Effectiveness and efficiency of different weight machine-based strength training programmes for patients with hip or knee osteoarthritis: a protocol for a quasi-experimental controlled study in the context of health services research

Inga Krauss,[1] Gerhard Müller,[2] Benjamin Steinhilber,[1,3] Georg Haupt,[1] Pia Janssen,[1] Peter Martus[4]

CrossMark

For numbered affiliations see end of article.

Correspondence to
Dr Inga Krauss; inga.krauss@med.uni-tuebingen.de

ABSTRACT

Introduction Osteoarthritis is a chronic musculoskeletal disease with a major impact on the individual and the healthcare system. As there is no cure, therapy aims for symptom release and reduction of disease progression. Physical exercises have been defined as a core treatment for osteoarthritis. However, research questions related to dose response, sustainability of effects, economic efficiency and safety are still open and will be evaluated in this trial, investigating a progressive weight machine-based strength training.

Methods and analysis This is a quasi-experimental controlled trial in the context of health services research. The intervention group (n=300) is recruited from participants of an offer for insurants of a health insurance company suffering from hip or knee osteoarthritis. Potential participants of the control group are selected and written to from the insurance database according to predefined matching criteria. The final statistical twins from the control responders will be determined via propensity score matching (n=300). The training intervention comprises 24 supervised mandatory sessions (2/week) and another 12 facultative sessions (1/week). Exercises include resistance training for the lower extremity and core muscles by use of weight machines and small training devices. The training offer is available at two sites. They differ with respect to the weight machines in use resulting in different dosage parameters. Primary outcomes are self-reported pain and function immediately after the 12-week intervention period. Health-related quality of life, self-efficacy, cost utility and safety will be evaluated as secondary outcomes. Secondary analysis will be undertaken with two strata related to study site. Participants will be followed up 6, 12 and 24 months after baseline.

Trial registration number German Clinical Trial Register DRKS00009257. Pre-results.

Summary

► The dosage parameters of the investigated strength training programme provide therapeutic validity for study results.
► The design of this study by use of insurance data and propensity score matching allows an adequate control group selection in a real-life situation.
► The selection of secondary outcome measures (short- and mid-term effects on economic efficiency, health-related quality of life, self-efficacy, adherence and safety) will contribute to a relevant knowledge gain in the field of exercise therapy in osteoarthritis sustainability along a 2-year period.
► This study is conducted in a real-life situation and therefore closes the gap between efficacy and effectiveness.
► This study is not designed to capture long-term follow-up data.

INTRODUCTION

Osteoarthritis (OA) is one of the eight most common conditions worldwide and the prevalent musculoskeletal disease in the elderly.[1] According to the German health reporting system of 2010, approximately 27% of the women and 18% of the men reported a previous OA diagnosis.[2] The knee and hip joints are the most commonly affected weight-bearing joints. OA goes along with symptoms such as joint stiffness, crepitus, occasional effusion, variable degrees of inflammation and functional impairment and worsens general health status and health-related quality of life.[2–4] OA therefore has a relevant impact on the

BMJ   1

individual and the healthcare system, inducing considerable direct and indirect healthcare costs.[2] Cumulative cost of absence from work, medical costs and community and social services in Europe is currently estimated at 0.5% of gross national product.[5] [6] The economic burden will further increase according to the demographic change in industrialised countries.[3] [7] In consequence, development and evaluation of effective and cost-efficient treatment strategies are of particular importance to counteract the aforementioned increase of personal as well as economic burden of OA.

Joint replacement may be the final therapeutic option in hip or knee OA. In earlier stages, conservative therapeutic interventions are important to reduce pain and increase function and health-related quality of life.[6] [8] The use of pharmacological therapies in chronic diseases is limited because of known and significant side effects.[5] Non-pharmaceutical interventions such as physical exercise programme are therefore of particular importance. Treatment effects are similar to simple analgesics and oral non-steroidal anti-inflammatory drugs. However, reports on known side effects such as pain increase or falls are rare; thus, exercise appears to be a safe intervention.[6] [9–11]

Different land-based intervention types have shown to be effective in terms of pain reduction and improvement of physical functioning in the short term in subjects with hip or knee OA.[10] [12–14] In consequence, international guidelines and expert panels recommend exercises for increasing strength and aerobic capacity as well as flexibility and neuromuscular training as first-line strategies for the treatment of hip and knee OA.[15–18]

Despite given evidence and consistent recommendations, implications for research highlight the need for further studies to determine dose-response relationship of physical exercises on pain, physical functioning and other outcome measures,[12 13 17] to assess intervention effects on self-efficacy, quality of life, need for joint replacement and use of analgesic medications[19 20] and to assess efficiency of the intervention from an economic perspective.[17] These research topics will be addressed in a quasi-experimental controlled study in the context of health services research. The outlined study evaluates a supervised weight machine-based hip and knee strength training for patients with hip and/or knee OA in comparison to general care. The intervention is delivered at two study sites with differing exercise modes.

The primary aim of this superiority trial is to determine whether a 12-week progressive training programme decreases symptom-related pain and increases physical function in adult subjects in comparison to a matched OA control group in the short term. Effectiveness is quantified by the subscales pain and physical function of the Western Ontario McMaster Universities Osteoarthritis Index (WOMAC Index 3.1 NRS).

Secondary outcomes will be differentiated into a short-term perspective (3 months), a mid-term perspective (6 months) and a long-term perspective (follow-up data 12 and 24 months after the initiation of the intervention). The aims are to determine whether a 12-week progressive supervised weight machine-based strength training programme with a complementary 12-week facultative intervention period...

