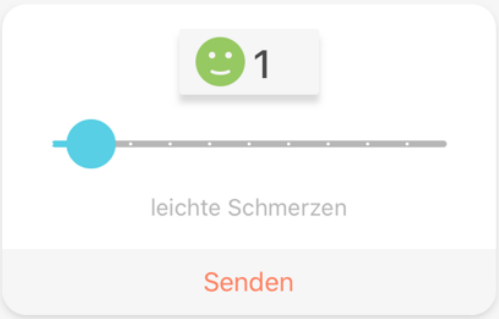
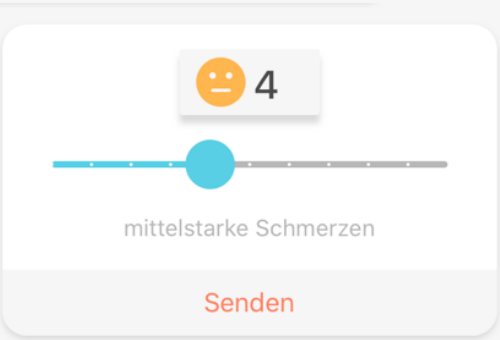
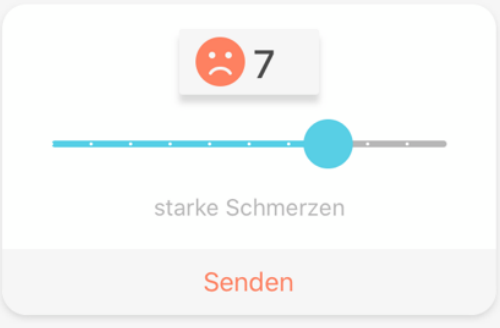
- Pain
- Mobility
- Stability
- Coordination
- Total Functional Score

No laboratory tests were conducted. Subsequently, all variables are tabulated:

Pain score

Pain intensity was measured using a verbal-numerical rating scale (VNRS). The following scale and description were used:

Scale range	Description	Image example

0	keine Schmerzen	
1-3	leichte Schmerzen	
4-6	mittelstarke Schmerzen	
7-9	starke Schmerzen	

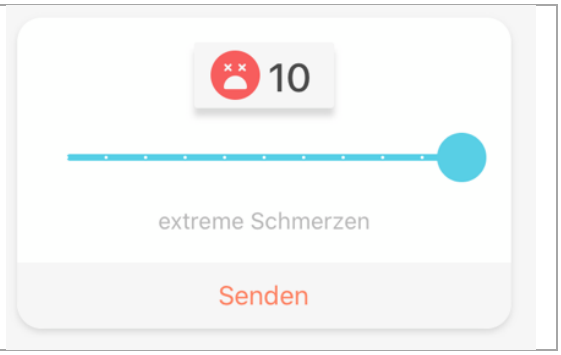
10	extreme Schmerzen	

Table 2: Overview of pain VNRS retrieved by the app

Mobility, Stability, Coordination and Total Functional Score

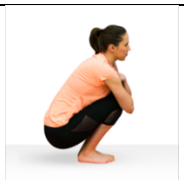
Mobility, stability, and coordination scores are calculated based on a patient´s reported ability to perform the following list of movements without pain and to meet the defined criteria. The scores for a patient feedback "yes" (confirming to meet the defined criteria without pain) are defined in the following table. Patient feedback "no" is scored with 0. Total score and individual scores for mobility, stability, and coordination are calculated. Score values range from 0 to 100. The movements within this assessment were selected and adapted (where required) from established movement screens:

- SFMA – selective functional movement assessment (Cook et al., 2006a, b[12])
- FMS – functional movement screen (Cook 2010[13])
- DFT - Dorsiflexion Lunge Test (Bennell et al., 1998[14])
- Other tests (Tong et al., 2014[15]; Magee, 2014[16]; Schellenberg et al., 2007[17]).

Nr	Movement	Image	Criteria	Total Score	Mobility Score	Stability Score	Coordination Score

[12] Cook, G., Burton, L., & Hogenboom, B. (2006a) 'The use of fundamental movements as an assessment of function - Part 1', NAJSPT, 1(2), pp- 62-72
Cook, G., Burton, L., & Hogenboom, B. (2006b). 'Pre-participation screening: the use of fundamental movements as an assessment of function - Part 2', NAJSPT, 1(3), pp. 132-139.
[13] Cook, G. Movement: Functional Movement Systems: Screening, Assessment and Corrective Strategies. Aptos, CA: Target Publications; 2010.
[14] Bennell, K.L., Talbot, R.C., Wajswelner, H., Techovanich, W., Kelly, D.H., Hall, A.J. Intra-rater and inter-rater reliability of a weight-bearing lunge measure of ankle dorsiflexion. Aust J Physiother. 1998;44(3):175-180.
[15] Tong, T.K., Wu, S., Nie, J. (2014) 'Sport-specific endurance plank test for evaluation of global core muscle function.' Phys Ther Sport, 15(1), pp. 58-63.
[16] Magee, D.J. (2014) Orthopedic Physical Assessment. St Louis, Mo: Saunders Elsevier.
[17] Schellenberg, K.L., Lang, J.M., Chan, K.M., Burnham, R.S. (2007) 'A clinical tool for office assessment of lumbar spine stabilization endurance: prone and supine bridge maneuvers.' Am J Phys Med Rehabil, 86(5), pp. 380-386.

1	Cervical flexion		Chin touches sternum	6,6	10	0	0
2a	Cervical rotation - left		Chin touches left collar bone	3,3	5	0	0
2b	Cervical rotation - right	See 2a	Chin touches right collar bone	3,3	5	0	0
3	Multisegmental Flexion		Fingertips touch ground with extended knees	6,6	10	0	0
4a	Multisegmental Rotation - left		Heterolateral shoulder is visible in mirror with mutisegmental rotation, no lateralflexion for compensation	3,3	5	0	0
4b	Multisegmental Rotation - right	See 4a	Heterolateral shoulder is visible in mirror with mutisegmental rotation, no lateralflexion for compensation	3,3	5	0	0
5	Multisegmental Extension (wall)		Heels, pelvis, shoulder blades and back of the head are in contact with wall simultaneously	6,6	10	0	0
6	Extension + full shoulder anteversion		Full abduction of shoulder is possible while heels, buttocks, shoulder blades, back of the head or hands stay in contact with wall	6,6	10	0	20
7a	Knee to Wall - left		Knee touches the wall with approx. 10 cm distance between toes and wall	3,3	5	0	0

7b	Knee to Wall - right	See 7a	Knee touches the wall with approx. 10 cm distance between toes and wall	3,3	5	0	0
8a	In-Line Lunge - left		Backward lunge with knee touching the ground is possible, no evasive movements	3,3	5	0	10
8b	In-Line Lunge - right	See 8a	Backward lunge with knee touching the ground is possible, no evasive movements	3,3	5	0	10
9	Deep Squat		Buttocks touch heels, knees stay together	6,6	10	0	20
10	Heelsit		Buttocks touch heels, knees stay together	6,6	10	0	0
11a	Bird Dog - left		Position held at least 10 seconds	3,3	0	10	10
11b	Bird Dog - right	See 11a	Position held at least 10 seconds	3,3	0	10	10
12	Forearm Plank		Position held at least 60 seconds	6,6	0	20	0
13a	Side Plank – left		Position held at least 30 seconds	3,3	0	10	0
13b	Side Plank - right	See 13a	Position held at least 30 seconds	3,3	0	10	0

14	Shoulder Bridge		Position held at least 60 seconds	6,6	0	20	0
15a	One Legged Stand - left		One legged stand held for at least 10 seconds, no evasive movements of upper body	3,3	0	10	10
15b	One Legged Stand - right	See 15a	One legged stand held for at least 10 seconds, no evasive movements of upper body	3,3	0	10	10
Maximum Score				**100**	**100**	**100**	**100**

Table 3: Overview of mobility, stability, coordination and total functional score retrieved by the app

Score evaluation. Resulting scores of total functional ability, mobility, stability and coordination are categorized as follows:

Score range	Ability within category
0-40	Low
41-70	Medium
71-90	High
91-100	Very high

Table 4: Score evaluation. Resulting scores of total functional ability, mobility, stability and coordination were categorized as indicated.

Application schedule

Patients were asked to exercise regularly, ideally daily but at least 3 times per week, using the training plan provided in the Vivira App. The duration of the exercise regimen is not limited, for the purpose of this study a 14-week window for analysis was chosen. The time of day for exercising was the patient's decision, no relation to meals had to be observed.

Persons responsible for the measurements

Patients inserted their data into the relevant part of the app´s CRF.

