

# Designing and evaluating an online reinforcement learning agent for physical exercise recommendations in N-of-1 trials

Dominik Meier<sup>1</sup>  
Ipek Ensari<sup>2</sup>  
Stefan Konigorski<sup>1,2,3\*</sup>

DOMINIK.MEIER@STUDENT.HPI.DE  
IPEK.ENSARI@MSSM.EDU  
STEFAN.KONIGORSKI@HPI.DE

<sup>1</sup> *Hasso Plattner Institute for Digital Engineering, University of Potsdam, Potsdam, Germany*

<sup>2</sup> *Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, USA*

<sup>3</sup> *Department of Statistics, Harvard University, Cambridge MA, USA*

## Abstract

Personalized adaptive interventions offer the opportunity to increase patient benefits, however, there are challenges in their planning and implementation. Once implemented, it is an important question whether personalized adaptive interventions are indeed clinically more effective compared to a fixed gold standard intervention. In this paper, we present an innovative N-of-1 trial study design testing whether implementing a personalized intervention by an online reinforcement learning agent is feasible and effective. Throughout, we use a new study on physical exercise recommendations to reduce pain in endometriosis for illustration. We describe the design of a contextual bandit recommendation agent and evaluate the agent in simulation studies. The results show that, first, implementing a personalized intervention by an online reinforcement learning agent is feasible. Second, such adaptive interventions have the potential to improve patients' benefits even if only few observations are available. As one challenge, they add complexity to the design and implementation process. In order to quantify the expected benefit, data from previous interventional studies is required. We expect our approach to be transferable to other interventions and clinical interventions.

**Keywords:** N-of-1 Trials, Reinforcement Learning, Contextual Bandits, Physical Exercise Recommendations, Endometriosis

## 1. Introduction

Chronic diseases are frequently associated with between-patient variability in their symptomatic manifestations and other clinical features. Therefore, one-size-fits-all interventions and strategies are not ideal and yield heterogeneous results. This has led to a shift in study designs in recent years focusing on patients' individual responses, like N-of-1 trials (Lillie et al., 2011). N-of-1 trials are multi-crossover randomized controlled trials within one individual that in their classic set-up compare two interventions applied in a pre-determined crossover sequence. Then, individual intervention effects can be estimated and if a series of N-of-1 trials is performed with multiple participants, population-level intervention effects can also be estimated (Zucker et al., 1997). This approach can be particularly promising for chronic conditions without an established cure or treatment, making patient self-management an important disease management component. While N-of-1 trials are the gold standard study design for estimating individual causal intervention effects, comparing more than two interventions can be inefficient as each additional intervention condition linearly extends the total study duration. In recent work, models for adaptive N-of-1 trials have been proposed where the intervention allocation over time can depend on intermediate analyses of the trial (Senarathne et al., 2020; Malenica et al., 2021; Shrestha and Jain, 2021).

**Contributions.** As a first contribution, we describe an innovative novel study design that evaluates in an N-of-1 trial whether a personalized intervention is superior compared to a generic intervention, where

---

\* Corresponding author

the personalized intervention is itself an adaptive reinforcement learning (RL) agent which learns the best of multiple interventions in a given context. To the best of our knowledge, incorporating an RL agent as one intervention in an N-of-1 trial has not been reported before. Second, we describe the process of designing and evaluating an RL agent tailored to this study design. For this, we use a trial for illustration that investigates physical exercise recommendations to reduce pain in patients with endometriosis. A priori, it is not clear if a personalized RL agent is able to learn in a setting such as our presented pilot study. Our results show that implementing a personalized intervention by an online reinforcement learning agent is both feasible and can be effective, even if only few data points are available and learning is done on the individual level without pooling across participants.

Endometriosis is an estrogen-mediated inflammatory disease characterized by the growth of endometrial-like tissue outside the uterus, leading to painful adhesions and lesions. It is associated with debilitating pelvic pain and infertility, and a 7-year delay in diagnosis, in part due to significant between-patient symptomatic variability (Tamareisis et al., 2014; Schliep et al., 2015; Fourquet et al., 2019). Despite recent efforts, endometriosis remains poorly understood and is not well-managed, and there is a critical need to identify novel methods for disease management with a focus on pain reduction (Rogers et al., 2009). There is promising evidence on physical exercise for endometriosis pain management, however, it is unclear which physical activity with which intensity and duration is best (Bonochoer et al., 2014). In this setting, using adaptive methodology, which allows to directly incorporate individual responses into further intervention decisions, offers a great potential to increase the efficiency of the trial. Here, we combine the N-of-1 trial design with an adaptive intervention arm, designed to maximize patient’s benefit. In the adaptive arm, personalized physical exercise recommendations are given by an online RL agent.