   I. improves muscle strength (no control comparator),
  II. improves health-related quality of life and/or global rating of health,
 III. decreases pain,
 IV. improves physical function,
  V. improves self-efficacy and physical activity levels,
 VI. delays time to surgery
VII. is efficient with regard to personal and/or social costs.

Aside from (I), all other secondary outcomes are compared with the control group.

In the context of a secondary analysis, outcomes will be differentiated into different intervention modes (mode of muscle contraction and number of repetitions for each muscle group) according to the instrument settings of the used strength machines at the two study sites being involved in the study. Exercise adherence, patient satisfaction and safety are further outcomes. They will be assessed throughout the intervention period for the intervention group only.

Methods and analysis
Design
This is a quasi-experimental two-centre controlled trial of a 12+12 week strength training programme with an 18-month follow-up in the context of health services research. All participants are insurance holders of the large statutory health insurance company AOK Baden-Wuerttemberg (AOK-BW) with approximately four million members in the federal state in which the study is conducted. Participants of the intervention group (IG) will be compared with a matched-pair control group (CO) to investigate the effects of an exercise intervention on clinical and economic outcomes. Matched pairs between IG and CO are determined by means of the propensity score method. Participants of IG are further allocated into a machine-based concentric/concentric strengthening programme (IG-SG) versus a machine-based concentric/eccentric strengthening programme (IG-Z) according to their place of residence and therefore study site (centre Z or SG). Measurements will be taken at baseline (t0), 3 months (t3), 6 months (t6), 12 months (t12) and 24 months (t24). Economic data will further be evaluated retrospectively for 2 years prior to baseline.
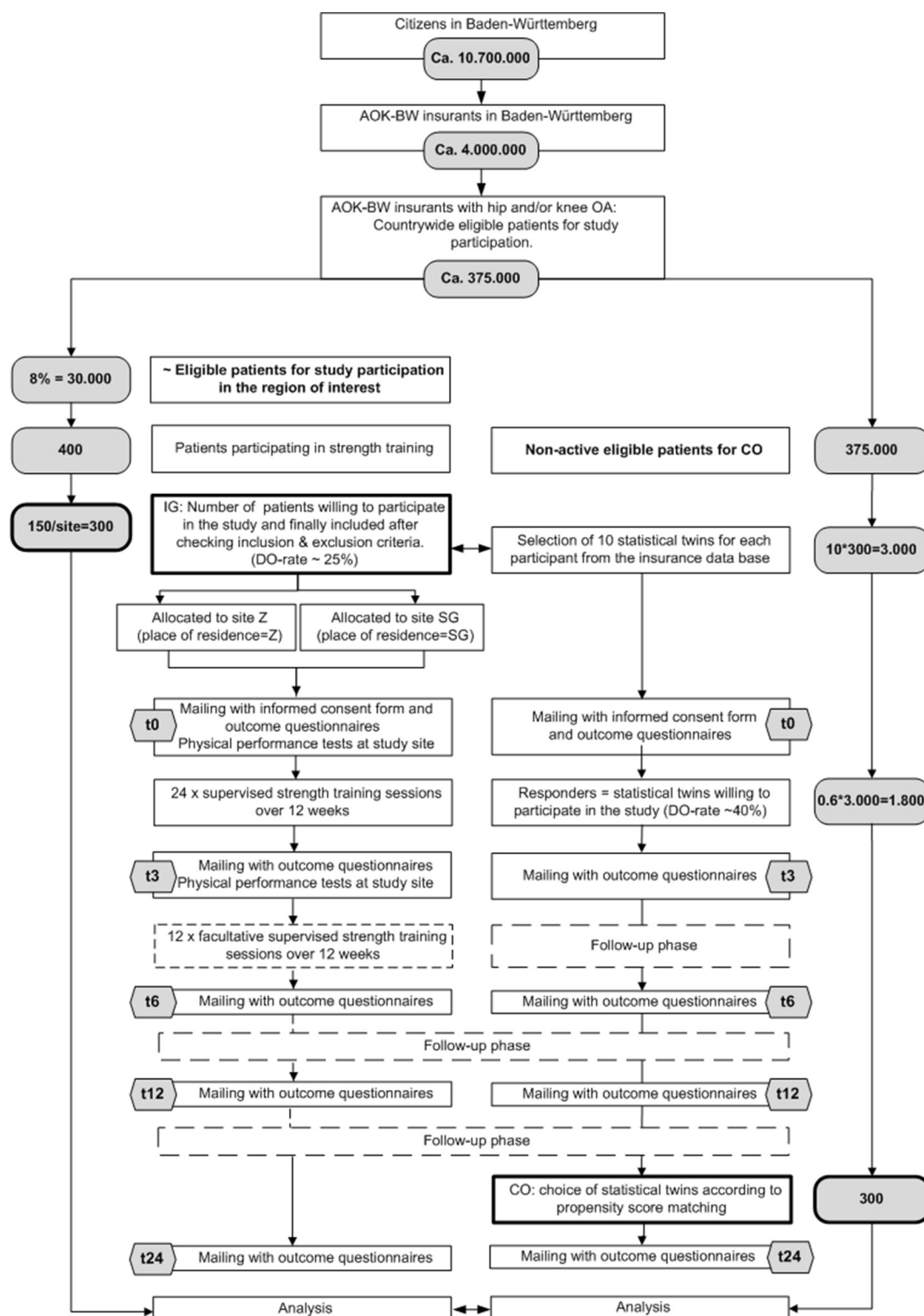
Figure 1 Study flow diagram.

