Author(s): Heikkinen, Risto; Waller, Benjamin; Munukka, Matti; Multanen, Juhani; Heinonen, Ari; Karvanen, Juha

Title: Impact or No Impact for Women With Mild Knee Osteoarthritis : A Bayesian Meta-Analysis of Two Randomized Controlled Trials With Contrasting Interventions

Arthritis Care & Research
Vol. 74, No. 7, July 2022, pp 1133–1141
DOI 10.1002/acr.24553

AMERICAN COLLEGE of RHEUMATOLOGY
Empowering Rheumatology Professionals

Impact or No Impact for Women With Mild Knee Osteoarthritis: A Bayesian Meta-Analysis of Two Randomized Controlled Trials With Contrasting Interventions

Risto Heikkinen,[1] Benjamin Waller,[2] Matti Munukka,[1] Juhani Multanen,[1] Ari Heinonen,[1] and Juha Karvanen[1]

Objective. To predict the probability of a benefit from 2 contrasting exercise programs for women with a new diagnosis of mild knee osteoarthritis, and to estimate the short- and long-term effects of aquatic resistance training (ART) and high-impact aerobic land training (HLT) compared with a control.

Methods. Original data sets from 2 previously conducted randomized controlled trials were combined and used in a Bayesian meta-analysis. Group differences in multiple response variables were estimated. Variables included cardiorespiratory fitness, dynamic maximum leg muscle power, maximal isometric knee extension and flexion force, pain, other symptoms, and quality of life. The statistical model included a latent commitment variable for each female participant.

Results. ART had a 55–71% probability of benefits in the outcome variables, and as the main effect, the intervention outperformed the control in cardiorespiratory fitness, with a probability of 71% immediately after the intervention period. HLT had a 46–63% probability of benefits after intervention with the outcome variables, but differently from ART; the positive effects of physical performance fade away during the follow-up period. Overall, the differences between groups were small, and the variation in the predictions between individuals was high.

Conclusion. Both interventions had benefits, but ART has a slightly higher probability of long-term benefits on physical performance. Because of high individual variation and no clear advantage of one training method over the other, personal preferences should be considered in the selection of the exercise program to ensure highest commitment to training.

INTRODUCTION

Exercise is one of the cornerstones in the management of osteoarthritis (OA) of the hip and knee (1). Since 2002 and unchanged with new evidence, it has been known that exercise is effective for the management of pain and impaired function in hip and knee OA (2,3). While there is strong evidence for positive effects from land-based neuromuscular exercise and muscular strength, aerobic, and aquatic exercise, there is no consensus on which type of exercise is superior. Recent systematic reviews have been unable to separate the different training environments (4,5), with only a consensus that the training should focus on a specific outcome and be completed 3 times a week (6).

In randomized controlled trials (RCTs) assessing the effect of exercises, the inference is often based on frequentist statistical analysis with group means and *P* values. A *P* value answers the question, "Under the null hypothesis, what is the probability to obtain this or more extreme result." A more important question not answered by *P* values (7) is, "What is the probability that each exercise program would be beneficial for each participant individually, and which exercise program should they choose?"

We use Bayesian analysis (8) to combine the relevant information from 2 different studies and calculate the probabilities that support the decision-making on exercise program recommendation. The first aim of the study is to calculate Bayesian posterior distributions and compare different exercise programs' probabilities

Supported by the Academy of Finland (grant 311877 under the Decision Analytics Utilizing Causal Models and Multiobjective Optimization [DEMO] research area), the Finnish Cultural Foundation, and the Central Finland Regional Fund.

[1]Risto Heikkinen, MSc, Matti Munukka, PhD, Juhani Multanen, PhD, Ari Heinonen, PhD, Juha Karvanen, DSc (Tech): University of Jyväskylä, Jyväskylä, Finland; [2]Benjamin Waller, PhD: University of Reykjavik, Reykjavik, Iceland.

Dr. Karvanen has received consulting fees and/or speaking fees from Biogen Finland and Tale (less than $10,000 each). No other disclosures relevant to this article were reported.

Address correspondence to Risto Heikkinen, MSc, University of Jyväskylä, Department of Mathematics and Statistics, PO Box 35 Jyvaskyla, Keski-Suomi 40014, Finland. Email: risto.heikkinen@statisti.fi.

Submitted for publication July 7, 2020; accepted in revised form January 5, 2021.

SIGNIFICANCE & INNOVATIONS

- We quantify the probability of a benefit from aquatic resistance training and high-impact land training in multiple measures of physical performance, symptoms, and quality of life for patients with mild knee osteoarthritis (OA).
- As the group differences are small compared to the variation between the individuals, patients should choose the exercise according to their preferences to ensure highest commitment to training.
- Medical professionals can make improved personal recommendations on training for individuals with knee OA based on predictive probability calculations that show tradeoffs between different outcomes.

of being beneficial to a new patient in physical performance, symptoms, and quality of life. The second aim is to improve the understanding about uncertainties and individual variation in predictions of exercise responses.