Frequency and timing of measurements

The following table displays how frequency and timing of measurements were performed:

Timing	Efficacy variable measured	Comment
Start of app usage (t=0)	Pain intensity (baseline)	This measurement was obligatory, it is the baseline pain score that is used as a reference
Every 7 days	Pain intensity (follow-up) Quality of life	Patients were encouraged to provide these measurements; however, measurement were not obligatory
Every 4 weeks	Total functional score Mobility score Strength score Coordination score	

Table 5: Frequency and timing of measurement

Persons responsible for evaluation of clinical outcomes

Prof. Dr. med. K. Weise, Prof. Dr. med. H.P. Zenner
Clinical Research Organization H.P. Zenner Clinical, Tübingen

Dr. med. Leo Benning, MPH
University Emergency Department, University Medical Center Freiburg

Procedures (blinding, anonymization)

Blinding was not feasible for the patients, but the physicians responsible for this report, were blinded because they received the data completely anonymously. The data were collected in the database of the app and then read out, pseudo anonymized, and made available for statistical analysis. At the end of statistical analysis, resulting metadata were completely anonymized and forwarded to the evaluators.

Appropriateness of measurements

The scientific basis of appropriateness was investigated for the primary efficacy variable pain. A number of well-validated instruments exists, including the

- Numeric Rating Scale (NRS)
- Visual Analogue Scale (VAS)
- Verbal Descriptor Scale
- Verbal Numerical Rating Scale (VNRS)

Among these the Visual Analogue Scales (Price et al. 1983[18]) are the gold standard. Outside a typical controlled research setting, however, assessment of pain is more complex because it is based not only on an individual's (i) perception of pain but also on his (ii) personal disability as well as his personal (iii) temporal (time pressure), (iv) spatial and (v) social circumstances. (ii)-(v) can in particular worsen compliance and thus drop off rate and reliability (Flaherty, 1996[19]). Thus, measurement of pain poses a challenge to clinicians, especially in the presently investigated motorically handicapped people in a remote setting on its own with the app. Using the presently investigated APP in a remote setting and thus in the absence of a personal support comprehension difficulty may add to non-compliance. Thus, Flaherty (1996[20]) postulated that pain should be assessed choosing instruments that take into consideration the current level of pain but also practical and personal factors. Therefore, in selecting an instrument for this study, we recognized the subjective nature of pain and the potential limitations of the motorically handicapped population targeted. Furthermore, potential incompliance reasons were taken into consideration as a high drop off rate would severely interfere with the test result reliability.

As early as 1983 Price et al. 1983[21] (N=50, high methodological and statistical quality) validated Visual Analogue Scales VAS as ratio scale measures for chronic and experimental pain. Chronic pain patients and healthy volunteers made VAS sensory and affective responses to 6 noxious thermal stimuli (43,45,47,48, 49 and 51 degrees C). The power functions were predictive of estimated ratios of sensation or affect produced by pairs of standard temperatures (e.g. 47 and 49*C), thereby providing direct evidence for ratio scaling properties of VAS. VAS sensory intensity responses to experimental pain, VAS sensory intensity responses to different levels of chronic pain, and direct temperature (experimental pain) matches to 3 levels of chronic pain and were all internally consistent, thereby demonstrating the valid use of VAS for the measurement of and comparison between chronic pain and experimental heat pain.

As already mentioned, apart from controlled research settings there may be practical limitations of the use of VAS and patient preferences that influence compliance and thus reliability. Patients requiring postinterventional physiotherapy may be handicapped. E.g. they

[18] Price, Donald D., Patricia A. McGrath, Amir Rafii, und Barbara Buckingham. „The validation of visual analogue scales as ratio scale measures for chronic and experimental pain". Pain 17, Nr. 1 (1983): 45–56.
[19] Flaherty, S. A. (1996). Pain measurement tools for clinical practice and research. Journal of the American Association of Nurse Anesthetists, 64, 133–140.
[20] Flaherty, S. A. (1996). Pain measurement tools for clinical practice and research. Journal of the American Association of Nurse Anesthetists, 64, 133–140.
[21] Price, Donald D., Patricia A. McGrath, Amir Rafii, und Barbara Buckingham. „The validation of visual analogue scales as ratio scale measures for chronic and experimental pain". Pain 17, Nr. 1 (1983): 45–56.

may suffer from a reduced motoricity. This may directly influence their capacity to handle analog scales. In addition, an uncomfortable environment and the absence of personal support might enhance the stress level and further influence compliance and thus reliability. Thus, a literature review was performed comparing VAS with the alternative instruments. Furthermore, the literature review investigates the reliability of pain assessment instruments in various stress situation or when used with patients with limited motoricity.

In a systematic literature review Hjermstad et al. 2011[22] compared Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults. Fifty-four of 239 papers were included. Eight versions of the NRS (NRS-6 to NRS-101) were used in 37 studies; a total of 41 NRSs were tested. Twenty-four different descriptors (15 for the NRSs) were used to anchor the extremes. When compared with the VAS and VRS, NRSs had better compliance in 15 of 19 studies reporting this, and were the recommended tool in 11 studies on the basis of higher compliance rates, better responsiveness and ease of use, and good applicability relative to VAS/VRS. Twenty-nine studies gave no preference. Many studies showed wide distributions of NRS scores within each category of the VRSs. Overall, NRS and VAS scores corresponded, with a few exceptions of systematically higher VAS scores.

In an additional study Price et al. 1994[23] (N=33, high methodological and statistical quality) provided a comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. Both VAS and numerical rating scales NRS produced reliably different stimulus response functions for pain sensation intensity as compared to pain unpleasantness and both provided consistent measures of experimental and clinical pain intensity.

Thus, Numerical Rating Scales and Verbal Rating Scales have been shown to be able to replace VAS. To investigate this aspect specifically, literature was searched dealing with pain assessment instruments in non-controlled settings or on patients characterized by limited motoricity.

Motoricity problems are well known in children. Thus, in a pediatric clinical study (McGrath et al. 1996 [24]) on practical limitations of pain scales and patient preference (N=123, high methodological and statistical quality) severe difficulty filling in a VAS was a result of fine motor problems or being immobilized, visual problems or comprehension difficulty totaling to 16 juvenile patients (13%). No patient in the study had comprehension difficulty with VNRS. The majority of patients (59.4%) found that using a pain scale helped them describe the severity of their pain. Twenty-two patients (17.89%) found VAS easier, 43 patients (34.96%) found VNRS

[22] Hjermstad, Marianne Jensen, Peter M. Fayers, Dagny F. Haugen, Augusto Caraceni, Geoffrey W. Hanks, Jon H. Loge, Robin Fainsinger, Nina Aass, und Stein Kaasa. „Studies Comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for Assessment of Pain Intensity in Adults: A Systematic Literature Review". *Journal of Pain and Symptom Management* 41, Nr. 6 (1. Juni 2011): 1073–93. https://doi.org/10.1016/j.jpainsymman.2010.08.016.

[23] Price, Donald D., Francis M. Bush, Stephen Long, und Stephen W. Harkins. „A Comparison of Pain Measurement Characteristics of Mechanical Visual Analogue and Simple Numerical Rating Scales". *Pain* 56, Nr. 2 (1. Februar 1994): 217–26. https://doi.org/10.1016/0304-3959(94)90097-3.

[24] McGrath, Patricia A., Cheryl E. Seifert, Kathy N. Speechley, John C. Booth, Larry Stitt, und Margaret C. Gibson. „A new analogue scale for assessing children's pain: an initial validation study". *Pain* 64, Nr. 3 (1996): 435–443.

easier to use. Fifty-eight patients (47.15%) said there was no difference. VAS and VNRS are not interchangeable in
assessing an individual patient's pain over time. VNRS has practical advantages over VAS in this setting.

The pain assessment using VNRS and VAS was investigated in a 3-month cross-sectional study by Ismail et al. 2015[25] (N=130, high methodological and statistical quality). They evaluated the agreement between VNRS and VAS in measuring acute pain in prehospital setting and to identify the preference among paramedics and patients. Their aim was to evaluate the agreement between VNRS and VAS in measuring acute pain in hospital and non-hospital setting and to identify the preference
among paramedics and patients. There was a strong correlation between VNRS and VAS at the scene ($r = 0.865$; $p < 0.001$), as well as on arrival at the hospital ($r = 0.933$; $p < 0.001$). Kappa coefficient values and Bland-Altman analysis indicates good agreement between both scales for measuring acute pain. VNRS was the preferred method to measure acute pain by patients and paramedics. As a conclusion VAS performs as well as VNRS in assessing pain. VAS and VNRS must not be used interchangeably to assess pain; either method should be used consistently.

Mohan et al. 2008[26] conducted a prospective observational study on the Visual Analogue Scale (VAS) and Verbal Numerical Rating Scale (VNRS) as pain assessment tools in an emergency department (N=123, moderate methodological quality, descriptive statistics). There was a strong correlation between VAS and VNRS ($r_s=0.93$). There was a better compliance for VNRS in those patients who expressed a preference for one pain scale over the other. VNRS has practical advantages over VAS in this setting.

Summary. VAS can be replaced by VNRS. VNRS is a validated instrument, that has practical advantages over VAS in non-laboratory settings.