Recruiting and participants

Recruitment of active study sample (IG)

The active study sample consists of eligible participants of a pilot project by the AOK-BW. Prerequisites for participation are (1) membership in the insurance company and (2) hip and/or knee OA (figure 1). Further specific inclusion and exclusion criteria for study participation are listed in box 1. The exercise programme is offered at two locations of the country. The recruitment period is 1 year. In this time frame, the intervention will be offered to approximately 200 participants at each study site (Z and SG). Local insurants with hip and/or knee OA can participate provided that their physicians prescribe the healthcare offer.

Box 1 Inclusion and exclusion criteria

INCLUSION CRITERIA

► Self-reported lifetime prevalence of hip and/or knee OA diagnosed by a medical practitioner.
► Insurance holder of the insurance company offering the exercise programme since 2 or more years.
► Physical and mental ability to participate in the interventional programme and to answer self-administered questionnaires.

EXCLUSION CRITERIA

► Significant established osteoporosis requiring treatment, previous spontaneous or low-impact fracture.
► Comorbidities leading to major impairments in everyday life and representing contraindications for physical activities.
► Hypertension stages 2–3: re-read systolic blood pressure >160 prior maximum strength testing, re-read diastolic blood pressure >100 prior maximum strength testing (intervention group (IG) only; blood pressure reading at study site prior to testing).
► Artificial joint replacement at the knee and/or hip joint within the last 6 months.
► Artificial joint replacement at the knee and/or hip joint with unstable anchoring.
► Artificial joint replacement at the knee and/or hip joint with radiological signs of implant loosening.
► Current pain at rest or with activity due to artificial joint replacement at the knee and/or hip joint.
► Luxation as an adverse event of artificial hip replacement.
► Acute joint inflammation at the knee and/or hip joint.
► Musculoskeletal surgery at the lower extremity within the last 3 months.
► Regular use of gait aids.
► Self-reported acute illness.
► Insufficient German language ability for self-administered questionnaires (IG).
► Current employment in the healthcare insurance.
► <15 points on the WOMAC Index subscale pain and <15 points on the WOMAC Index subscale physical function (scale range 0–100 with low values indicates less pain and improved function). This exclusion criterion is valid for the primary data analysis only.

Prior to this, practitioners in the region of the two study sites are introduced into the new offer. They are requested to check comorbidities as well as OA diagnosis before transcribing patients to the healthcare offer. Interested patients then register for the exercise programme at one study site and get informed about its accompanying scientific evaluation. Further details including the study information sheet for participants and the informed consent form will be delivered by mail. This mail will be sent to each participant prior to the first training and test session. The mail further includes the self-administered questionnaires. It will be

mentioned that study participation is not a presupposition for exercise participation. Eligibility of the patients to enter the exercise programme and to participate in the study will finally be determined by checking the inclusion and exclusion criteria for the exercise programme during the patient's first visit (box 1) and retrospectively by information recorded via the self-administered questionnaire at t0 (SAQ, table 1). Patients refraining from the study receive the same exercise programme and isometric peak torque test for feedback but do not enter the statistical evaluation. The benefit for study participants is the reception of preferential appointments for the start of the intervention.

Feasibility of recruitment for IG
We assume a prestudy drop-out rate of 25% due to lack of interest in the programme or in study participation or exclusion according to study criteria. Thus, 150 participants at each study site will be allocated to the trial.

Recruitment of passive study sample (control group CO)
The matched-pair control group will be recruited from the country-wide insurance database. From this, they will be selected using the propensity score method. The propensity score is the probability of treatment assignment conditional on observed baseline characteristics.[21] The detailed procedure will be outlined in the online supplementary file 1.

Feasibility of recruitment for CO
The selection of statistical twins is designed according to a successful procedure of a previous study of the insurance company. In this study, on the efficacy of an exercise programme for patients with back pain, a responder rate of the statistical twins of 60% could be achieved.[22]

Blinding
Blinding of subjects to treatment is not possible as treatment exposure is evident. Blinding of assessors of questionnaires is not applicable, as outcome measures are self-administered by the participants. Blinding of assessors for strength measures is not applicable as strength measures are quantified in the active intervention group only (see outcome measures, table 1). Blinding for the analysis of primary outcomes is warranted as random identification is blinded for the responsible statistician from t0 to t3. Blinding for further outcomes is not warranted as interim reports and results of primary analysis will unblind results.

Interventions

Active intervention: weight machine-based strengthening programme (IG-SG, IG-Z)

General aspects: All IG participants are requested to refrain from seeking other forms of treatment during the 12-week intervention period from t0 to t3. Subsequently, subjects are offered to participate in the preservation exercise programme over another 12 weeks (one session/week) from t3 to t6. After this supervised intervention period (t3 or t6), participants are further advised to continue physical exercises autonomously. In this regard, participants may attend further offers of the health insurance company.