MATERIALS AND METHODS

Study design. This study utilized data from our 2 previous registered RCT studies, AquaRehab (ISRCTN: 65346593) and LuRu (ISRCTN: 58314639). The data sets were collected from January 2012 to May 2013, and March 2008 to April 2010 for AquaRehab and LuRu, respectively. Both studies had an exercise intervention group (aquatic resistance training [ART] in AquaRehab, high-impact aerobic land training [HLT] in LuRu), and both had a nonintervention control group. The study protocols of AquaRehab (9) and LuRu (10) can be found elsewhere and were followed without changes. Included participants were women ages 50–68 years with mild knee OA, body mass index of <35, and no medical reason preventing participation in intensive exercise. Mild knee OA was classified as experiencing knee pain on most days during the last 12 months, not exceeding 5 of 10 on a visual analog scale (0 = "no pain at all," and 10 = "worst pain imaginable"), with radiographic changes in tibiofemoral joint grades I (possible osteophytes) or II (definite osteophytes, possible joint space narrowing) according to the Kellgren/Lawrence classification (11). The design of both studies followed the Consolidated Standards of Reporting Trials (CONSORT) recommendations (12). Both AquaRehab (Dnro 19U/2011) and LuRu (Dnro1E/2008) study protocols were approved by the Ethics Committee of the Central Finland Health Care District and conform to the Declaration of Helsinki. Written informed consent was obtained from all participants prior to enrollment.

Subject recruitment and randomization. The recruitment methods and eligibility criteria for AquaRehab (13) and LuRu (10) are described elsewhere. Inclusion criteria in these 2 RCTs were otherwise similar except for age (AquaRehab: age range 60–68 years; LuRu age range: 50–66 years). The subjects in both studies were randomly allocated into 1 of the 2 arms of the study. Principal investigators were blinded to group allocation. The recruitment process is presented as a flow chart in Figure 1 of Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24553.

Interventions. Participants in the AquaRehab intervention group participated in ART lasting 1 hour, 3 times a week for 4 months. Variable resistance equipment was used to progress training intensity with 3 resistance levels: barefoot; small Thera-Band resistance fins (Hygenic Corporation); and large Hydro-boots resistance boots (Hydro-Tone Fitness Systems). Training intensity was set at "as hard and fast as possible." A full description of the training program, its progression, and daily training program can be found elsewhere (13).

Participants in the LuRu intervention group participated in supervised HLT, multidirectional aerobic and step-aerobic jumping lasting 55 minutes, 3 times a week for 12 months. The loading was gradually increased after 3 months by progressively raising the height of the fences from 5–20 cm in aerobic exercises, and the height of the step benches from 10–20 cm in jumping exercises. More detailed exercise protocol is provided elsewhere (10).

The control groups in both studies maintained usual care and were asked to continue their leisure time activities. The controls were offered 2 sessions consisting of 1 hour of light stretching, relaxation, and social interaction in AquaRehab, and a social group meeting every third month in LuRu.

Outcome measures. Measurement protocols were identical in both studies. In this study, we chose to use the secondary outcomes from both studies because these are more clinically applicable than the primary outcomes that required quantitative magnetic resonance imaging and dual-energy X-ray absorptiometry. Cardiorespiratory fitness (VO_2 peak, ml/kg/minute) was calculated from the UKK 2 km walk test (14). Maximal isometric knee extension and flexion force (in newtons) of the affected knee was measured using an adjustable Good Strength dynamometer chair (Metitur) (15). Dynamic maximum leg muscle power (in watts) was examined by measuring peak instantaneous power production during the take-off phase of counter movement jump performed on a custom-made force plate (University of Jyväskylä, Finland). Self-reported pain, other symptoms, and quality of life were measured using the 3 domains of the Finnish version of the Knee Injury and Osteoarthritis Outcome Score (KOOS) (16). Scores were transformed into a score of 0–100, with a score of 0 indicating extreme knee problems, and 100 indicating no knee problems (17).

Statistical analysis. The changes in the response variables from baseline to the end of the intervention and from

Table 1. Group baseline measurements*

	AquaRehab		LuRu		Combined, control (n = 83)
	Control (n = 43)	Intervention (n = 42)	Control (n = 40)	Intervention (n = 35)	
Age, years	63.9 ± 2.2	63.7 ± 2.4	58.2 ± 4.3	57.4 ± 4.2	61.2 ± 4.4
Height, cm	163 ± 4.7	163 ± 5.3	162 ± 4.5	166 ± 5.9	163 ± 4.6
Body mass, kg	69.8 ± 10.6	68.6 ± 9.9	68.8 ± 11.3	72.6 ± 8.4	69.3 ± 10.9
BMI, kg/m^2	26.1 ± 3.2	25.7 ± 3.6	26.1 ± 4.0	26.4 ± 2.7	26.1 ± 3.6
Vo_2 peak, ml/kg/minute	24.9 ± 4.8	24.5 ± 5.6	29 ± 4.3	29.1 ± 3.9	26.9 ± 5.0
Power, W	1,663 ± 285	1,612 ± 260	1,798 ± 341	1,975 ± 382.9	1,728 ± 318
Force, N					
Extension	353 ± 78.9	333 ± 61.7	413 ± 74.6	408 ± 102	382 ± 82.2
Flexion	170 ± 43.1	165 ± 51.3	178 ± 54.5	189 ± 54.8	174 ± 48.8
KOOS score (range 0–100)					
Pain	82.2 ± 12.0	80.3 ± 10.3	86.9 ± 7.2	85.9 ± 10.3	84.5 ± 10.2
Symptoms	75 ± 14.3	74.1 ± 13.0	82.7 ± 10.1	78.3 ± 12.0	78.7 ± 13.0
QoL	70.6 ± 20.3	65.5 ± 17.4	78.5 ± 15.2	76.3 ± 15.1	74.4 ± 18.4