Primary efficacy variable

Pain VNRS

Descriptive efficacy variables

Mobility, Stability, Coordination and Total Functional Score

Data Quality Assurance

Data provided by the patient are not lost and totally reflect the entered data. Part of the data depends entirely on the voluntary nature of data entry by patients - patients can discontinue

[25] Ismail, Ahmad K., Muhammad A. Abdul Ghafar, Noor Shaza A. Shamsuddin, Nurul A. Roslan, Hilwani Kaharuddin, und Nik A. Nik Muhamad. „The assessment of acute pain in pre-hospital care using verbal numerical rating and visual analogue scales". The Journal of emergency medicine 49, Nr. 3 (2015): 287–293.
[26] Mohan, Helen, John Ryan, Brendan Whelan, und Abel Wakai. „The end of the line? The Visual Analogue Scale and Verbal Numerical Rating Scale as pain assessment tools in the emergency department". Emergency Medicine Journal 27, Nr. 5 (2010): 372–375.

entering data at any time. Thus, data quality assurance for this subset of data over the entire process depends on the voluntary nature of patients´cooperation.

12. STUDY PATIENTS

Disposition of Patients

As the study population at hand was enrolled prior to the home exercise program being subject to a prescription by physicians and other authorized health care providers, the enrollment was primarily based on self-selection through out-of-pocket pay, or the use of a voucher code, as handed out over the period of the data collection to evaluate the digital home exercises at hand. 2'517 patients, who met the inclusion criteria, provided at least two data points necessary for the intraindividual control over a period of twelve weeks with measurements after two, four, eight, and twelve weeks

Protocol Deviations

Patients who entered the study even though they did not satisfy the entry criteria
None

Patients who developed withdrawal criteria during the study but were not withdrawn
None

Dropouts

There was a significant number of dropouts in the course of time. This was expected since the present study is no prospective study. The overall retention rate is 17% after two weeks, 10% after four weeks, 4% after eight weeks and 3% after twelve weeks. This large number of dropouts may introduce bias. Thus, drawing conclusions based only on patients who completed the study can be misleading.

13. EFFICACY EVALUATION

Data Sets Analysed

N=2517 patients were analyzed.

Baseline Characteristic

Demographic characteristics on sex, age, pain area and pain duration were constructed to investigate differences of age groups in pain duration and pain area. The sample population

consists of 63% female participants, the average age is 47.08 years (SD 14.61 years). As patients were allowed to skip the question about their duration, 1'065 data entries (42,3%) cannot be stratified according to pain duration, respectively. Baseline demographics are displayed in the subsequent Table 6.

		Reported Pain Area				Pain Duration			
	N	Lower Back	Upper Back	Hip	Knee	Acute	Subacute	Chronic	Not specified
All	**2.517**	**1.278**	**458**	**312**	**469**	**290**	**234**	**928**	**1.065**
Age Group									
18 - 35	**714**	312	196	58	148	110	85	243	276
36 - 45	**436**	255	71	52	58	59	46	145	186
46 - 55	**639**	340	118	94	87	69	48	236	286
56 - 65	**522**	268	46	82	126	38	36	230	218
66 - 75	**172**	82	23	22	45	13	18	68	73
75+	**25**	15	2	4	4	1	1	6	17
n/a	**9**	6	2	-	1	-	-	-	9
Sex									
Female	**1.586**	770	318	211	287	182	138	638	628
Male	**931**	508	140	101	182	108	96	290	437

Table *6*: Baseline characteristics of the study population.

14. Measurements of Treatment Adherence

Adherence to digital therapeutics has often been described prone to high loss to follow up. Adherence was assessed as retention rate on the basis of whether patients were actively using the home exercise program at pre-defined thresholds (two weeks, four weeks, eight weeks, twelve weeks), given they met the inclusion criteria.

The overall retention rate is 17% after two weeks, 10% after four weeks, 4% after eight weeks and 3% after twelve weeks. This high attrition is present in all sub-populations, no difference in loss to follow up patterns could be detected. Yet, a total attrition can be observed in patients with pain in the lower back and a non-specified pain duration, upper back with acute and subacute pain durations and knee with a non-specified pain duration (Table 7). Nonetheless, we see a tendency towards higher retention rates among patients with chronic pain (Figure 1).

Indication Subset	**Pain Duration**	**Initial (N, %)**	**Week 2 (N, %)**	**Week 4 (N, %)**	**Week 8 (N, %)**	**Week 12 (N, %)**
All	All	2517(100%)	418(17%)	255(10%)	107(4%)	68(3%)
Lower Back	All	1278(100%)	202(16%)	119(9%)	57(4%)	33(3%)
	Acute	144(100%)	30(21%)	16(11%)	10(7%)	5(3%)

	Subacute	107(100%)	26(24%)	17(16%)	4(4%)	3(3%)
	Chronic	443(100%)	120(27%)	69(16%)	39(9%)	23(5%)
	Not specified	584(100%)	26(4%)	17(3%)	4(1%)	2(1%)
Upper Back	All	458(100%)	81(18%)	46(10%)	17(4%)	9(2%)
	Acute	74(100%)	15(20%)	11(15%)	4(5%)	0(0%)
	Subacute	50(100%)	15(30%)	4(8%)	2(4%)	0(0%)
	Chronic	170(100%)	47(28%)	26(15%)	9(5%)	7(4%)
	Not specified	164(100%)	4(2%)	5(3%)	2(1%)	2(1%)
Hip	All	312(100%)	62(20%)	44(14%)	23(7%)	7(2%)
	Acute	27(100%)	6(22%)	2(7%)	2(7%)	1(4%)
	Subacute	29(100%)	11(38%)	8(28%)	4(14%)	1(3%)
	Chronic	138(100%)	42(30%)	33(24%)	17(12%)	4(3%)
	Not specified	118(100%)	3(3%)	1(1%)	0(0%)	1(1%)
Knee	All	469(100%)	73(16%)	46(10%)	27(6%)	19(4%)
	Acute	45(100%)	9(20%)	3(7%)	1(2%)	2(4%)
	Subacute	48(100%)	10(21%)	6(12%)	2(4%)	3(6%)
	Chronic	177(100%)	51(29%)	31(18%)	22(12%)	14(8%)
	Not specified	199(100%)	3(2%)	6(3%)	2(1%)	0(0%)

Table 7: User Retention across indication subsets and reported pain duration.

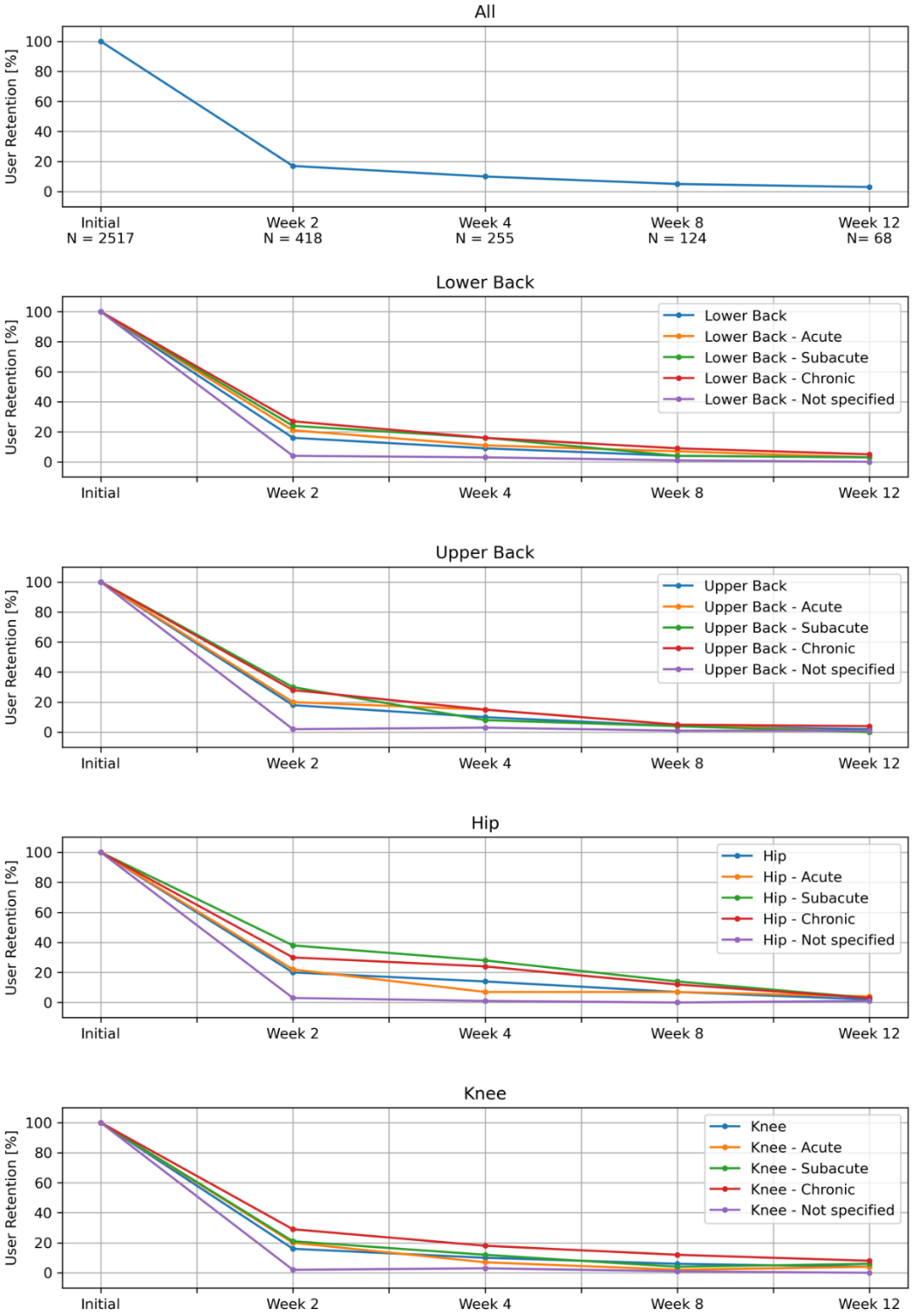


Figure 1: Retention rate for different pain areas and durations.