The rest of the paper is structured as follows: In Section 2, we reference previous studies on exercise and pain in endometriosis and give background for our RL approach. In Section 3, we present a conceptual causal model of our assumed effects of exercise on endometriosis-related pain that serves as rationale for our selected interventions and study design and informs the design of the RL agent. Section 4 describes our study setup in detail. Section 5 describes

the requirements for the usage of an agent to select intervention decisions in this scenario, and Section 6 explains our chosen architecture, followed by an evaluation of the agent in simulation studies described in Section 7. We conclude with a discussion in Section 8.

## 2. Related work

### 2.1. Exercise interventions in endometriosis

Despite its prevalence of 1 in 10 women being affected worldwide, endometriosis is poorly managed and without well-established treatment or cure (Cramer and Missmer, 2002; Jarrell et al., 2005; Stratton et al., 2008; National Institutes of Health, NIH). Current medical intervention efforts in endometriosis focus on pharmacological symptom management, which is associated with unpleasant side effects and inadequate efficacy (Rogers et al., 2009; Sinatra, 2010; Krebs et al., 2018). Patients report a variety of strategies for symptom self-management. Of these, physical exercise remains the most promising and is associated with consistently favorable effects (Gonçalves et al., 2017; Armour et al., 2019a,b; Ensari et al., 2022). However, data are scarce and suggest variability across patients with respect to exercise preferences and contextual factors that influence efficacy (Bonochoer et al., 2014; Ricci et al., 2016; Belavy et al., 2021). Previous work further indicates that regular exercisers are more likely to experience decreased pain responses after exercise bouts and that a wide range of modalities are utilized, including aerobic, calisthenic, and stretching-oriented activities (Ensari et al., 2022).

For building interventions in our study, we follow the standard exercise prescription method within clinical exercise physiology which considers 4 components: type, duration, intensity, frequency (Franklin, 2021). Physiologists typically update the individual’s prescription at each follow-up appointment based on the individual’s responses. We propose to automate this learning and updating process using RL.

### 2.2. Contextual bandits in adaptive trials

Adaptive designs allow responding to collected data quickly and efficiently (Pallmann et al., 2018). Having learned adaptive sequences, Just-in-Time Adaptive Interventions can be developed which provide interventions to patients based on contextual features (Nahum-Shani et al., 2018). In the adaptive recommendations’ field, studies have mainly focused

on investigating which personalized messaging can help to improve the proximal outcome. For example, the HeartSteps project used an RL algorithm based on Thompson Sampling to personalize notifications for sedentary patients (Liao et al., 2020). In another study, Boltzmann sampling was used to personalize messages for sedentary patients with diabetes (Yom-Tov et al., 2017). In the SleepBandits project, patients had the possibility to self-experiment with different interventions on their sleep. Recommendations were provided based on Thompson Sampling, which led to increased sampling of beneficial conditions (Daskalova et al., 2020).

### 3. Conceptual model for pain modulation via exercise

The overall goal of this study is to design an RL agent that can efficiently identify the optimal dose (w.r.t. type, intensity, duration) of exercise for pain reduction for a given individual. To achieve this, the RL agent needs to learn quickly with limited data and determine the optimal exercise intervention per individual. This requires consideration of the temporality of the pain modulating effects of exercise based on expert knowledge. Exercise has both acute (i.e., immediate) and delayed (i.e., distal) effects acting through various possible pathways, and we depict these timelines and our assumed causal graph in Figure 1. We focus on immediate effects. Nevertheless, we give a brief overview of all effects in the following.

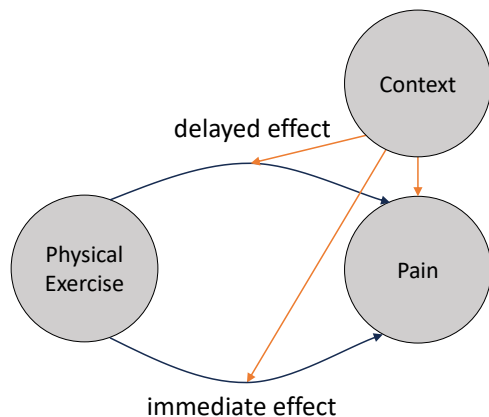


Figure 1: Causal graph with immediate and delayed effects. Context moderates effect on pain but also influences pain.