* Values are the mean ± SD. Only participants with at least 1 observed value in response variables have been included. BMI = body mass index; KOOS = Knee Injury and Osteoarthritis Outcome Score; N = newtons; QoL = quality of life; W = watts.

baseline to the end of the follow-up period were compared between the ART group, the HLT group, and the combined control group. In the Bayesian model, the 14-variate response variable (changes in the 7 secondary outcomes, at 2 time points, postintervention and follow-up) was explained by the exercise group effect and a personal effect modifier. The modifier is a latent variable that has not been measured but is estimated from the correlation structure of the data. This personal effect includes commitment, i.e., adherence and compliance with training, plus all other personal factors, for example, age and comorbidities that could cause systematic variation to one's exercise effects between individuals within the group. The modifier is defined so that the population average is 1, i.e., for a woman with average intervention effect, the latent coefficient has a value of 1. For example, if a woman has 5% higher than average differences in all outcome variables, her modifier has a value of 1.05. The minimum for the modifier is 0, indicating possible poor adherence and compliance with the intervention. It is assumed that other unmeasured personal factors do not change the sign of the intervention effect. A gamma distribution with mean 1 and variance 0.2 was chosen to describe our prior knowledge on this individual variation. This distribution has quantiles (Q) (Q[0.025] = 0.33 and Q[0.975] = 2.05) describing range of typical values based on our prior knowledge. The personal effect modifier was fixed to 0 for the control group, as there is no intervention effect. The error terms were modeled by the 14-variate normal distribution.

Missing values in the response variables were assumed to be missing at random and were handled as unknown parameters in the model, i.e., missing values were imputed parallel with estimation of the parameters of interest. The statistical modeling was carried out using R (18) and RStan (19). More information about the implemented Bayesian model, including mathematical

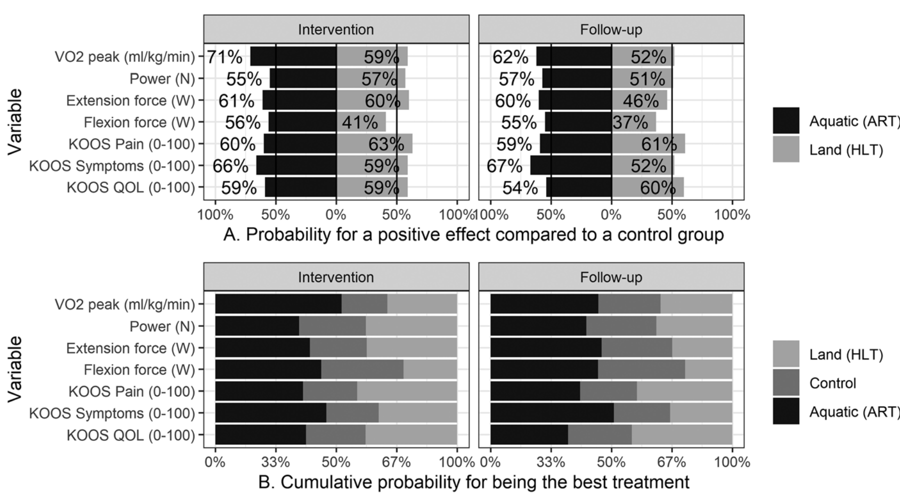


Figure 1. **A**, Probability for a positive effect compared to a control group. **B**, Cumulative probability for being the best treatment. ART = aquatic resistance training; HLT = high-impact aerobic land training; KOOS = Knee Injury and Osteoarthritis Outcome Score; QOL = quality of life.