15. Efficacy Results

Primary efficacy variable pain

Overall analysis

Prior to the formation of indication-specific strata, we see a substantial reduction in self-reported pain scores across all two, four, eight, and twelve weeks, $T(2516) = 2728.27$, $p < .05$. Self-reported pain scores at the start were on average 5.19 out of 10 (SD = 1.96), after two weeks 3.72 out of 10 (SD = 2.06), after four weeks 3.39 out of 10 (SD = 2.35), after eight weeks 3.19 out of 10 (SD = 2.44), and after twelve weeks 3.35 out of 10 (SD = 2.38). These differences are illustrated in the subsequent Figure 2 and Table 8:

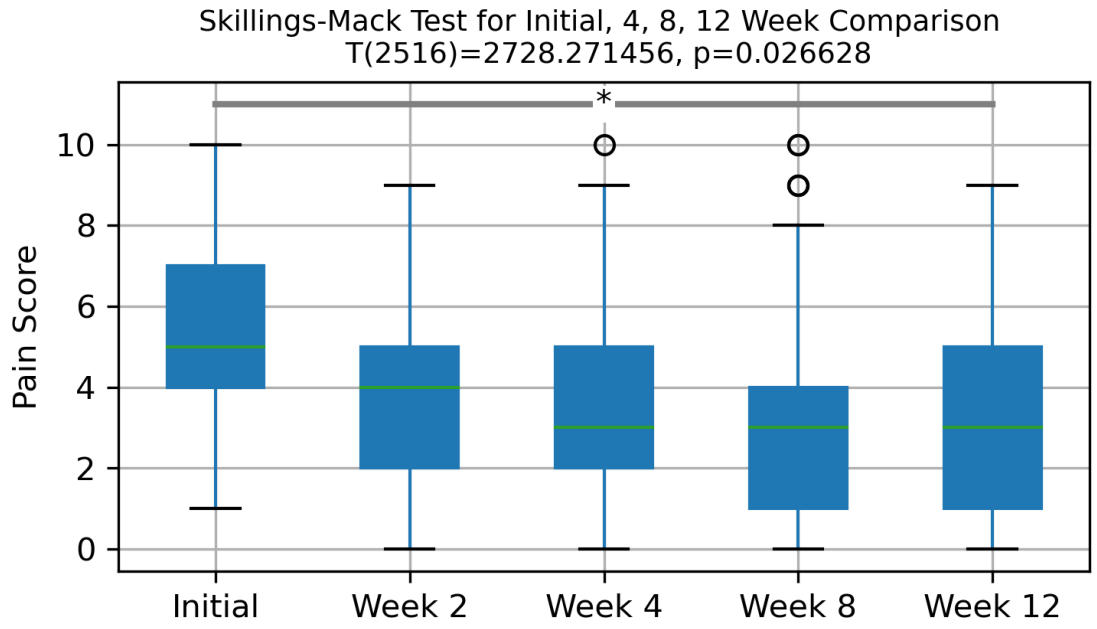


Figure 2: Average self-reported pain score for each retention time period for all pain areas. Center line (green), median; boxplots limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers; p < 0.05 = *, p < 0.005 = **, p < 0.0005 = *** for the Skillings-Mack Test.

Pain Area and Duration	Initial			Week 2			Week 4			Week 8			Week 12			Skillings-Mack Test			
	N	Mean	STD	N	Mean	STD	N	Mean	STD	N	Mean	STD	N	Mean	STD	Test Values	P Values	DF	Adjusted P
All	2.517	5,19	1,96	418	3,72	2,06	255	3,39	2,35	124	3,19	2,44	68	3,35	2,38	2.728,27	0,001	2.516	*
Lower Back	1.278	5,33	1,98	202	3,97	2,04	119	3,63	2,38	57	3,58	2,41	33	4,12	2,63	1.361,13	0,039	1.271	0,823
Acute	144	4,47	1,76	30	3,23	2,10	16	2,19	1,56	10	2,60	2,12	5	2,80	3,27	156,39	0,210	143	1,000
Subacute	107	5,00	1,54	26	3,54	1,73	17	3,71	1,99	4	3,00	1,63	3	3,00	1,00	115,34	0,252	106	1,000
Chronic	443	5,37	1,84	120	4,20	2,05	69	4,12	2,39	39	4,00	2,52	23	4,35	2,62	523,17	0,003	439	0,073
Not specified	584	5,57	2,13	26	4,19	2,02	17	2,94	2,77	4	2,50	2,08	2	6,50	0,71	571,83	0,588	580	1,000
Upper Back	458	5,19	1,84	81	3,65	2,11	46	2,91	2,03	17	3,65	2,98	9	3,67	2,50	487,45	0,157	457	1,000
Acute	74	4,57	1,76	15	3,13	2,53	11	3,18	1,89	4	3,50	2,52	-						
Subacute	50	4,96	1,74	15	4,27	2,22	4	2,25	2,22	2	1,50	2,12	-						
Chronic	170	5,15	1,73	47	3,57	1,93	26	2,81	2,00	9	3,00	2,40	7	2,86	2,04	187,02	0,163	169	1,000
Not specified	164	5,57	1,93	4	4,25	2,22	5	3,40	2,79	2	9,00	1,41	2	6,50	2,12	159,81	0,556	163	1,000
Hip	312	5,21	1,97	62	3,87	2,08	44	3,93	2,43	23	3,04	2,46	7	3,14	2,04	353,05	0,050	311	1,000
Acute	27	4,48	1,70	6	3,00	2,00	2	4,00	-	2	0,50	0,71	1	5,00					
Subacute	29	4,31	1,69	11	4,09	1,97	8	4,00	1,51	4	3,75	2,99	1	1,00					
Chronic	138	5,22	1,86	42	3,93	2,12	33	3,94	2,73	17	3,18	2,38	4	2,75	2,06	174,65	0,016	137	0,345
Not specified	118	5,58	2,11	3	4,00	2,65	1	3,00		-			1	5,00					
Knee	469	4,80	1,99	73	2,97	1,91	46	2,72	2,33	27	2,22	1,91	19	1,95	1,18	508,86	0,088	467	1,000
Acute	45	4,44	1,93	9	2,22	1,56	3	1,00	1,73	1	1,00		2	2,00	-				
Subacute	48	4,48	1,54	10	3,80	1,32	6	1,00	0,89	2	1,50	0,71	3	2,00	1,00	48,58	0,409	47	1,000
Chronic	177	4,57	1,92	51	3,06	2,01	31	2,97	2,24	22	2,41	2,02	14	1,93	1,33	217,42	0,016	175	0,338
Not specified	199	5,17	2,10	3	1,00	1,00	6	4,00	3,03	2	1,50	2,12	-						

Table 8: Self-reported pain scores and changes across indication subsets and reported pain duration by retained days. $p < 0.05$ = *, $p < 0.005$ = **, $p < 0.0005$ = *** for the Skillings-Mack Test, adjusted p values were calculated using Bonferroni corrections.