Strength training is structured into motor learning (weeks 1–2), strength endurance training (weeks 3–5), maximum strength training (weeks 6–12) and a facultative 12-week strength preservation phase (online supplementary file 2). It comprises exercises for hip, knee and core muscles. Training starts with a 5 min ergometer warm-up, followed by 30 s stretches of the subsequently loaded muscle groups. The strength training consists of two sets for each muscle group performed consecutively with the first set using weight machines for strength training (online supplementary file 3), followed by the second set applying functional exercises in supine, sitting and standing postures. Elastic rubber bands, exercise balls, weight cuffs and subjects' own body weight are used as training devices (online supplementary file 4). While the functional exercises are identical at the two study sites, different kinds of strength training machines are used (study site SG: FREI AG, Kirchzarten, Germany; study site Z: Proxomed Medizintechnik, Alzenau, Germany). SG uses strength training machines where both agonistic and antagonistic muscles are trained in the same repetition in a concentric-concentric (con-con) mode. Z uses strength training weight machines where only the agonistic muscle group is trained separately in a concentric-eccentric (con-ecc) mode (online supplementary file 3). The exercises performed are standardised (trainers receive a standard operating procedure) and matched between the two study sites. Matching includes overall configuration (height adjustment, adjustment to range of motion), movement velocity (2–3 s for each movement direction), number of repetitions and perceived exertion after the training set. However, the different exercise modes (concentric/concentric vs concentric/eccentric) at the two study sites lead to different overall workloads per muscle group. Participants at study site Z perform about twice the workload (load×range of motion of a repetition) for a muscle group. This also leads to more time spent for each exercise session at study site Z in comparison to study site SG. Thus, the complete training duration varies between 40 and 60 min according to training site. However, pretesting indicated that, although training duration and workload differ, the perceived exertion ratings were similar for each strength training machine exercise when performing with the same percentage of a patient's isometric maximum voluntary contraction (iMVC) value.

For hip and knee exercises using weight machines, training intensity is determined according to the test results of the iMVC testing that is applied prior to the first exercise session of the programme (see also: outcome measures, physical performance tests). No iMVC strength testing will be performed for back extension and flexion to avoid adverse events related to spontaneous vertebrae fractures in subjects at higher risk for osteoporosis. Exercise loads for core muscles and for subjects who cannot perform iMVC testing for the lower extremity because of pain or subjective discomfort are guided via subjective exhaustion after the exercises (online supplementary files 2 and 5). This procedure is also used to guide progression of training loads within the intervention period. Training loads of functional exercises can be adapted by use of elastic rubber bands with different resiliencies, weight cuffs, subjects own body weight and varying lever arms of the moving extremity.[23]

Control intervention: general care (CO)

Participants of the control group receive general care and may attend other offers of the health insurance company. Interventions are not restricted. However, physical activity along the study period is monitored via SAQs.

Collection points

Participants will be assessed at baseline (t0) within (t3) and after the 24-week intervention period (t6). Two follow-ups are conducted by mail 6 (t12) and 18 (t24) months after the end of the intervention period (figure 1). Economic evaluation will further compare data 6 (mt6), 12 (mt12), 18 (mt18) and 24 (mt24) months retrospectively prior to t0.

Outcome measures (table 1)

Self-administered questionnaires

Unless otherwise stated, measures are conducted at baseline and all follow-ups.

Western Ontario McMaster Universities Osteoarthritis Index

The Western Ontario McMaster Universities Osteoarthritis Index (WOMAC NRS 3.1 German Index) is a disease-specific instrument used to evaluate self-reported pain, stiffness and functional impairment. It is a valid, reliable and responsive score, easy to complete, simple to score and available in multiple language forms and scaling formats.[24]

OMERACT-OARSI

The OMERACT-OARSI Set of Responder Criteria according to the Osteoarthritis Research Society International (OARSI) and the Outcome Measures in

Table 1 Outcome measures

Characteristics and confounders	Description and instrument	Data source	Sample	Collection points
Patient's characteristics	Date of birth, gender, ethnicity, BMI (height, weight), site (s) of OA diagnosis, date of first OA diagnosis, labour situation (working, retired, unemployed, in rehabilitation status)	SAQ, IDB	IG, CO	t0, t3, t6, t12, t24
Primary outcome measure	**Clinical outcomes**			
Pain	WOMAC Index 3.1 German (11-box NRS): subscale pain	SAQ	IG, CO	t0, t3
Function	WOMAC Index 3.1 German (11-box NRS): subscale physical functioning	SAQ	IG, CO	t0, t3
Secondary outcome measures	**Clinical outcomes**		IG, CO	
Stiffness, disease specific impairment	WOMAC Index 3.1 German (11-box NRS): subscale stiffness, overall score	SAQ	IG, CO	t0, t3, t6, t12, t24
Pain, physical function	WOMAC Index 3.1 German (11-box NRS): subscale pain, subscale physical function	SAQ	IG, CO	t6, t12, t24
Health-related quality of life	VR-12 including VR-6D utility index (4-week time slot)	SAQ	IG, CO	t0, t3, t6, t12, t24
General self-efficacy	General Self-Efficacy Scale	SAQ	IG, CO	t0, t3, t6, t12, t24
Response to exercise	OMERACT-OARSI set of responder criteria: composite score with minimum absolute and relative change levels for pain or pain and function	SAQ	IG, CO	t0, t3, t6, t12, t24
Physical activity status	Habitual physical and sports activity status	SAQ	IG, CO	t0, t3, t6, t12, t24
Time to surgery	Endpoint 'elective joint replacement'	IDB	IG, CO	t0, t3, t6, t12, t24
	Patient satisfaction			
Patient satisfaction	Modified version of the ZUF-8 Questionnaire to assess patient satisfaction	SAQ	IG	t3, t6
	Functional performance measures			
Knee extension/flexion Hip abduction/ adduction Trunk flexion/extension Closed kinetic chain leg extension	Isometric peak torque (1 RPM) using weight machines.	PPT	IG	t0, t3, t6
	Economic data			
Unspecific and specific (OA related) healthcare costs	Outpatient costs, hospital costs, costs related to drugs, adjuvants and physical modalities, rehabilitation costs, sick pay	IDB	IG, CO	tm24, tm18, tm12, tm6 t0, t3, t6, t12, t24
Unspecific/specific (OA related) periods of disability	Days of disability (overall and related to OA)	IDB	IG, CO	tm24, tm18, tm12, tm6 t0, t3, t6, t12, t24
Intervention-related costs	Costs for human and physical resources/session	IDB	IG	t3, t6, t12, t24
	Adherence to exercise			
Training adherence	Summarised number of attended training sessions according to training log	SAQ	IG	t3
	Safety evaluation			