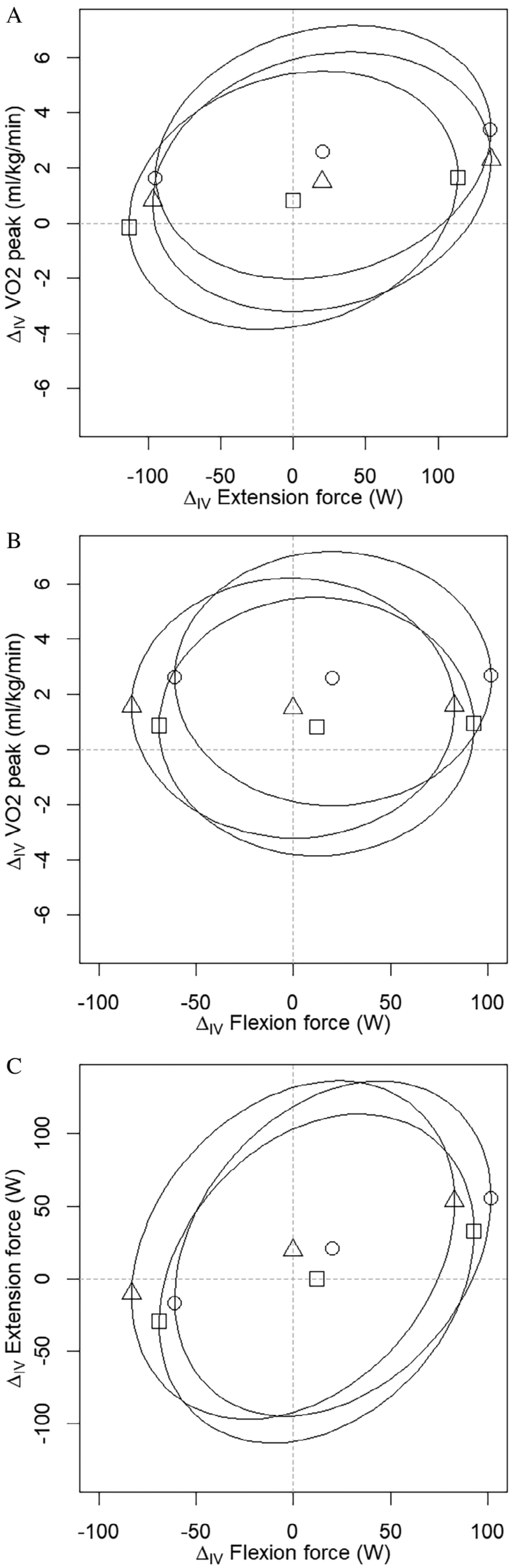


Figure 2. The mean change between baseline and postintervention and 2-dimensional 90% Bayes region for the prediction of extension force and Vo_2 peak (**A**), flexion force and Vo_2 peak (**B**), and flexion force and extension force (**C**). The values in the top right indicate a positive outcome in both variables. There is 1 mean point and Bayes region for each group. Circles represent the aquatic resistance training group, triangles represent the high-impact aerobic land training group, and squares represent the control group. W = watts.

formulas, is given in Supplementary Appendix A, available on the *Arthritis Care & Research* website at http://onlinelibrary.wiley.com/doi/10.1002/acr.24553.

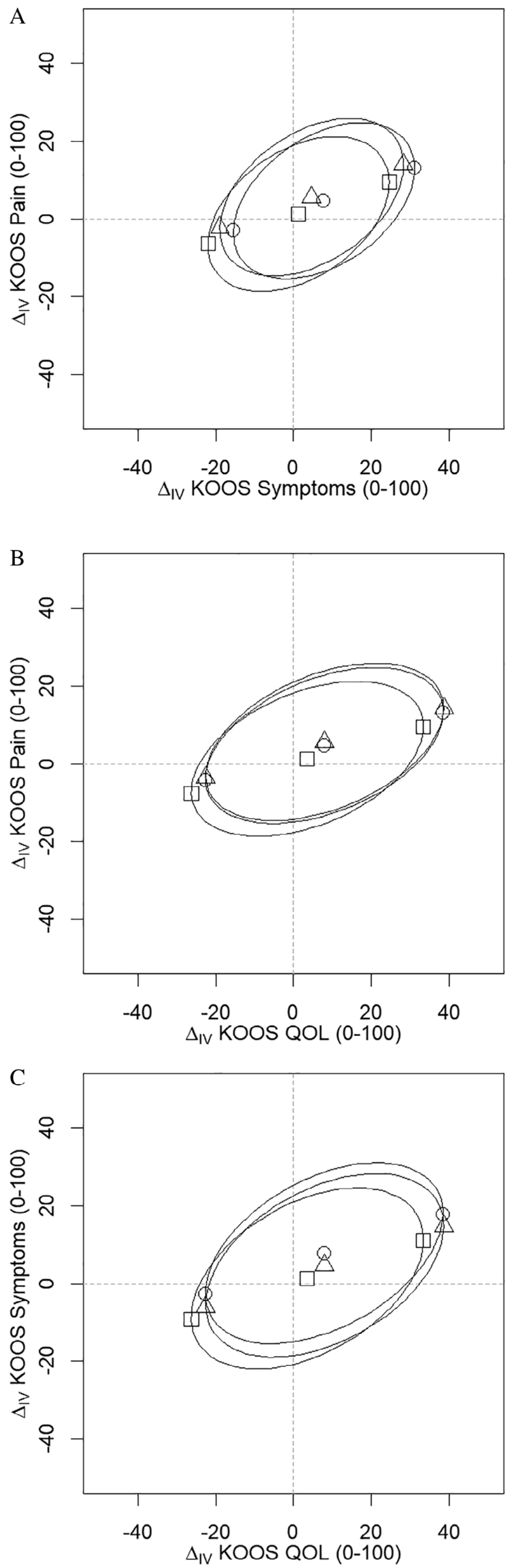
The Bayesian analysis results are posterior probability distributions (later posteriors) for the unknown parameters. These posteriors can be used to predict the response variables for a new female participant under different exercise programs. The predictive posterior distributions for pairs of the response variables are visualized by 2-dimensional ellipses that describe the 90% Bayes regions for estimated future values in different groups. With 90% probability, the response of a new patient (a randomly selected female patient with average commitment from the background population) will lie inside the 90% Bayes region. More details of these ellipses are given in Figure 2 of Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24553.

The predictive posterior distributions are summarized as probabilities of benefits. An intervention is considered beneficial if a randomly selected member of the exercise group with an average commitment has a larger change in response variable than a randomly selected member of the control group. Exact 50% probability of benefit is equal to throwing a coin when one predicts whether the exercise program will lead to a better result than the control treatment.