Post-hoc analysis comparing of sequential data entry points

To investigate the effect of different assessment times and, consequently, different durations of exposure to the digital home exercise program, we calculated further post-hoc tests. We used the Bonferroni method to adjust for family-wise error. First, we calculated a Wilcoxon signed-rank test to investigate to what degree a change in pain reduction occurred in patients that provided self-reported data at the initial assessment and after using "Vivira" for two weeks. We found a significant difference between the initial assessment (median = 5.0) and the assessment after two weeks (median = 4.0), T (679) = 22092, $p < 0.001$. Second, we calculated a Kruskal-Wallis' test showing that the self-reported pain values differed significantly between the initial (Median = 5.0), two-week (Median = 3.0), and four-week assessments (Median = 3.0), T (254) = 86.18, $p < 0.001$. Third, we calculated a Kruskal-Wallis' test showing that the self-reported pain values differed significantly between the initial (Median = 5.0), two-week (Median = 3.0), four-week (Median = 3.0), and eight-week assessments, T (88) = 34.12, $p < 0.001$. Finally, we calculated a one-way ANOVA (since this subsample was normally distributed and had equal variance indicated by Bartlett's test) showing a non-significant difference in self-reported pain for the initial (mean = 4.83, std = 2.08), two-week (mean = 3.63, std = 2.33), four-week (mean = 3.31, std = 2.46), eight-week (mean = 3.09, std = 2.54), and twelve-week (mean = 3.09, std = 2.66) assessments, $F(34) = 3.17$, $p = 0.06$. The subsequent Figure 3 illustrates these findings and highlights that, given the relevant loss to follow up, shorter exercise periods also show an overall effectiveness on pain score reduction.

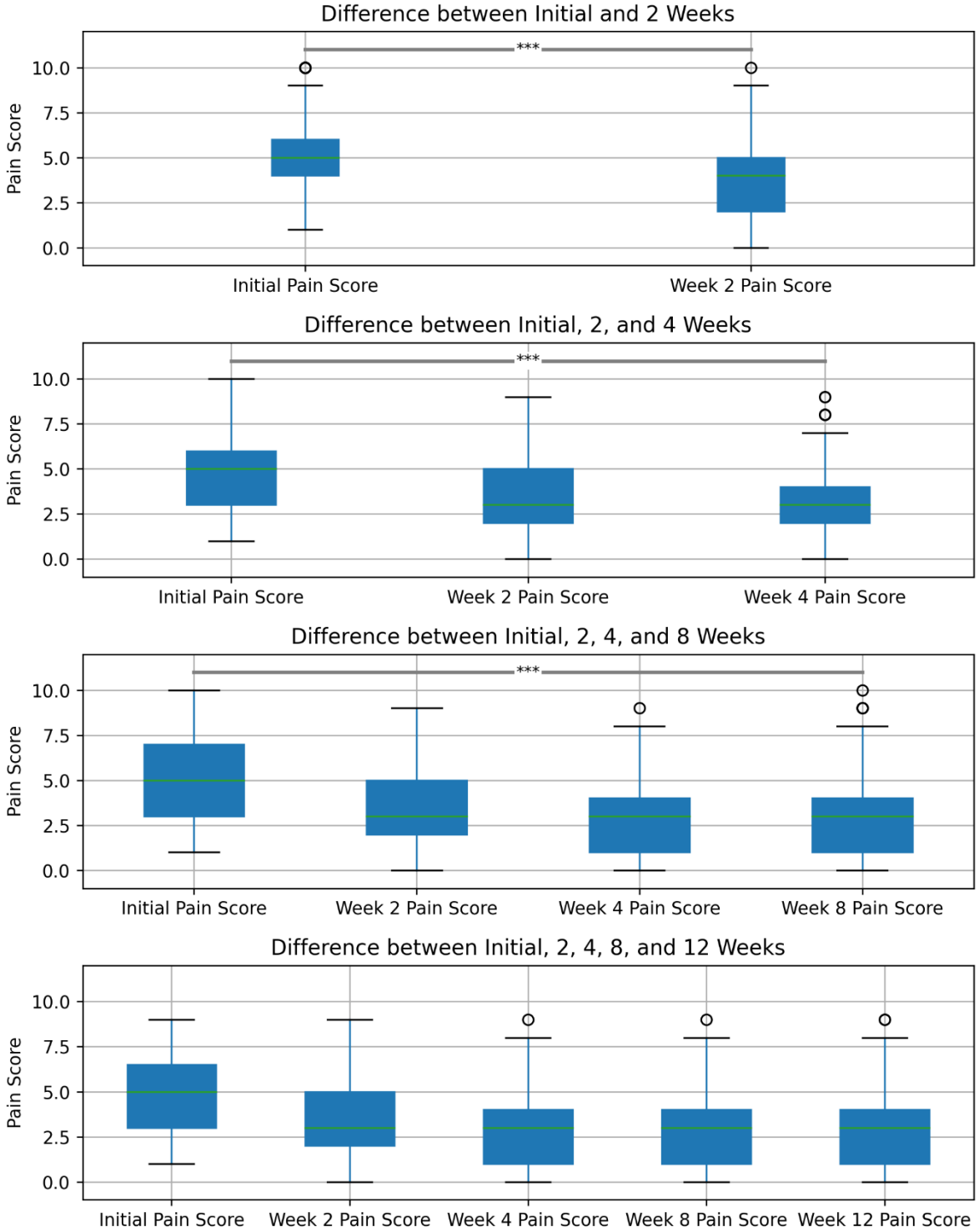


Figure 3: Post-hoc results for self-reported pain scores when comparing different assessment times. Center line (green), median; boxplots limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers; $p < 0.05$ = *, $p < 0.005$ = **, $p < 0.0005$ = *** for the Wilcoxon signed-rank test

(initial and two weeks), the Kruskall-Willis test (initial, two and four weeks and initial, two, four and eight weeks) and the one-way ANOVA (initial, two, four, eight and twelve weeks).

Efficacy subset analysis

Stratum-specific changes in pain intensity

After stratifying the available data for pain area and pain duration as a secondary analysis, we see a comparable response pattern across all pain areas. Patients with lower back pain report a reduction of their initial pain score from 5.33 to 4.12 after twelve weeks of exercises, $T(1271) = 1361.13$, $p = .823$. The sub-population of patients with chronic lower back pain show a particularly highlighted improvement from 5.37 to 4.35, T(439) = 523.17, p = 0.073. Similarly, patients with upper back pain report a reduction of their pain intensity from 5.19 to 3.67 after the completion of the exercise program, T(457) = 478.45, p = 1.0. On a comparable trajectory is the pain score change in patients with hip pain. Here, we see a reduction from a baseline pain score of 5.21 to 3.14 after twelve weeks, T(311) = 353.05, p = 1.0. Lastly, patients with knee pain improve from a baseline of 4.8 to 1.95 after completing the exercise program, T(467) = 508,86, p = 1.0. As the employed Skillings-Mack test does not return values for lacking blocks, no sub-stratum analyses for acute and subacute upper back pain, acute, subacute and non-specified hip pain and acute and non-specified knee pain are reported (Figure 4).