Continued

Table 1 Continued

Characteristics and confounders	Description and instrument	Data source	Sample	Collection points
Adverse events and side effects	Summarised number and details of adverse events and side effects according to training log	SAQ	IG	t3

Data source: self-administered questionnaire via mail (SAQ), physical performance test at study site (PPT) and insurance database (IDB).
Sample: participants (IG), matched pairs control (CO).
BMI, body mass index; OA, osteoarthritis; OARSI, Osteoarthritis Research Society International; OMERACT, Outcome Measures in Rheumatoid Arthritis Clinical Trials; VR-12, Veterans RAND 12 Item Health Survey; WOMAC Index, Western Ontario McMaster Universities Osteoarthritis Index.

Rheumatoid Arthritis Clinical Trials (OMERACT) is a composite score with minimum absolute and relative change levels for pain or pain and function. Input variables for pain and function will be derived from the WOMAC subscales pain and functional impairment.[25] According to this dichotomous score, participants of the intervention are categorised into responders to exercise or non-responders to exercise.

Veterans RAND 12-Item Health Survey
The Veterans RAND 12-Item Health Survey (VR-12) was derived from the Veterans RAND 36 Item Health Survey (VR-36) [26] . It comprises eight different scales, four of them related to physical health (physical functioning, role-physical, bodily pain, general health) and four of them related to mental health (vitality, social functioning, role-emotional, mental health). Higher scores indicate better health. The VR-6D is a utility measure derived from the VR-12. Utility values capture personal preferences for health states on a scale from 0 to 1, where 0 denotes death and 1 optimum health, with negative values for states worse than death. A German version of the questionnaire with the 4-week time slot will be used for analysis.[27]

General Self-Efficacy Scale
General self-efficacy will be measured with the German version of the General Self-Efficacy Scale with a 4-point Likert scale ('not right' vs 'definitely right'). Ten items are designed to tap this construct. Each item refers to successful coping and implies an internal-stable attribution of success. The scale ranges from 10 to 40 points with 40 points indicating maximum self-efficacy.[28]

Habitual physical activity status
Habitual physical activity status will be quantified with regard to sports activities as well as physical activities in everyday life in the previous 4 weeks. Questions are similar to those being used in the German Health Update 2012. Sports activities are differentiated into none, less than 2 hours/week, 2–4 hours/week or more; physical activities are defined as a state of sweating or being out of breath aside of the above-mentioned sports activities. Their frequency is requested in days per week, and their duration should be specified into less than 10 min/day, 10–30 min/day, 30–60 min/day or more.[29 30]

Patient satisfaction
Patient satisfaction (t3, t6) will be quantified by participants of the IG with a 4-four point Likert scale using a modified and shortened version of the ZUF-8.[31–33] Modifications were made in order to adapt the 8-item instrument to the outpatient setting.

Patient characteristics
The following physical charateristics were determined: body mass index (height, weight), site(s) of OA diagnosis and artificial joint replacement (hip or knee and time of event).

Safety evaluation (IG only, t3 and t6)
The questionnaires at t3 and t6 include a section where subjects can report on any symptom related to the exercise programme that occured within the preceding intervention phase. Further questions ask for frequency, duration and intensity as well as possible reasons for the onset of symptoms. Aside from this study-related safety report, participants are urged to contact their personal trainers or physicians in case of adverse events and side effects.

Adherence to exercise (IG only, t3 and t6)
Adherence is quantified by the number of attended training sessions in relation to all offered training sessions. Within each session, the number and type of exercises, the number of repetitions and the subjective exhaustion after each strength machine-based exercise are documented via the software of the strength training machine. The questionnaires at t3 and t6 further ask for reasons of non-compliance in case of non-attendance.

Sociodemographic data (IDB)
The following sociodemographic data will be read out of the database: date of birth, gender and labour situation (working, retired, unemployed, in rehabilitation status), complexity of work (from 1 low to 4 high), level of education (1 no graduation to 4 high school), highest level of educational attainment (from

1 no qualification to 6 doctoral degree) and contractual form (permanent/fixed term contract, full time/part time).

Economic data (IDB)

Patient-specific economic data comprise (1) unspecific and (2) specific OA-related healthcare costs for the diagnosis of hip and/or knee OA, (3) specific/unspecific disability days, (4) intervention-related costs for the exercise programme.

1. Unspecific healthcare costs (overall costs): sick pay, hospital costs, outpatient costs, costs related to periods of disability and costs related to drugs, physical modalities and adjuvants.
2. Specific diagnosis (hip/knee OA) related healthcare costs: sick pay, hospital costs, outpatient costs, costs related to periods of disability and costs related to disease-related drugs, physical modalities and adjuvants such as walkers, cranks or orthotics.
3. Specific (hip/knee OA related)/unspecific disability days.
4. Intervention-related costs include human and physical resources that are required for the training programme. Intervention-related costs will be added to the above-mentioned diagnosis related costs for all participants of the intervention groups according to the number of their scheduled training session (24 mandatory units+another 12 optional units).