#### 3.1. Immediate effects of exercise on pain

There are various psychological and physiological factors that can explain the immediate (i.e., lasting up to several hours post-exercise bout) pain modulating effects of exercise in endometriosis.

**Psychological effects** Psychological factors such as a sense of accomplishment and/or self-efficacy, accompanied by an overall positive mood state, can favorably modulate post-exercise pain responses (Middelkamp et al., 2017). Similarly, the distraction hypothesis of exercise has been demonstrated for pain and related outcomes (Nolen-Hoeksema et al., 1993; Villemure and Bushnell, 2002; Mikkelsen et al., 2017).

**Physiological effects** Evidence suggests acute improvements in different types (e.g., pressure, thermal, cold, intensity) of pain sensitivity post-exercise (Nau-gle et al., 2012), accompanied by changes in immediate hormone profile (West and Phillips, 2012) and autonomic responses (e.g., reduced pain receptor reactivity) (Uzawa et al., 2023). Similarly, some types of pain (e.g., nociceptive, somatic) can respond favorably to muscular relaxation and enhanced movement in the pain-impacted area (Louw et al., 2016). As endometriosis pain often originates from multiple body areas, the type of the pain varies (Ensari et al., 2020).

#### 3.2. Delayed effects of exercise on pain

Besides immediate effects, there is evidence for delayed effects of exercise on chronic pain.

**Improved physical function** A wider range of musculoskeletal and other physiological improvements are expected in the long term with exercise, which subsequently can improve pain sensitivity and reactivity and associated quality of life (Ambrose and Golightly, 2015; Rice et al., 2019)

**Improved endocrine function** Especially relevant for endometriosis, there is strong evidence for exercise-mediated endocrine system, particularly estrogen and other sex hormones, for pain modulation in menstrual pain disorders (Jaleel et al., 2022).

**Improved inflammatory markers** Different exercise intensities and modalities have been shown to improve inflammatory marker profiles (Bote et al., 2013) and inflammatory responses (Athanasidou et al., 2023). This is of particular relevance to endometriosis given its inflammatory nature.

**Other neuromodulator adaption** Existing evidence suggests that physical activity modulates central nervous system excitability and inhibition, and that chronic exercise can downregulate endogenous pain pathways (Sluka et al., 2018).

### 3.3. Moderation through context

We expect moderation of both immediate and delayed effects by the current context of the patient, for which we assess past and current (i) intensity and duration of activities and (ii) endometriosis-related pain.

Because we expect pain measurements over time to be autocorrelated, context also influences pain in our model. The context variables will be used to personalize the exercise recommendations by the agent. Pain areas can vary across patients and over time. Here, we consider pain in the lower abdominal (including pelvis), groin and hip areas as primary endometriosis-related pain outcomes. As secondary outcomes, we consider other body locations tracked by the patients, which is accomplished through use of a visual body map similar to the visual body map in the McGill Pain Scale (Melzack, 1975) that allows tracking all body areas to indicate pain location, as well as intensity and type of the pain.

## 4. Study design of the N-of-1 trial

In this section, we describe the proposed pilot N-of-1 trial, see Figure 2 for a visualization. After an initial baseline phase of one week, patients are randomized into an A-B or B-A design, where each A or B phase is 2 weeks long. The baseline phase serves to preconfigure the agent using tracked datapoints. An initial assessment is used to define the set of exercises available for recommendation. In Phase A, patients receive generic physical exercise recommendations based on the published PA Guidelines for adults (U.S. Dept. of Health and Human Services, 2018). In Phase B, patients receive RL-generated personalized exercise recommendations. All exercise recommendations are delivered through an mHealth app, which allows integrating real-time data from patients into the recommendations. Of note, while typical physical exercise guidelines assume exercises done 3–5 times per week, in this pilot study we propose daily exercises so that more observations can be made under each intervention, providing an easier setting for the RL agent to learn. As outcomes, patients will self-track current

symptoms (including pain location, type, intensity), daily functioning, health behavior, and exercise in the morning, evening, before, and after their exercise through the app. The primary outcome for the RL agent is pain measured on a visual analog scale (VAR). In addition, a weekly PROMIS pain interference questionnaire<sup>1</sup> is administered. The study has been reviewed and approved by the Mount Sinai IRB Board (STUDY-23-00721).