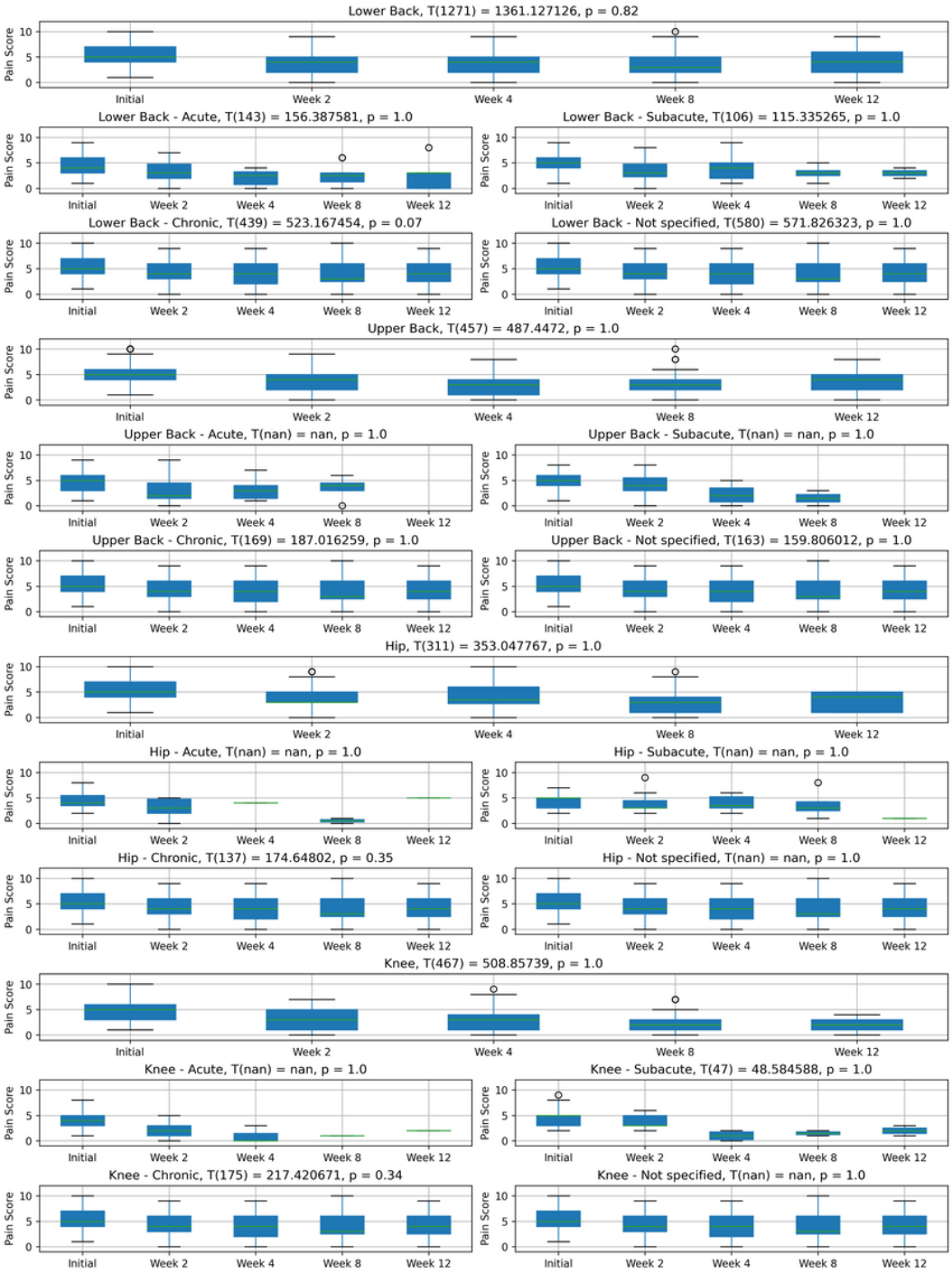


Figure 4: Average self-reported pain score for each retention time period for all pain areas. Center line (green), median; boxplots limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers; $p < 0.05$ = *, $p < 0.005$ = **, $p < 0.0005$ = *** for the Skillings-Mack Test. 'nan' = not a number, null value.

Descriptive variables/Functional scores

Functional scores

As secondary outcome, the improvement of a set of functional scores was assessed. Lower and upper back showed significant improvement for strength and mobility as well as total functional score (Table 9). This finding is consistent over all intervals of submitted functional scores studied, with the exception of upper back, which does not achieve significant improvement in strength score between their first and fourth submission (Table 9). For coordination upper back and lower back did not show significant improvement across any intervals of submitted scores studied, with the exception of upper back between first and fourth submission of functional scores where a significant improvement in coordination score is observed (Table 9). Knee and hip show significant improvement in mobility and coordination as well as total functional score between first and second submission of functional scores but do not show significant improvement in strength across any completed submission (Table 9). For hip and knee no significant improvement could be shown for mobility, coordination and total functional score could be shown between first and third and first and fourth submission of functional scores (Table 9).

		Total Score					Strength Score					Mobility Score					Coordination Score				
Matched Comparison	Pain Area	N	Retained Days (IQR)	Initial (IQR)	Last (IQR)	Test	N	Retained Days (IQR)	Initial (IQR)	Last (IQR)	Test	N	Retained Days (IQR)	Initial (IQR)	Last (IQR)	Test	N	Retained Days (IQR)	Initial (IQR)	Last (IQR)	Test
1st and 2nd Entry																					
	Lower Back	132	29 (20,5; 38,5)	60 (43; 75)	71,5 (53; 81,5)	***	132	29 (20,5; 38,5)	60 (30; 80)	70 (40; 100)	***	132	29 (20,5; 38,5)	60 (47,5; 80)	70 (55; 80)	***	132	29 (20,5; 38,5)	70 (40; 80)	80 (55; 80)	ns
	Upper Back	38	29 (20,5; 38,5)	65 (43; 80)	71,5 (60; 83)	***	38	29 (20,5; 38,5)	60 (40; 80)	70 (60; 100)	*	38	29 (20,5; 38,5)	62,5 (50; 75)	70 (60; 90)	***	38	29 (20,5; 38,5)	80 (50; 80)	80 (60; 100)	ns
	Hip	40	29 (20,5; 38,5)	67 (43; 77)	70 (55; 80)	*	40	29 (20,5; 38,5)	60 (40; 100)	80 (55; 100)	ns	40	29 (20,5; 38,5)	60 (45; 77,5)	70 (50; 80)	*	40	29 (20,5; 38,5)	60 (35; 80)	80 (50; 85)	*
	Knee	47	29 (20,5; 38,5)	70 (50; 83)	80 (57; 87)	**	47	29 (20,5; 38,5)	70 (50; 90)	80 (60; 100)	ns	47	29 (20,5; 38,5)	60 (50; 80)	70 (55; 85)	**	47	29 (20,5; 38,5)	60 (40; 80)	70 (50; 80)	*
1st and 3nd Entry																					
	Lower Back	48	59 (48; 80)	60 (43; 75)	78,5 (60; 87)	***	48	59 (48; 80)	60 (30; 80)	80 (60; 100)	***	48	59 (48; 80)	60 (47,5; 80)	75 (60; 85)	**	48	59 (48; 80)	70 (40; 80)	80 (60; 100)	ns
	Upper Back	15	59 (48; 80)	65 (43; 80)	73 (63; 87)	*	15	59 (48; 80)	60 (40; 80)	80 (60; 100)	*	15	59 (48; 80)	62,5 (50; 75)	75 (60; 90)	*	15	59 (48; 80)	80 (50; 80)	80 (60; 100)	*
	Hip	16	59 (48; 80)	67 (43; 77)	60 (41,5; 81,5)	ns	16	59 (48; 80)	60 (40; 100)	60 (45; 95)	ns	16	59 (48; 80)	60 (45; 77,5)	65 (40; 82,5)	ns	16	59 (48; 80)	60 (35; 80)	70 (50; 90)	ns
	Knee	20	59 (48; 80)	70 (50; 83)	80 (73; 83)	ns	20	59 (48; 80)	70 (50; 90)	80 (60; 100)	ns	20	59 (48; 80)	60 (50; 80)	80 (72,5; 82,5)	ns	20	59 (48; 80)	60 (40; 80)	80 (60; 80)	ns
1st and 4th Entry																					
	Lower Back	25	88,5 (72; 112)	60 (43; 75)	80 (67; 87)	*	25	88,5 (72; 112)	60 (30; 80)	80 (60; 100)	*	25	88,5 (72; 112)	60 (47,5; 80)	70 (65; 90)	*	25	88,5 (72; 112)	70 (40; 80)	80 (80; 80)	ns
	Upper Back	8	88,5 (72; 112)	65 (43; 80)	81,5 (67; 96,5)	*	8	88,5 (72; 112)	60 (40; 80)	80 (50; 100)	ns	8	88,5 (72; 112)	62,5 (50; 75)	82,5 (75; 95)	*	8	88,5 (72; 112)	80 (50; 80)	80 (60; 100)	ns
	Hip	5	88,5 (72; 112)	67 (43; 77)	67 (63; 80)	ns	5	88,5 (72; 112)	60 (40; 100)	60 (60; 80)	ns	5	88,5 (72; 112)	60 (45; 77,5)	70 (65; 70)	ns	5	88,5 (72; 112)	60 (35, 80)	80 (60; 100)	ns
	Knee	13	88,5 (72; 112)	70 (50; 83)	80 (73; 87)	ns	13	88,5 (72; 112)	70 (50; 90)	90 (60; 100)	ns	13	88,5 (72; 112)	60 (50; 80)	80 (70; 85)	ns	13	88,5 (72; 112)	60 (40; 80)	80 (70; 80)	ns

Table 9: Self-reported functional scores and changes across indication subsets by retained days. Matched comparisons between the first and second, first and third, and first and fourth completed functional assessment. Adjusted for familywise error, $p < 0.0167$ = *, $p < 0.00167$ = **, $p < 0.000167$ = ***, using Bonferroni

Conclusion

Pain score analyses showed

- a statistically significant reduction of pain over time through the use of the App.
- Consistent users of the app show a better pain reduction than less frequent users.
- Effects were
    - statistically significant,
    - of medium to large size and achieved a
    - clinically sufficient clinical effect strength.

16.Statistical methods/Analytics

Descriptive statistical methods were used in the analysis of these data. Quantitative variables were described by measures of location (means, medians) and dispersion (standard deviations). Moreover, changes from baseline were calculated and the pertinent measures of location and dispersion were determined, too. Furthermore, 95 % confidence intervals were calculated for the changes from baseline and the effect sizes (absolute changes divided by common standard deviation) were additionally determined. For qualitative variables the corresponding absolute and relative frequencies were presented. Moreover, the results were presented by graphical methods.

The primary hypothesis test for the assessment of pain score changes is a non-parametric, two-sided Skillings-Mack test, which has been outlined elsewhere in detail[27]. In brief, it allows the analysis of unbalanced an incomplete block designs with relevant missing data by design or by random.

For secondary analyses of pain scores and functional scores, a Wilcoxon signed-rank test, a Kruskall-Willis test and a one-way ANOVA were employed. Distributions were assessed using Bartlett's test. Corrections for familywise errors were performed using the Bonferroni procedure.

Adjustments for Covariates. Covariate analyses were not performed.