RESULTS

Baseline statistics. Participants who had at least 1 observed value in 7 outcome variables in either the postintervention or the follow-up measurement were included in the analysis. Within this population there were 10% of values missing in 14 response variables of 160 participants. Group sizes were 42 for the ART intervention group, 35 for the HLT intervention group, and 83 (AquaRehab n = 43, and LuRu n = 40) for the control group.

The average values and SDs of the background variables and response variables at baseline for the different study groups are summarized in Table 1. Participants in the AquaRehab study were on average ~6 years older and less active compared to participants in the LuRu study. For the response variables in the statistical model, differences from baseline, the group averages together with SDs, are presented in Tables 1–3 of Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24553.



Probabilities of benefits. The estimated probabilities of benefits are summarized in Figure 1A. These probabilities have been calculated for both time periods, from baseline to postintervention and from baseline to the end of the follow-up. The highest probabilities for benefits after intervention are seen in Vo_2 peak and symptoms in favor of ART. From the posteriors, it was calculated that ART intervention led to higher Vo_2 peak than the control, with a probability of 71%, and to better KOOS symptoms score, with a probability of 66%. All probability calculations are based on the parameter estimates of the Bayesian model and presented in Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24553.

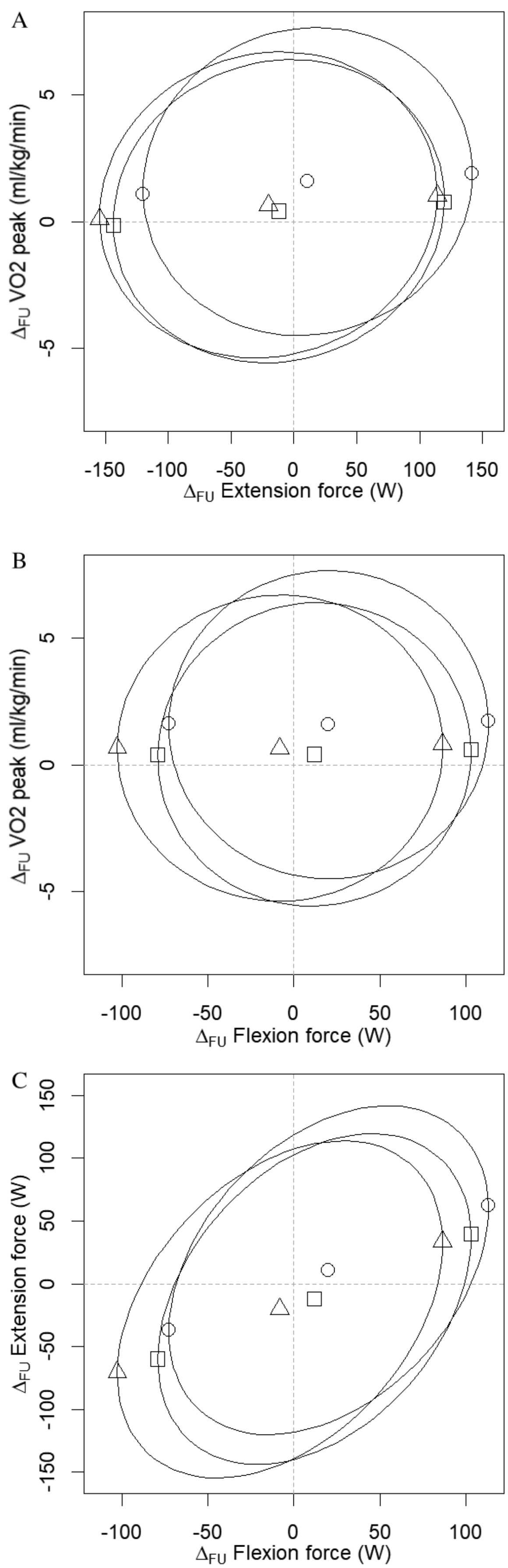
All variables other than flexion force have close to a 60% probability to have benefits after intervention with the HLT exercise program. However, the benefits in physical performance variables faded away during the follow-up period, ending up close to a 50% probability with most of the variables. The long-term benefits in flexion strength have only a 37% probability with HLT, meaning that the intervention is likely to be ineffective.

The effects for ART remained during the follow-up. As the highest long-term effect, with a probability of 67%, the ART led to fewer symptoms after the follow-up compared to controls. In addition, it was calculated that with a probability of 65%, ART is better than HLT for reducing symptoms in the long term (not shown in Figure 1). In all of the 7 variables, ART is likely to be beneficial even though the probabilities are not much higher than 50%.

In addition, Figure 1B reports the calculated probabilities for the most effective treatment from posteriors for 3 female participants with typical commitment. For instance, on average, the best result in reducing pain right after intervention is achieved by ART, with a probability of 36%, by HLT, with a probability of 41%, and by the control treatment, with a probability of 23%.