## 5. Requirements for a personalized agent inside an N-of-1 trial

In the following section, we describe requirements for the RL agent. For this, we adapt the PCS framework (Trella et al., 2022) to a personalized RL agent inside an intervention arm for an N-of-1 trial.

### 5.1. Personalization (P)

Few data points need to be sufficient for effective personalization. Phase B of our pilot study will only include 14 decision points for the agent over two weeks. These few decision points make efficient personalization challenging. To overcome this, we use the interventional data from the baseline phase, as well as from the A phase in case of an A-B randomization, for agent pre-configuration.

### 5.2. Computability (C)

Model updates and recommendations need to be feasible. This requires timely collection of contextual information from patients and timely intervention decisions. This is guaranteed by integrating the agent in an mHealth platform when implementing the study.

### 5.3. Stability (S)

The agent may never harm patients. We use the beginning assessment to define a safe subset of possible interventions to ensure this. Additionally, the agent needs to be robust in real-life scenarios. This includes robustness against model misspecification, missing-data and non-adherence to exercise recommendations. The pilot format of our study requires that the data generated by the agent are usable for post-trial inference and for informing the implementation of a follow-up study.

1. [https://cde.nida.nih.gov/sites/nida\\_cde/files/PROMIS%20Adult%20ShortForm%20v1.0%20Pain%20Interference%206b.pdf](https://cde.nida.nih.gov/sites/nida_cde/files/PROMIS%20Adult%20ShortForm%20v1.0%20Pain%20Interference%206b.pdf)

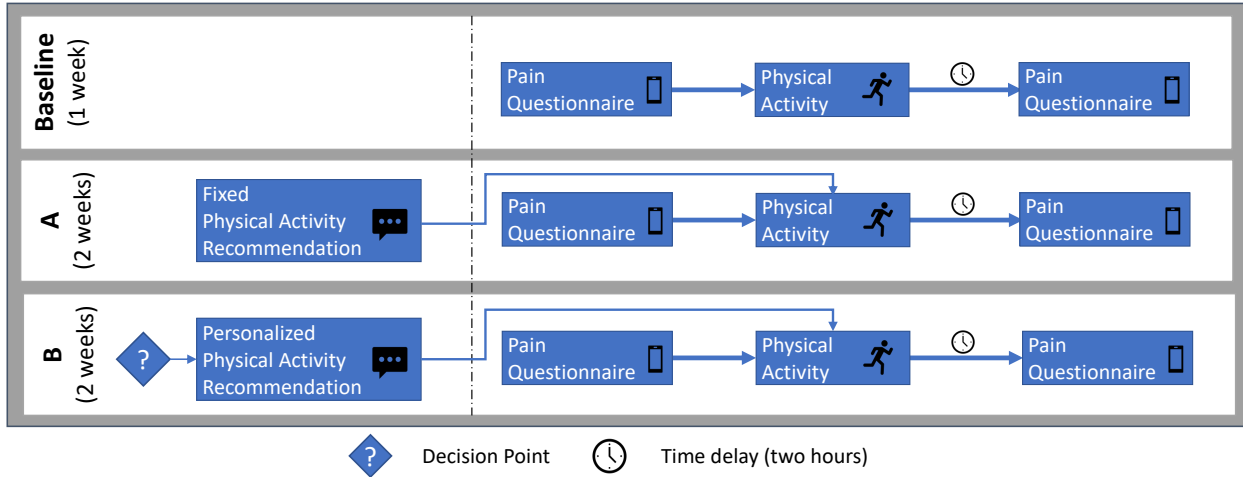


Figure 2: Five weeks randomized pilot study design with Baseline followed by randomized A-B or B-A.

In the baseline phase, no recommendations are given, but patients can track exercising. In Phase A, fixed exercise schedules are recommended, which is compared to Phase B, in which personalized exercise recommendations are provided by a Reinforcement Learning agent. The agent uses data points from all previous phases to make a personalized recommendation. Throughout the five weeks, data will be collected using wearables and questionnaires for later analysis.