Handling of Dropouts or Missing Data. Dropouts were not counted as therapy failures but the remaining smaller group was used for further calculations. The same applies to missing data. The authors are aware of the fact that ignoring the patients who dropped out of the study and drawing conclusions based only on patients who completed the study can be misleading. A large number of dropouts may introduce bias. Details on the handling of dropouts and missing data are described in the respective chapters.

[27] Chatfield M, Mander A. The Skillings-Mack test (Friedman test when there are missing data). Stata Journal. 2009;9(2):299-305.

Interim Analyses and Data Monitoring. No interim analysis was performed. In this retrospective study data monitoring without codebreaking was n/a.

Multicentre Studies. Not applicable to this observational study. Patients were enrolled decentralized due to the remote nature of the exercise program. Access to the program was granted through payment or voucher code activation; a regional stratification was not feasible.

Multiple Comparisons/Multiplicity. Alpha adjustments to correct for familywise errors were performed using Bonferroni's procedure, where applicable.

Use of an "Efficacy Subset" of Patients. Not applicable, for the rationale of exploratory stratifications, see respective chapter.

Active-Control Studies Intended to Show Equivalence: Not applicable due to the observational nature of this study. Refer to other study reports for confirmatory trial results.

Examination of Subgroups.
As outlined, a stratification was carried out for the different body areas and pain durations examined; respective stratum-specific estimates are presented.

Tabulation of individual response data

For data protection reasons, the data was evaluated in pseudonymized form. Thus it is not possible to present the results of individual patients. K-anonymity is assumed to be given as a high homogeneity of data entries creates high barriers for any re-identification of participants.

Dose and relationships to response

This was investigated through multiple follow-up time points and the presentation of interval-specific endpoints. See the relevant result chapter.

Verum treatment (Vivira app) and other therapy (physiotherapy) interactions

Not applicable due to the observational nature of this study. Hence, no causal inference could be made.

By-patient displays

Because the data was evaluated in pseudonymized form for data protection reasons, it is not possible to present the results of individual patients individually. K-anonymity is assumed to be given as a high homogeneity of data entries creates high barriers for any re-identification of participants.

17. Efficacy discussion and conclusions

A digital home exercise can lead to significant improvements in pain scores

As exercise is known to effectively address unspecific and degenerative musculoskeletal pain[28], a digitally guided home exercise program was a priori considered a practical therapeutic intervention to address this spectrum of conditions. The overall analysis of the dataset at hand supports this assumption and shows a significant improvement of self-reported pain scores, based on a VNRS. Additional post-hoc analyses show significant improvements between the initial and the two week assessment, the initial, the two week and the four week assessment and the initial, the two week, the four week and the eight week assessment. Yet, it fails to show significant improvements between all assessment time points. We conclude from these analyses that an indicator for an overall improvement in pain scores is given and that shorter periods of exposure to the home exercise program yield significant pain score improvements over the abbreviated time points (i.e. up until eight weeks). Yet, conclusions based on this dataset need to be considered carefully, as a high rate of attrition is prone to bias.

Secondary analyses of sub-population do not yield relevant pain score reductions

An exploratory stratification across different pain areas (i.e. upper back, lower back, hip and knee) and different pain durations (i.e. acute, subacute and chronic pain) did not yield significant improvements in pain scores reported. However, to perform this analysis correctly, repeated corrections for familywise errors were required. Therefore, a significantly lower alpha level had to be applied. From the insignificant improvements, however, we see a tendency towards a relevant improvement in pain scores for lower back (unadjusted $p = 0.039$), hip (unadjusted $p = 0.05$) and knee (unadjusted $p = 0.088$). These data suggest a more nuanced response to a home exercise program across different pain areas, but the available data did not provide a sufficient density to thoroughly investigate this issue.

Functional improvements show a differential pattern

With the exception of hip and knee, significant improvements in strength and mobility could be detected between the first and the second assessment of the functional ability. Patients with hip and knee pain, however, show a significant response in increasing their coordination. This indicates a secondary benefit of the examined digital home exercise program. Interestingly, patients with lower back pain improve particularly sustained over an extended period of time

[28] Foster NE, Anema JR, Cherkin D, et al. Prevention and treatment of low back pain: evidence, challenges, and promising directions. The Lancet. 2018;391(10137):2368-2383.

Hayden JA, van Tulder MW, Tomlinson G. Systematic review: strategies for using exercise therapy to improve outcomes in chronic low back pain. Ann Intern Med. 2005;142(9):776-785.

van Gool CH, Penninx BW, Kempen GI, et al. Effects of exercise adherence on physical function among overweight older adults with knee osteoarthritis. Arthritis Rheum. 2005;53(1):24-32.

(median follow up of 88,5 days, IQR 72-112) in the dimensions of strength and mobility. We interpret this as an indicator for a differential functional response to the respective exercise programs. As the transformation of the functional test results (i.e. test could be completed successfully or test could not be completed successfully) into a discrete score (i.e. mobility strength coordination and total score) is solely based on expert consensus, a thorough validation of the assessment is required. Therefore, and because of the limited data availability a careful interpretation of these results is warranted.

Adherence rates are within the expected range of a digital therapeutic

Adherence rates to digital therapeutics have proven to show high rates of attrition. Baumel et al.[29], for example, report an average adherence to mental health digital therapeutics of < 10% after 30 days of use. Similarly, Fleming et al.[30] report in a systematic review on the intensity of use of digital therapeutics in mental health, a sustained use (i.e. completion of a program or continuation for more than 6 weeks) between 0.5-28.6%. The here presented retention rates are within this spectrum; only the spectrum of hip pain reaches a retention rate of 14% after four weeks. After twelve weeks, i.e. after the completion of the exercise program, an average retention rate of 3% could be demonstrated.

Conclusion

Innovative therapeutic means are required to address the increasing burden of disease from musculoskeletal conditions. This study presents early observational use data on the effectiveness of a digital home exercise program on the overall self-reported pain score reduction and demonstrates significant improvement in its primary analysis. Yet, stratum-specific pain reductions did not reach the adjusted level of significance. Significant functional improvements, particularly in the domains of strength and mobility, could be demonstrated for upper and lower back pain, but not for hip and knee pain. Yet, coordination improved significantly in patients with hip and knee pain. Interestingly, chronic back pain profited from the extended use and showed significant increases in strength and mobility scores after a median of 88.5 days. Adherence shows to be low, but within the spectrum of the available literature. Nonetheless, the presented data warrant a careful interpretation and further analyses are required to substantiate the early indicators of a therapeutic benefit of the examined digital home exercise program.

[29] Baumel A, Muench F, Edan S, Kane JM. Objective User Engagement With Mental Health Apps: Systematic Search and Panel-Based Usage Analysis. J Med Internet Res. 2019;21(9):e14567.

[30] Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the Trial: Systematic Review of Real-World Uptake and Engagement With Digital Self-Help Interventions for Depression, Low Mood, or Anxiety. J Med Internet Res. 2018;20(6):e199.

18. SAFETY EVALUATION

Adverse Events (AEs)

The app's software does not offer the option of an active surveillance of AEs.
PMS includes the option to record complaints and repairs. There is no field service that visits users. When complaints and repairs were returned, there was no indication of AE relevant hazards and AEs. It was also not necessary to file a notification of hazard or AEs with the supervisory authorities. There was also no recall action. The preclinical risk analysis revealed the following foreseeable hazards and AEs:

Potential hazards

a) Incorrectly conducting an exercise or conducting an unsuited exercise, due to

i) misunderstanding of the instructions provided by Vivira (unclear handling/ operation)

ii) non-compliance with instructions on intended use (incorrect user-feedback) (unclear handling/ operation)

b) The contraindication for use is the same as for any training program: If users are not in a health condition that enables them to train on their own, without the presence of a doctor or therapist, then they shall not use Vivira

b) There is a risk that users incorrectly configure their training programs by accidentally providing incorrect answers to onboarding or other questions or by providing incorrect configuration answers on purpose.

c) There is a risk that software updates go unnoticed by the user, leading to users working with outdated, possibly faulty versions of the software

d) Users need to be able to follow a simple smartphone application communication interface. If they are not able to, and therefore provide incorrect input to the Vivira application, they may incorrectly configure their training program (unclear handling/ operation)

d) Distractions in the home (where Vivira is meant to be used) can occur

The subsequent table summarizes these hazards und describes the foreseeable AEs:

Potential hazard	Potential adverse reaction (AR)	Potential serious AR (y/n)
a) Incorrectly conducting an exercise or conducting an unsuited exercise, due to i) misunderstanding of the instructions provided by Vivira (unclear handling/ operation)	Injury to bone or connective tissue	n
ii) non-compliance with instructions on intended use (incorrect user-feedback) (unclear handling/ operation)	Injury to bone or connective tissue	n
b) If users are not in a health condition that enables them to train on their own, without the presence of a doctor or therapist, then they shall not use Vivira	Injury to bone or connective tissue. Specifically: risk of e.g., luxation if users with artificial joints use Vivira espite the inadequacy of condition	y

a) There is a risk that users incorrectly configure their training programs byi) accidentally providing incorrect answers to Onboarding questions or other questions	Injury to bone or connective tissue	n
ii) by providing incorrect configuration answers on purpose	Injury to bone or connective tissue	n
b) There is a risk that software updates go unnoticed by the user, leading to users working with outdated, possibly faulty versions of the software	Injury to bone or connective tissue	n
Users need to be able to follow a simple smartphone application communication interface. If they are not able to, and therefore provide incorrect input to the Vivira application, they may incorrectly configure their training program (unclear handling/ operation)	Injury to bone or connective tissue	n
Distractions in the home (where Vivira is meant to be used) can occur	Injury to bone or connective tissue	n
Currently, users with certain disabilities, cannot use Vivira, such as blind people	n/a	n/a

Table 10: Hazards and foreseeable AEs

Safety Conclusions

From the PMS data, authority databases data, data from preclinical risk analysis and expert opinion it can be concluded that the risk of AEs is low.