Personal effect modifiers. The estimated expected values for the latent personal effect modifiers were distributed between 0.58 and 1.53. This result indicates that the most committed patient had 53% more beneficial intervention than a patient with average commitment, and the least committed got 42% less benefits compared to average commitment. Even though we are interpreting this latent variable primarily as commitment, it also includes all other causes for the individual variation in benefits of

Figure 3. The mean change between baseline and postintervention and 2-dimensional 90% Bayes region for the prediction of Knee Injury and Osteoarthritis Outcome Score (KOOS) symptoms and KOOS pain (**A**), KOOS quality of life (QoL) and KOOS pain (**B**), and KOOS QoL and KOOS symptoms (**C**). The values in the top right indicate a positive outcome in both variables. There is 1 mean point and Bayes region for each group. Circles represent the aquatic resistance training group, triangles represent the high-impact aerobic land training group, and squares represent the control group.



the exercise program. For ART, the commitment had an average of 0.93 and SD of 0.19. For HLT, the average was 0.92 and the SD 0.15, thus there were no major differences between the groups regarding to the variability of commitment.

Pairwise predictions of variables. The ellipses in the Figures visualize the posteriors. The major axis of the ellipse is tilted for predictions that are correlated, such as the flexion force and the extension force. The ellipses also show the amount of individual variation in predictions of differences in the outcome variables. For a clear visualization, the variable power, which has the least differences between groups among physical performance variables, has been dropped out of these ellipse figures.

The ellipses in Figure 2 show the predictions for the differences after the intervention period in the physical performance variables. The overlapping ellipses indicate that the groupwise predictions have a lot of individual variation. The largest differences between the groups are seen on the variables Vo_2 peak and flexion force, where the ART group has the highest predicted improvement.

The predictions after the intervention period in the KOOS domains in Figure 3 indicate that all differences are correlated with each other. Both exercise groups have higher predictions than the control group on average on all variables, but there are a lot of individual variations.

The main finding in the follow-up predictions in Figure 4 is that there are not any signs of benefits of HLT exercise compared to the control group in physical performance variables. Vo_2 peak and extension force show the biggest differences in the ART group compared to other groups.

Figure 5 presents the follow-up predictions in the KOOS domains. The main difference compared to postintervention results is the lack of correlation between symptoms and other variables. Both exercise groups have slightly higher predictions than the control group on average, but there are a lot of individual differences.

DISCUSSION

This was the first Bayesian meta-analysis that estimates posterior probability distributions for exercise interventions in the management of OA. Differently from *P* values, posteriors allow us to calculate the probability of a hypothesis given the data.

Figure 4. The mean change between baseline and post–follow-up and 2-dimensional 90% Bayes region for the prediction of extension force and Vo_2 peak (**A**), flexion force and Vo_2 peak (**B**), and flexion force and extension force (**C**). The values in the top right indicate a positive outcome in both variables. There is 1 mean point and Bayes region for each group. Circles represent the aquatic resistance training group, triangles represent the high-impact aerobic land training group, and squares represent the control group. W = watts.

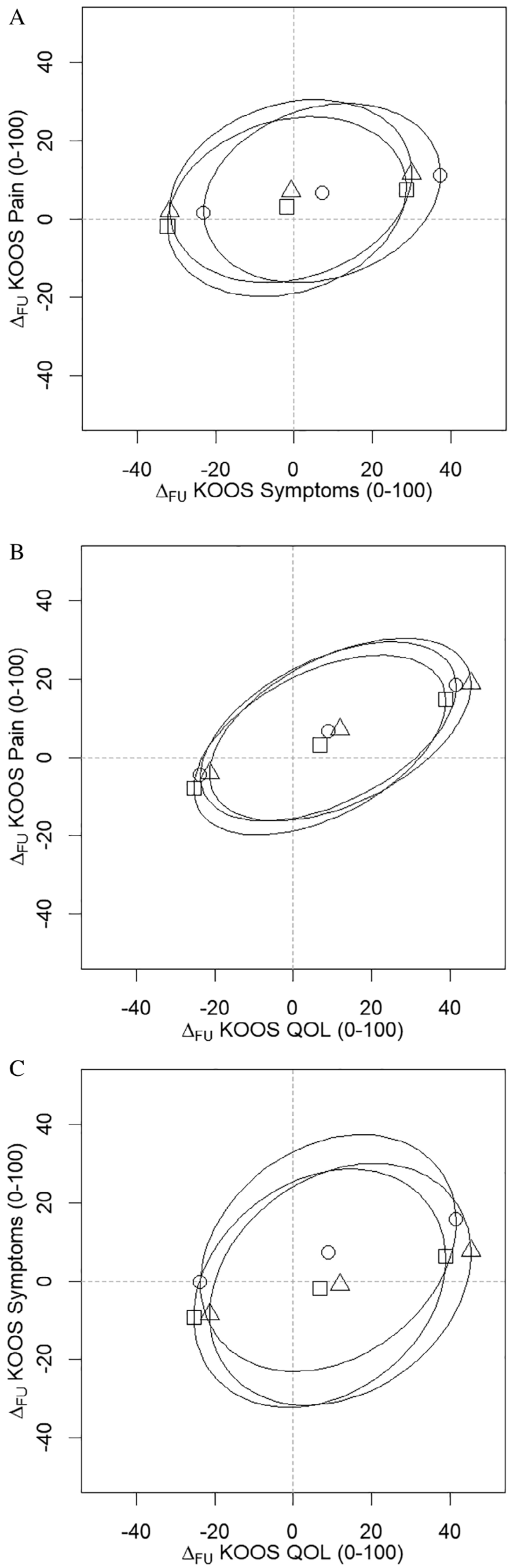


Figure 5. The mean change between baseline and post–follow-up and 2-dimensional 90% Bayes region for the prediction of Knee Injury and Osteoarthritis Outcome Score (KOOS) symptoms and KOOS pain (**A**), KOOS quality of life (QOL) and KOOS pain (**B**), and KOOS QOL and KOOS symptoms (**C**). The values in the top right indicate a positive outcome in both variables. There is 1 mean point and Bayes region for each group. Circles represent the aquatic resistance training group, triangles represent the high-impact aerobic land training group, and squares represent the control group.