Physical performance tests (study sites, IG only, t0, t3, t6 (facultative))

Maximum strength will be quantified bilaterally at the same time for hip abduction, hip adduction, knee extension and knee flexion and closed kinetic chain leg extension using iMVC (online supplementary file 3). Prior testing, a blood pressure reading is conducted. In case of exceeding the thresholds (box 1), subjects have to be excluded from strength testing and further study participation. Otherwise, subjects accomplish a 5 min warm-up on a bicycle ergometer (50–100 watts) followed by some stretching exercises. Each trial is initialised by a sub-maximum test trial to allow participants to get used to the movement and measurement procedure. Participants are then asked to contract their muscles by gradually increasing intensity within 5 s until the maximum is reached. After a short rest period (45 s), the contraction is repeated two more times. The trial will be repeated if the execution is of low quality in terms of jerky muscle contraction. Finally, the highest peak torque value is used for data analysis.

Statistics
Sample size

The empirical basis of the sample size estimation was retrieved from the paper of Krauss et al.[34] In this paper, intraindividual differences of WOMAC subscale pain and WOMAC subscale physical function identically showed a standardised difference of about 0.5 between the intervention and the control group (pain, intervention: −8.5±13.9, control: −1.3±15.3, physical function, intervention: −8.4±13.4, control: −2.1±12.9). We do not assume a relevant efficacy-effectiveness gap in the study described in this protocol because the intervention under study is a pilot offer for insurance holders at two sites of the country only, with highly experienced and trained providers in a resource-intensive 'ideal setting'.[35] We thus also assume standardised differences of 0.5 for this study according to the above-mentioned effect sizes of the randomised controlled trial by Krauss et al.

Due to the fact that we want to test the two endpoints (pain and physical function) simultaneously, we choose a level of significance of 0.025 (two sided, Bonferroni correction) and a power of 0.90 for each test. This leads to a sample size of 101 evaluable subjects per group in the parallel group design (nQuery release 7.0). We assume a prestudy drop-out rate of 25% due to non-eligibility or non-willingness of subscribers to the healthcare offer to participate in the study. We further assume a ~30% drop-out rate after inclusion. The intervention will therefore be offered to 400 participants; of them, 300 subjects should be allocated to the trial. This sample size estimation takes into account the secondary separate analysis of the factor study site/training mode with two levels. The control group includes one statistical twin for each participant. This leads to a final sample size of 300 subjects in the control group (figure 1).

We refrain from analysing our data using a matched-pair design as propensity score matching does not guarantee that individual pairs will be well-matched on the full set of covariates, only that groups of individuals with similar propensity scores will have similar covariate distributions.[36] Despite this fact, propensity score matching including, among other variables, baseline WOMAC scores as matching factors may reduce variance of outcome measures and therefore probably increase the power of the study further if the propensity score is included as covariate in the statistical evaluation model. However, it is difficult to make a guess about this effect, and thus, we conservatively calculated the sample size obtained for the parallel group design.

Statistical analysis of clinical endpoints

Controls and cases will be included into a linear model with the propensity score as covariate. Additionally, we will use models for longitudinal data using each measurement point (instead of only t0 and t3 for the primary analysis). The level of significance will be 0.025 (two sided, Bonferroni correction) for both primary endpoints. Secondary endpoints will be analysed analogously without claiming confirmatory interpretation of p-values. Descriptive analysis will include absolute and percentage frequencies for categorical variables, means, medians, SD, quartiles and ranges for quantitative variables and medians, quartiles and ranges for ordinal variables. For the main results, two-sided 95% CIs will

be given additionally to significance tests. For percentages, exact CI for proportions based on the binomial distribution will be given.

Exploratory, prognostic factors will be analysed using multiple regression models (linear regression) to identify potential responders to the training.

Correlation analyses will use Pearson product moment correlation coefficient (normally distributed variables) and Spearman correlation (non-normally distributed variables).

The primary analysis will be done on the full set of participants recruited in consideration of the exclusion criteria 'limited pain and impairment in physical function' (WOMAC Index subscales pain and physical function with values below 15) except for participants refusing consent and usage of data during the study and obviously wrongly diagnosed participants (intent-to-treat population). The following secondary analyses are planned: (1) analysis of primary outcome in all subjects with adherence to the study protocol as to be defined by a blind data review (per protocol population); (2) analysis in the full set of eligible and included participants irrespective of their limitations in pain and physical function at baseline (t0); this analysis is done in order to improve external validity of the study even so we do not expect the same effect size in these participants; (3) analysis including study site as an additional factor to compare the effectiveness of both training approaches; and (4) separate analysis for both study sites and inspection of interactions between type of treatment and other prognostic factors. Secondary analyses will only claim a local confirmatory interpretation of p-values if the primary analyses are able to detect group differences in a statistical manner. Unplanned analyses will be clearly defined as 'exploratory'.

Missing values will be imputed using multiple imputation approaches; complete case and last observation forward analysis will be performed as a sensitivity analysis. No interim analysis, except for administrative purposes, will be performed. All statistical analysis will be done using the software SPSS and R in the newest release.

Statistical analysis of economic endpoints

Cost-efficiency of the intervention will be quantified on the expanded perspective of the payer. If a dominant strategy does not exist (ie, lower costs and higher health-related effects in the intervention group), the costs will be related to the effects. The costs (including the supplementary costs of the pilot offer for participants) will be related to the differences between the groups (double-difference method, 2 years preintervention and 2 years postintervention) in quality-adjusted life-years (QALYs, equation 1) and health-related effects (WOMAC Index: equation 2). The costs of the intervention will further be related to the differences in costs (double-difference method of the healthcare costs (unspecific healthcare costs (overall costs), specific healthcare costs and the

costs for days of disability (human capital approach): equation 3). The time to surgery will be estimated using the Kaplan-Meier method. For the economic evaluation, all participants have to be members of the insurance for the period from 24 months before baseline (tm24) up to 24 month after baseline (t24).