19. DISCUSSION AND OVERALL CONCLUSION

To investigate efficacy of the Vivira app a retrospective, intraindividual controlled study with N=2517 patients suffering from pain of the musculoskeletal system was performed. Descriptive statistical methods were used in the analysis of these data. Quantitative variables were described by measures of location (means, medians) and dispersion (standard deviations). Moreover, changes from baseline were calculated and the pertinent measures of location and dispersion were determined, too. Furthermore, 95 % confidence intervals were calculated for the changes from baseline and the effect sizes were additionally determined. For qualitative variables the corresponding absolute and relative frequencies were presented.

The overall analysis of the dataset at hand supports the assumption that Vivira´s digital home exercises can lead to significant improvements in pain scores. Vivira shows a significant improvement of self-reported pain scores, based on a VNRS. Additional post-hoc analyses show significant improvements between the initial and the two week assessment, the initial, the two week and the four week assessment and the initial, the two week, the four week and the eight week assessment.

An exploratory stratification across different pain areas (i.e. upper back, lower back, hip and knee) and different pain durations (i.e. acute, subacute and chronic pain) showed statistically insignificant improvements. We see a tendency towards a relevant improvement in pain scores for lower back (unadjusted p = 0.039), hip (unadjusted p = 0.05) and knee (unadjusted p = 0.088).

With the exception of hip and knee, significant improvements in strength and mobility could be detected between the first and the second assessment of the functional ability. Patients with hip and knee pain, however, show a significant response in increasing their coordination. This indicates a secondary benefit of the examined digital home exercise program. Interestingly, patients with lower back pain improve particularly sustained over an extended period of time (median follow up of 88,5 days, IQR 72-112) in the dimensions of strength and mobility.

Conclusion

This study presents early observational use data on the effectiveness of a digital home exercise program on the overall self-reported pain score reduction and demonstrates significant improvement in its primary analysis. Significant functional improvements, particularly in the domains of strength and mobility, could be demonstrated for upper and lower back pain, but not for hip and knee pain. Yet, coordination improved significantly in patients with hip and knee pain. Interestingly, chronic back pain profited from the extended use and showed significant increases in strength and mobility scores after a median of 88.5 days. Nonetheless, the presented data warrant a careful interpretation and further analyses are required to substantiate the early indicators of a therapeutic benefit of the examined digital home exercise program.

There was a significant number of drop-outs in the course of time. This was expected since the present study is no prospective study. The overall retention rate is 17% after two weeks, 10% after four weeks, 4% after eight weeks and 3% after twelve weeks. A high attrition rate is prone to bias. Thus, these preliminary data warrant a careful interpretation of its clinical implications.

The app's software does not offer the option of an active surveillance of AEs. Vivira's PMS, however, includes the option to record complaints and repairs. These provided no indication of AE relevant hazards and AEs. It was also not necessary to file a notification of hazard or AEs with the supervisory authorities. There was also no recall action. The preclinical risk analysis revealed foreseeable hazards and AEs. However, from the PMS data, authority databases data, data from preclinical risk analysis and expert opinion it can be concluded that the risk of AEs is low.

20. REFERENCES

See footnotes.

21. APPENDIX

List and description of investigators

Dr. med. Leo Benning, MPH Health Care Supply Research and Data Mining Working Group, University Emergency Department (UNZ)	University Medical Center Freiburg im Breisgau	Statistics and statistical report
Professor K. Weise, M.D., Dr. med. Medical specialist in Orthopedic Surgery Professor of Orthopedic Surgery & Traumatology Professor H.P. Zenner, M.D., Dr. med. Medical specialist in ORL, Head&Neck Surgery Professor of Head&Neck Surgery	Orthopedic Evaluation Center Tübingen Clinical Research Organization H.P. Zenner Clinical GmbH&CoKG	Study report

Curriculum Vitae Dr. med. Leo Benning, MPH
Academic Education: Medical Studies, Albert-Ludwigs-Universität of Freiburg (MD), Medical Doctorate (Dr. med.), Master of Public Health, Harvard School of Public Health (MPH)
Professional Career: Residency Training Internal Medicine, University Medical Center Freiburg (2020-2022), Head of Clinical, Vivira Health Lab GmbH (2021-2022)
Clinical Study Statistics: Quasi-experimental interventional study (G-BA Innovationsfond, 2021), Clinical Data Mining and high-dimensional data analysis for Health Care Supply Research in urgent and emergency care, observational studies on statutory health insurance (SHI) routine data.
Publications: 5 Medline-indexed publications.

Curriculum Vitae Prof. Dr. med. Kuno Weise

Clinical Education. M.D., Dr. med., University of Tübingen, Germany.
Clinical Career. Residency General Surgery followed by a residency in traumatology. 1977 Board Certified General Surgeon, 1982 Board Certified Orthopedic Surgeon, 1982 Consultant (Oberarzt) Orthopedic Surgery BG Unfallklinik Tübingen. 1990 Vice Chairman BG Unfallklinik Tübingen, 1993 Chair Orthopedic Surgery and Traumatology, University of Leipzig, 1996 Chair Orthopedic Surgery and Traumatology, University of Tübingen and Chairman BG Unfallklinik Tübingen, 2010 -today Director Institute of Orthopedic Surgery and Traumatology Evaluation, Tübingen, Germany.
Honorary Offices. 2007 President German Association of Orthopedic Surgery & Traumatology.
Evaluation. Examiner Clinical Trials biocontact prostheses, Director and Examiner Clinical Trials chondrocyte transfer, 2010 – today Director Institute of Orthopedic Surgery and Traumatology Evaluation, Tübingen, Germany.
Publications. More than 100 Medline listed publications

Curriculum Vitae Professor Hans P. Zenner

Clinical Education. 1972 M.D. University of Mainz, Germany; Studies in Paris, France. 1974 Dr. med. in oncological research.
Clinical Carreer. 1973 Medical Assistant at University of Heidelberg, Germany. 1974-1976 Post-Doc at the Dept. of Biochemistry, University of Würzburg, Germany. 1976 Residency Dept. of ORL, Head&Neck Surgery, University of Würzburg; 1981 Habilitation in Oncology, Oberarzt (Consultant) and Docent University of Würzburg. 1985 Visiting Scientist at Michigan University Ann Arbor, USA, and at Washington University St. Louis, USA. 1986 Professor of ORL, Head&Neck Surgery, University of Würzburg; 1988-2016 Distinguished Professor of ORL, HeadNeck Surgery, and Chairman Dept. of ORL, Head&Neck Surgery, University of Tübingen, Germany.
Clinical Studies. 1986-2016 Director Clinical Trials according to MPG and AMG. 2017-2018 Chairman AMG/MPG Ethics Commission Tübingen, before that member.

Honorary Offices. 2009 and 2010 President German Association of Physicians and Scientists (GDNÄ). Since 2010 Board Member, German Natl. Acad. Sci. Leopoldina.
Awards. Various awards including Leibniz Prize, Humboldt Medalist. Several International Memberships of Honor and several Doctorates honoris causa.
Clinical Research Focus. Group leader Development and Clinical Trials Medical Devices including the Tübingen Titanium Prostheses (TTP), totally implantable hearing aids, cochlear implant technology, and audiological devices.
Publications. More than 300 Medline listed publications, h index 61, 8 books, around 70 patent publications.

Publications based on the study

No paper has been published.

Patient Data Listings

Discontinued patients
See Chapter 13

Protocol deviations
None

Patients excluded from the efficacy analysis
None

Demographic data
See Chapter 13

Compliance Data
See Chapter 13

Individual Efficacy Response data
Not available

Adverse event listings
See Chapter 18

Case Report Forms

Integrated in the app`s software, no printed version available.

SIGNATURES

Name	Function		The authors agree with the contents of the report
Prof. Dr. Kuno Weise Professor of Orthopedic Surgery & Traumatology	Author		
Prof. Dr. med. H.P. Zenner, M.D., Professor of Head&Neck Surgery	Author