The Bayesian model combining information from 2 different studies offered a possibility to calculate probabilities for benefits in multiple outcome measurements. Overall, ART seems to be slightly more beneficial than HLT.

The main short-term (from baseline to postintervention) effect was in cardiorespiratory fitness (V_{O_2} peak), which had a 71% probability to have a positive short-term benefit with ART intervention. Other short-term probabilities varied between 55% and 66% in ART, and HLT had a >50% probability with all other variables except flexion force. Thus, neither of these interventions predicted overall harm to the patients.

Interestingly, ART is considered an aerobic training intervention that would not be specific enough to improve muscle strength (20). However, results indicate that ART had an equal probability of a positive short-term outcome on muscle strength and power compared to HLT, which showed a lower probability of improvement in V_{O_2} peak (Figure 2).

As the main long-term (from baseline to follow-up) effect, ART led to benefits in self-reported OA-related symptoms, with a probability of 67%. The benefits on physical performance in the HLT group faded out during the follow-up period. This can be explained by the fact that subjects' exercise and physical activity intensity decreased after the intervention, thus leading to a detraining effect (21). Short duration interventions (up to a year) have only a short-term effect. Therefore, other interventions, including lifestyle chances and education, may be necessary in these populations to maintain training effects. Interestingly, based on participants' feedback, some found it difficult to find suitable training options after the cessation of the intervention; for example, the ART group found available aquatic exercise groups easy and not effective enough.

Unsurprisingly, the long-term probabilities for pain and quality of life stayed at the same ~60% level with both exercise programs, i.e., there was only a slightly better than coin toss chance to get a better result than with control treatment. It is important to notice that patients with severe knee OA were not included, and therefore the opportunity for change in the KOOS domains was limited (ceiling effect).

The ART program took 4 months compared to the HLT, which took 12 months, which for many participants is a significant difference for commitment. A longer period of exercise could be predicted to produce larger improvements in outcomes that are easier to maintain, but this was not the case in the current study. We found that the 4-month intensive ART had a slightly higher probability of improvement, which is in line with previous recommendations for intervention durations (6). However, one primary aim for the HLT intervention was to impart an effect on bone traits, and thus a longer intervention would be more suitable. Therefore, careful match of desired outcome should be considered when choosing which intervention to recommend.

Our results show that both ART and HLT had a moderately high probability of benefits, but we cannot strongly differentiate between the 2, which is in line with comparisons between land and aquatic exercise (22). The Bayes regions for different treatments were largely overlapping (Figures 2–5), which indicates that uncertainty on the individual level is large compared to the average treatment effects. Approximately 3% of the uncertainty in predictions was due to uncertainty in the estimated model parameters, and ~97% due to variation between individuals (see Supplementary Appendix A, available at http://onlinelibrary.wiley.com/doi/10.1002/acr.24553), and we were unable to perform counterfactual reasoning on the individual level. There are many interacting individual factors that have an impact on the positive or negative outcome of an exercise intervention, which is shown by in the biopsychosocial model (23). Given that commitment interacts with outcome, we can recommend that the type of the intervention is not as important as the willingness and ability to commit to the intervention.

Based on these results, a woman with mild knee OA who especially wants to improve muscle power in long term, can decide to take ART exercise. On the other hand, there could be another woman, to whom quality of life variables are more important than muscle power and who does not feel that comfortable in a swimming hall. She can decide to take HLT exercise because quality of life variables are not expected to be major differences between the exercises. A third woman could be especially interested in improving V_{O_2} peak and extension force. Her exercise decision could be supported with predictions in Figure 4A. These examples demonstrate the multicriteria nature of decision-making, and Bayesian meta-analysis offers support for comparing tradeoffs between personally important criteria.

The analysis has some limitations. The sizes of the exercise groups are not large enough to allow for reliable estimation of the interaction with individual level background variables such as age and body mass index. The control group was a combination of 2 control groups in different studies that had some differences: patients in the AquaRehab control group were ~6 years older than in the LuRu group, and the AquaRehab's intervention took 4 months compared to LuRu's 12 months of intervention. However, the length of the follow-up period was the same: 12 months in both studies.

In conclusion, Bayesian meta-analysis shows potential as a decision-making tool that can help with choosing between different exercise programs. As common advice for a random female patient without further information, this analysis suggests a slight preference of ART because it is likely to cause benefits in the short term and the long term for every variable. From the quality of life point of view, there does not seem to be any difference as to which exercise program to choose. In addition, in physical performance variables, even the highest probability (71%) leaves a lot of room for individual differences, and it is not necessary to force patients to exercise against their preferences. Instead, this research could be continued within this Bayesian framework toward more personalized recommendations. In a sufficiently large study, this could be done by taking background variables into account and developing a multicriteria decision-making procedure that recommends an exercise program based on the background information of a new patient and her motivation for different kinds of exercises.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Mr. Heikkinen had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Heikkinen, Waller, Munukka, Heinonen, Karvanen.