## 6. Agent design

In this section, we describe our design decisions for the agent. To design our agent, we make two simplifying assumptions: First, based on the description in Section 3, we assume that the state is fully captured by the context variables and that there is no hidden state. Hence, the context moderates the effect of physical exercise on pain. Second, we assume that it is sufficient to pick the recommendation which is most effective at the current time point, and disregard long-term planning of recommendations. Similar assumptions have been made in many mHealth intervention settings (e.g., Liao et al. (2020); Figueroa et al. (2021)), and allow fast learning in few data point environments. Making these assumptions allows us to model the selection of a recommendation out of the predetermined set as an instance of a contextual multi-armed bandit problem. We design the RL agent as a contextual bandit using a Thompson sampling policy based on a Bayesian model. Next, we describe the action space, policy, and reward function of the agent.

### 6.1. Action space

The action space defines which interventions can be selected. Based on the beginning assessment of patients, we create a personalized set of physical exercise recommendations. Each action is defined by type, duration, and intensity.

### 6.2. Policy

We choose Thompson Sampling as underlying policy of the agent due to its widespread use in adaptive trials (e.g., Daskalova et al. (2020); Liao et al. (2020)) and empirically good performance in diverse settings (Chapelle and Li, 2011). Thompson Sampling uses the probability of being the optimal intervention in a given context as selection probability of an intervention. We define the optimal intervention as the intervention with the largest pain reduction, and estimate the probability by fitting a Bayesian model for prediction of the measured pain reduction  $\Delta pain$ . We assume that using the difference in pain measurements pre- and post-exercise allows us to only measure the immediate effect of the exercise and reduce autocorrelation in the time series, which we otherwise would have to incorporate in the model.

### 6.2.1. BAYESIAN MODEL

We model the pain reduction as a sum of four individual terms, which capture the different assumed effects of exercise type, intensity, duration, and pain. Let  $i = 1 \dots n$  denote patients,  $j = 1 \dots k$  denote the exercise recommendation in the personalized set of physical exercises, and let  $t = 1 \dots T$  denote the decision points. Each exercise is characterized by its type, intensity, and duration and we define

$$\text{Burden}_{ij} = \text{Intensity}_{ij} \cdot \text{Duration}_{ij},$$

where exercise intensity is manually assigned a value between 0 and 1 and duration is assessed in minutes and divided by the maximum duration of any exercise. As each exercise type is assigned an intensity, it affects burden through its intensity. This allows the agent to understand that some interventions are similar in intensity and duration. See [Table 1](#) for an example.

As inputs for the calculation of the expected pain reduction, we define

$\text{pain}_{i,t}$  = Current pain of patient  $i$  at  $t$ .

$mi_{i,t}$  = Mean intensity of last 3 exercises for  $i$  at  $t$

$md_{i,t}$  = Mean duration of last 3 exercises for  $i$  at  $t$

Our Bayesian linear regression model includes four summands:  $\tau_{\text{Type},i,j}$  models an intercept for each type of physical exercise per patient.  $y_{\text{Intensity},i,j,t}$  models the effect of the intensity of the physical exercise on pain. We model this as a linear interaction between the intensity of the proposed physical exercise and the mean intensity of the last three exercises. Similarly,  $y_{\text{Duration},i,j,t}$  models this effect for duration of exercise.  $y_{\text{Burden},i,j,t}$  models an interaction effect between current pain and burden. We use burden instead of individual coefficients for duration and intensity to decrease the number of parameters. This gives the following model formula:

$$\Delta \text{pain}_{i,j,t} \sim \text{Normal}(\mu_i, \sigma_i) \quad (1)$$

$$\mu_i = \tau_{\text{Type},i,j} + y_{\text{Intensity},i,j,t}$$

$$+ y_{\text{Duration},i,j,t} + y_{\text{Burden},i,j,t}$$

$$y_{\text{Intensity},i,j,t} = (\alpha_i + \beta_i \cdot mi_{i,t}) \cdot \text{Intensity}_{i,j}$$

$$y_{\text{Duration},i,j,t} = (\gamma_i + \delta_i \cdot md_{i,t}) \cdot \text{Duration}_{i,j}$$

$$y_{\text{Burden},i,j,t} = (\eta_i + \kappa_i \cdot \text{pain}_{i,t}) \cdot \text{Burden}_{i,j}$$

This model has 7 parameters plus the number of different types of activities for each patient. We

use  $\sigma \sim \text{Exponential}(1)$ , and  $\alpha, \beta, \gamma, \delta, \eta, \kappa, \tau \sim \text{Normal}(0, 1)$  as non-informative priors.