Equation 1: Cost-utility analysis=incremental cost utility ratio (ICUR)

$$ICUR = \frac{\Delta cost}{\Delta QALY_S}$$

Equation 2: Cost-effectiveness analysis=incremental cost-effectiveness ratio (ICER)

$$ICER = \frac{\Delta cost}{\Delta Effect}$$

Equation 3: Cost-benefit analysis=incremental cost-benefit ratio (ICBR)

$$ICBR = \frac{\Delta cost}{\Delta Benefit}$$

Ethics and dissemination

All participants of the exercise programme receive a postal mailing prior to the first training session. This mailing includes information on the aims and the content of the study, as well as on data privacy. The mailing further includes the informed consent form and the outcome questionnaire. From the insurance database, selected subjects for the control group receive a similar mailing. Interested persons are informed that they confirm their informed consent by returning the form and the pseudomysed outcome questionnaires by post. Participants are undeceived about the voluntariness of study participation at all times. Participants of the exercise programme are further informed that study participation is not a requirement for participation in this healthcare offer.

Ethical approval has been obtained from the Ethics Committee of the University of Tuebingen (vote number 421/2015BO1). Besides data privacy was approved by the data protection officer of the insurance company.

A manuscript with the results of the primary data analysis will be prepared for publication at first. Further manuscripts will be written on short-term, mid-term and long-term effects of secondary outcomes related to clinical and economic endpoints. Data on exercise adherence and safety aspects will be included. It is contrived to publish all manuscripts in peer-reviewed journals.

Study data are erased after the legal retention period according to privacy policy apart from data that were already analysed and gone down as study report and/or publication.

Timelines

Ethical approval was obtained in August 2015. Definition of exercise regime and training of exercise instructors has been undertaken from May 2015 to September 2015. Recruitment of participants for the study started in September 2015 (first patient in). The intervention period is scheduled for 1 year (last patient in: August 2016). All participants are expected to have completed the intervention period end of February 2017. All participants are expected to have completed the study by end of August 2018 (last patient out).

Discussion

Summary of the aim of the study

Non-pharmaceutical interventions are extremely important in the treatment of chronic diseases such as OA of the hip and knee. However, data on dose-response relationship and outcomes other than pain and function such as self-efficacy, quality of life, need for joint replacement and use of analgesic medications as well as economic measures are rare. The aim of the outlined study therefore is to address the above-mentioned research topics by evaluating a supervised weight machine-based hip and knee strength training for patients with hip and/or knee OA in the context of health services research.

Strength of the study and value of the results

Evidence is needed for exercise dosage recommendations, and therapeutic validity of exercise programme should be warranted.[36] The transparency of this study protocol with respect to patient eligibility, exercise setting as well as rationale and content of the therapeutic exercises gives consideration to this requirement and can therefore be identified as a strength of this study. Another strength is related to the choice and time point of outcome measures and its field of application: this study allows new insights into the effectiveness of an exercise intervention into patient care with respect to short-term as well as mid-term effects. The outcomes of the study are related to pain, physical function and economic measures on health-related quality of life, self-efficacy, adherence and safety. Therefore, a significant knowledge gain for the aforementioned aspects can be expected which are insufficiently investigated so far.[17 19 20]

Limitation of the study

A relevant open research question is related to the long-term effectiveness of an exercise intervention on symptoms as well as disease progression.[10 12 17] The two-year follow-up phase of this study will not allow sufficient data for statements on the effect of the intervention in the long run such as symptom release, time to surgery and disease-specific healthcare costs. This drawback of the study might be compensated by a protocol amendment for a 5-year follow-up if 24-month evaluations are positive in terms of study outcomes.

Author affiliations
^{1}Department of Sports Medicine, Medical Clinic, University Hospital Tübingen, Tübingen, Germany
^{2}Allgemeine Ortskrankenkasse AOK Baden-Wuerttemberg, Stuttgart, Germany
^{3}Institute of Occupational and Social Medicine and Health Services Research, University Hospital Tübingen, Tübingen, Germany
^{4}Institute for Clinical Epidemiology and Applied Biometry, University Hospital Tübingen, Tübingen, Germany

Handling editor Karim Khan

REFERENCES

1. World Health Organization. *The global burden of disease: 2004 update.* Geneva: WHO, 2008.
2. Rabenberg M. Arthrose. *Gesundheitsberichterstattung des Bundes.* Berlin: Robert Koch-Institute, 2013:1–36.
3. Grotle M, Hagen KB, Natvig B, *et al.* Prevalence and burden of osteoarthritis: results from a population survey in Norway. *J Rheumatol* 2008;35:677–84.
4. Symmons D, Mathers C, Pfleger B. Global burden of osteoarthritis in the year 2000. http://www.who.int/healthinfo/statistics/bod_osteoarthritis.pdf. 2006.