Acquisition of data. Waller, Munukka, Multanen, Heinonen.

Analysis and interpretation of data. Heikkinen, Karvanen.

REFERENCES

1. Fransen M, Mcconnell S, Harmer AR, van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee. Cochrane Database Syst Rev 2015;2015.
2. Uthman OA, van Der Windt DA, Jordan JL, Dziedzic KS, Healey EL, Peat GM, et al. Exercise for lower limb osteoarthritis: systematic review incorporating trial sequential analysis and network meta-analysis. Br J Sports Med 2014;48:1579.
3. Goh SL, Persson MS, Stocks J, Hou Y, Welton NJ, Lin J, et al. Relative efficacy of different exercises for pain, function, performance and quality of life in knee and hip osteoarthritis: systematic review and network meta-analysis. Sport Med 2019;49:743–61.
4. Batterham SI, Heywood S, Keating JL. Systematic review and meta-analysis comparing land and aquatic exercise for people with hip or knee arthritis on function, mobility and other health outcomes. BMC Musculoskelet Disord 2011;12:123.
5. Dong R, Wu Y, Xu S, Zhang L, Ying J, Jin H, et al. Is aquatic exercise more effective than land-based exercise for knee osteoarthritis? Medicine (Baltimore) 2018;97:e13823.
6. Juhl C, Christensen R, Roos EM, Zhang W, Lund H. 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.
7. Wasserstein RL, Lazar NA. The ASA's statement on P-values: context, process, and purpose. Am Stat 2016;70:129–33.
8. Gelman A, Carling JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis, 3rd ed. Chapman & Hall / CRC; 2013.
9. Waller B, Munukka M, Multanen J, Rantalainen T, Pöyhönen T, Nieminen MT, et al. Effects of a progressive aquatic resistance

exercise program on the biochemical composition and morphology of cartilage in women with mild knee osteoarthritis: protocol for a randomised controlled trial. BMC Musculoskelet Disord 2013;14:1–14.

10. Multanen J, Nieminen MT, Häkkinen A, Kujala UM, Jämsä T, Kautiainen H, et al. Effects of high-impact training on bone and articular cartilage: 12-month randomized controlled quantitative MRI study. J Bone Miner Res 2014;29:192–201.
11. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. Arthritis Rheum 1986;29:1039–49.
12. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 2012;10:28–55.
13. Munukka M, Waller B, Rantalainen T, Häkkinen A, Nieminen MT, Lammentausta E, et al. Efficacy of progressive aquatic resistance training for tibiofemoral cartilage in postmenopausal women with mild knee osteoarthritis: a randomised controlled trial. Osteoarthr Cartil 2016;24:1708–17.
14. Laukkanen RM, Oja P, Pasanen ME, Vuori IM. A two-kilometer walking test: effect of walking speed on the prediction of maximal oxygen uptake. Scand J Med Sci Sports 2007;3:263–6.
15. Sipilä S, Multanen J, Kallinen M, Era P, Suominen H. Effects of strength and endurance training on isometric muscle strength and walking speed in elderly women. Acta Physiol Scand 1996;156: 457–64.
16. Multanen J, Honkanen M, Häkkinen A, Kiviranta I. Construct validity and reliability of the Finnish version of the Knee Injury and Osteoarthritis Outcome Score. BMC Musculoskelet Disord 2018;19:155.
17. Bekkers JE, de Windt TS, Raijmakers NJ, Dhert WJ, Saris DB. Validation of the Knee Injury and Osteoarthritis Outcome Score (KOOS) for the treatment of focal cartilage lesions. Osteoarthritis Cartilage 2009; 17:1434–9.
18. R Core Team. R: a language and environment for statistical computing. 2020. URL: https://www.r-project.org/.
19. Stan Development Team. {RStan}: the {R} interface to {Stan}. 2020. URL: http://mc-stan.org.
20. Lund H, Weile U, Christensen R, Rostock B, Downey A, Bartels EM, et al. A randomized controlled trial of aquatic and land-based exercise in patients with knee osteoarthritis. J Rehabil Med 2008;40:137–44.
21. Waller B, Munukka M, Kujala UM, Heinonen AO. Response to the comments on "Effects of high intensity aquatic resistance training on body composition and walking speed in women with mild knee osteoarthritis: a 4-month RCT with 12-month follow-up." Osteoarthritis Cartilage 2017;25:e19–20.
22. Heywood S, McClelland J, Mentiplay B, Geigle P, Rahmann A, Clark R. Effectiveness of aquatic exercise in improving lower limb strength in musculoskeletal conditions: a systematic review and meta-analysis. Arch Phys Med Rehabil 2017;98:173–86.
23. Kanavaki AM, Rushton A, Efstathiou N, Alrushud A, Klocke R, Abhishek A, et al. Barriers and facilitators of physical activity in knee and hip osteoarthritis: a systematic review of qualitative evidence. BMJ Open 2017;7:e017042.