### 6.3. Reward

We use  $\Delta \text{pain}_{i,j,t}$  as reward, which is the decrease in pain measured two hours after performing the exercise relative to the pain measured before the exercise on the VAR scale.

### 6.4. Implementation

The model was implemented and fitted using the NUTS Sampler provided by PYMC ([Salvatier et al., 2016](#)). To generate the probability of best intervention, we sampled the posterior predictive distribution with the parameters set to the patient's context to obtain an estimated pain reduction. The relative frequencies of the greatest pain reduction were then used as the selection distribution for Thompson Sampling.

## 7. Evaluation

In this section, we describe the evaluation of the performance of our proposed agent. Our goal is to assess whether we can confidently deploy the agent in the pilot study. As data from previous studies is not available for evaluating our proposed agent, we test robustness and personalization across different scenarios in a simulation study.

### 7.1. Simulation setup

The evaluation consists of three main steps. First, we define different scenarios, which each correspond to different assumptions on the environment. Second, to quantify performance, we translate the PCS requirements (see [Section 5](#)) into metrics. Third, we simulate the behavior of the agent for 100 patients for each of the seven scenarios and each of the two randomization schemes (A-B, B-A), for a total of 14 simulations, and report the performance of the agent with respect to the defined metrics.

For our simulation, we use a set of eight possible exercise recommendations shown in [Table 1](#). The simulation mirrors the study configuration. The baseline phase and the A phase in case of A-B randomization were simulated using a fixed sequence of these exercise recommendations and used as input data for the agent. Our simulation did not include missing values.

Type	Intensity	Duration	Comment
0	0.3	0.5 (30min)	Slow jogging
0	0.5	0.5 (30min)	Jogging
0	0.7	0.5 (30min)	Fast jogging
1	1	0.1 (6min)	HIIT
1	1	0.2 (12min)	HIIT
1	1	0.3 (18min)	HIIT
2	0.5	0.75 (45min)	Swimming
3	0.1	1 (60min)	Yoga

Table 1: Example set of physical exercise recommendations used for simulation

The Python code and results are available from the GitHub repository<sup>2</sup>.

## 7.2. Evaluation scenarios

In Section 3, we presented the assumed effects of exercise on pain. As we have little information about the true effects in the graph in Figure 1, we generate six different possible scenarios (I-VI) in order to test stability (see Subsection 5.3) across different environments. An additional Scenario VII is included to test the effects of informative non-adherence:

**Scenario I** Pain reduction is drawn independently of other variables from a Normal(0,1) distribution. This scenario is included as a null scenario.

**Scenario II** Pain reduction is calculated from the linear model specified in Equation 1. All parameters are drawn individually for each patient from a Normal(0,1) distribution, and are then fixed for this patient across the study. This environment is included to test the behavior of the agent when the underlying model is correctly specified.

**Scenarios III, IV, V, VI** Pain reduction in Scenarios III, IV, V, VI is defined as in Scenario II, but sets the parameters regarding type (Scenario III), or intensity (Scenario IV), or duration (Scenario V), or intensity and duration (Scenario VI) to 0. These environments are included to test robustness of the agent against partially violated assumptions of the Bayesian model.

**Scenario VII** To investigate the behavior of the agent when data is missing not at random, we in-

clude Scenario VII. Pain reduction is calculated like in Scenario II, but informative non-adherence using the following rationale is simulated: If the recommendation that is given by the agent would increase the pain (based on the linear model), this data point is dropped in 50% of the cases, and will not be used as feedback to the agent.

## 7.3. Metrics

We evaluate three key metrics that have similarly been proposed by Trella et al. (2022) to assess the personalization of an agent:

- Average of users’ average regret
- The 25th percentile of users’ average regret
- Average regret for multiple time points

We define the regret as the improvement over the non-adaptive arm, calculated as the difference between pain reduction under the fixed arm and the adaptive arm simulated with the same context variables:

$$\text{Regret}_{i,t} = \sum_{l=1}^t \Delta \text{pain}_{i,\text{fixed},l} - \sum_{l=1}^t \Delta \text{pain}_{i,\text{adaptive},l}$$

Since our pilot study is meant to inform further research, it is important to generate a dataset useful for post-trial inference. The greatest threat to meaningful post-trial inference is if some participants only receive the same intervention, due to fast convergence of the Bayesian model. We measure the diversity of interventions by five different metrics: For assessing the different types of exercise recommendations (discrete case), we calculate the Shannon entropy (Shannon, 1948) as a measure of randomness. For the continuous variables intensity and duration, we calculate the standard deviation. Additionally, we calculate the mean of the minimum and maximum probability for intervention selection.