5. Conaghan PG, Kloppenburg M, Schett G, *et al*. Osteoarthritis research priorities: a report from a EULAR ad hoc expert committee. *Ann Rheum Dis* 2014;73:1442–5.
6. Krauss I, Mueller G, Haupt G, *et al*. Effectiveness and efficiency of an 11-week exercise intervention for patients with hip or knee osteoarthritis: a protocol for a controlled study in the context of health services research. *BMC Public Health* 2016;16:367.
7. Sharif B, Kopec J, Bansback N, *et al*. Projecting the direct cost burden of osteoarthritis in Canada using a microsimulation model. *Osteoarthritis Cartilage* 2015;23:1654–63.
8. American College of Rheumatology Subcommittee on Osteoarthritis Guidelines: Recommendations for the medical management of osteoarthritis of the hip and knee: 2000 update. *Arthritis Rheum* 2000;43:1905–15.
9. Roddy E, Zhang W, Doherty M, *et al*. Evidence-based recommendations for the role of exercise in the management of osteoarthritis of the hip or knee–the MOVE consensus. *Rheumatology* 2005;44:67–73.
10. Fransen M, McConnell S, Hernandez-Molina G. Reichenbach S: Exercise for osteoarthritis of the hip. *Cochrane Database Syst Rev* 2014;4:CD007912.
11. Bennell KL, Dobson F, Hinman RS. Exercise in osteoarthritis: moving from prescription to adherence. *Best Pract Res Clin Rheumatol* 2014;28:93–117.
12. Fransen M. McConnell S: Exercise for osteoarthritis of the knee. *Cochrane Database Syst Rev* 2009.
13. Juhl C, Christensen R, Roos EM, *et al*. Impact of exercise type and dose on pain and disability in knee osteoarthritis: a systematic review and meta-regression analysis of randomized controlled trials. *Arthritis Rheumatol* 2014;66:622–36.
14. Zhang W, Nuki G, Moskowitz RW, *et al*. OARSI recommendations for the management of hip and knee osteoarthritis: part III: changes in evidence following systematic cumulative update of research published through January 2009. *Osteoarthritis Cartilage* 2010;18:476–99.
15. American Academy of Orthopaedic Surgeons. Treatment of Osteoarthritis (OA) of the Knee. http://www.aaos.org/Research/guidelines/GuidelineOAKnee.asp
16. Conaghan PG, Dickson J, Grant RL. Guideline Development Group. Care and management of osteoarthritis in adults: summary of NICE guidance. *BMJ* 2008;336:502–3.
17. Fernandes L, Hagen KB, Bijlsma JW, *et al*. EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. *Ann Rheum Dis* 2013;72:1125–35.
18. Zhang W, Moskowitz RW, Nuki G, *et al*. OARSI recommendations for the management of hip and knee osteoarthritis, part II: OARSI evidence-based, expert consensus guidelines. *Osteoarthritis Cartilage* 2008;16:137–62.
19. Bennell KL, Hinman RS. A review of the clinical evidence for exercise in osteoarthritis of the hip and knee. *J Sci Med Sport* 2011;14:4–9.
20. Lange AK, Vanwanseele B, Fiatarone Singh MA. Strength training for treatment of osteoarthritis of the knee: a systematic review. *Arthritis Rheum* 2008;59:1488–94.
21. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011;46:399–424.
22. Müller G. *Bericht zur Evaluationsstudie AOK-Rückenkonzept*, 2015. Ref Type: Personal Communication.
23. Haupt G, Janßen P, Krauß I, *et al*. *Das Tübinger Hüftkonzept*. Essen: hellblau, 2014.
24. Bellamy N. The WOMAC Knee and Hip Osteoarthritis Indices: development, validation, globalization and influence on the development of the AUSCAN Hand Osteoarthritis Indices. *Clin Exp Rheumatol* 2005;23:S148–S153.
25. Pham T, Van Der Heijde D, Lassere M, van der HD LM, Altman RD, *et al*. Outcome variables for osteoarthritis clinical trials: the OMERACT-OARSI set of responder criteria. *J Rheumatol* 2003;30:1648–54.
26. Usman Iqbal S, Rogers W, Selim A, *et al*. The Veterans Rand 12 Item Health Survey (Vr-12): What it is and how it is used. 2015. http://hosonline.org/surveys/hos/download/Veterans_RAND_12_Item_Health_Survey_VR-12_2007.pdf.
27. Buchholz I, Buchholz M, Kohlmann T. DRV Schriften Band 111: Der Veterans RAND Health Survey (VR-36, VR-12): Eine Alternative zum Short Form (SF-36, SF-12)? *26. Rehabilitationswissenschaftliches Kolloquium* 2017:60–2 http://forschung.deutsche-rentenversicherung.de/ForschPortalWeb/ressource?key=tagungsband_26_reha_kolloqu.pdf
28. Schwarzer R, Jerusalem M. Generalized Self-Efficacy scale. Weinman J, Wright S, Johnston M, *et al*. *Measures in health psychology: A users portfolio*. England 1995:35–7.
29. Robert Koch-Institut. *Faktenblatt zu GEDA 2012: Ergebnisse der Studie £Gesundheit in Deutschland aktuell 2012"* Sportliche Aktivität. Berlin: Robert Koch-Institut, 2014.
30. Robert Koch-Institut. *Faktenblatt zu GEDA 2012: Ergebnisse der Studie £Gesundheit in Deutschland aktuell 2012"* Körperliche Aktivität. Berlin: Robert Koch-Institut, 2014.
31. Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. *Eval Program Plann* 1982;5:233–7.
32. Schmidt J, Lamprecht F, Wittmann WW. [Satisfaction with inpatient management. Development of a questionnaire and initial validity studies]. *Psychother Psychosom Med Psychol* 1989;39:248–55.
33. Krauß I, Steinhilber B, Haupt G, *et al*. Exercise therapy in hip osteoarthritis–a randomized controlled trial. *Dtsch Arztebl Int* 2014;111:592–9.
34. Singal AG, Higgins PD, Waljee AK. A primer on effectiveness and efficacy trials. *Clin Transl Gastroenterol* 2014;5:e45.
35. Stuart EA. Matching methods for causal inference: a review and a look forward. *Stat Sci* 2010;25:1–21.
36. Hoogeboom TJ, Oosting E, Vriezekolk JE, *et al*. Therapeutic validity and effectiveness of preoperative exercise on functional recovery after joint replacement: a systematic review and meta-analysis. *PLoS One* 2012;7:e38031.