## 7.4. Results

As seen in Table 2, and Figure 3, the agent reaches a negative regret for all scenarios except Scenario I. Clear downward trends are visible for the median and quartiles. The regret in Scenario I is zero, since under the null scenario, the pain reduction is independent of the chosen intervention.

2. <https://github.com/HIALab/Reinforcement-learning-agents-in-N-of-1-trials>

In the B-A randomization, the agent only has seven data points from the baseline to learn the participant’s coefficients. We therefore observe higher regret (i.e., worse performance) compared to A-B randomization. We observe that regret is slightly negative in the 25% worst case. This shows that personalization not only works for the average patient but for most patients. The mean maximum selection probability across all environments is 44%, while the minimum is at 1%. This suggests that low values for regret do not come from selection of good interventions, but rather through discarding bad interventions.

In the non-adaptive arm with the generic intervention, the entropy is 1.23 for exercise type and the standard deviations are 0.26 for duration and 0.32 for type. In the adaptive arm applying the RL agent, we mostly observe smaller values for entropy and standard deviation. Especially the entropy of exercise type in Scenario II is low with 0.82 in the A-B design. This suggests that a correctly specified model leads to a fast model convergence. At the same time, this will create high beneficence for patients in pain reduction. The non-informative adherence in Scenario VII did worsen the 25% worst case performance slightly, as well as the mean performance in the B-A design. In the A-B design, the agent still performed similar to Scenario II.

## 8. Discussion

Here, we have presented an innovative N-of-1 trial study design comparing a personalized intervention by an online RL agent to a generic intervention. We focused on the study design and the design of the RL agent, while leaving the design of post-trial inference to follow-up studies. Regarding the RL agent design, the results of simulation studies show that employing such an RL agent can be feasible also when only few data points are available and learning is done on the individual level without pooling across patients.

Our proposed agent showed the ability to learn in different scenarios, therefore allowing to create study designs which are flexible across different plausible scenarios. In this low-sample setting and even when the set of possible interventions is large, we expect benefit from personalization.

Finding the right balance between specialized and general agents remains a challenge. Specialization can lead to higher optimization potential, e.g., through including prior knowledge, but at the same time might decrease robustness against model mis-

specification. Due to the limited data available for our pilot trial, we prioritized robustness and selected a more general agent.

Under the assumption that data is missing completely at random, missing data will slow down the learning of the agent due to less available data points, but would not interfere with the learning of the Bayesian model. Once our pilot study will be completed, we will investigate the behavior of the agent when data is missing not at random in this real data.

Our work has some limitations. For real-world applications, additional challenges such as missing data need to be considered. While we made simplifying assumptions in our simulation study, such challenges might make it necessary to include additional safety measures inside the agent. Also, due to time constraints for the total length of the study and our main goal of proofing the feasibility of integrating an RL agent inside an N-of-1 trial, we opted for daily exercise recommendations to collect more data points. This is more frequent than usual exercise frequencies in typical studies. However, we expect that our results generalize to longer studies with less frequent exercising, and are planning to reduce the frequency in follow-up studies. Finally, it will be interesting to investigate model misspecification not only when the model is misspecified (see Scenario I) but also when the priors in the Bayesian model are not well selected.

A natural extension and possible requirement in some situations to improve the learning rate in a low-sample setting is to include pooling, which has been effectively deployed in previous studies. However, it might also worsen personalization in highly heterogeneous environments (Tomkins et al., 2020; Trella et al., 2022). Judging how and to which extent pooling might improve the agent for our scenario is hard to estimate without additional data. Since we already observed benefit for participants without pooling, we decided to disregard it for this pilot study.

We are curious to see the benefit of personalizing agents for pain reduction in endometriosis and beyond, and to observe the increasing use of N-of-1 trials for their evaluation.

## References

Kirsten R. Ambrose and Yvonne M. Golightly. Physical exercise as non-pharmacological treatment of chronic pain: Why and when. *Best Practice & Research. Clinical Rheumatology*, 29(1):120–130, Feb



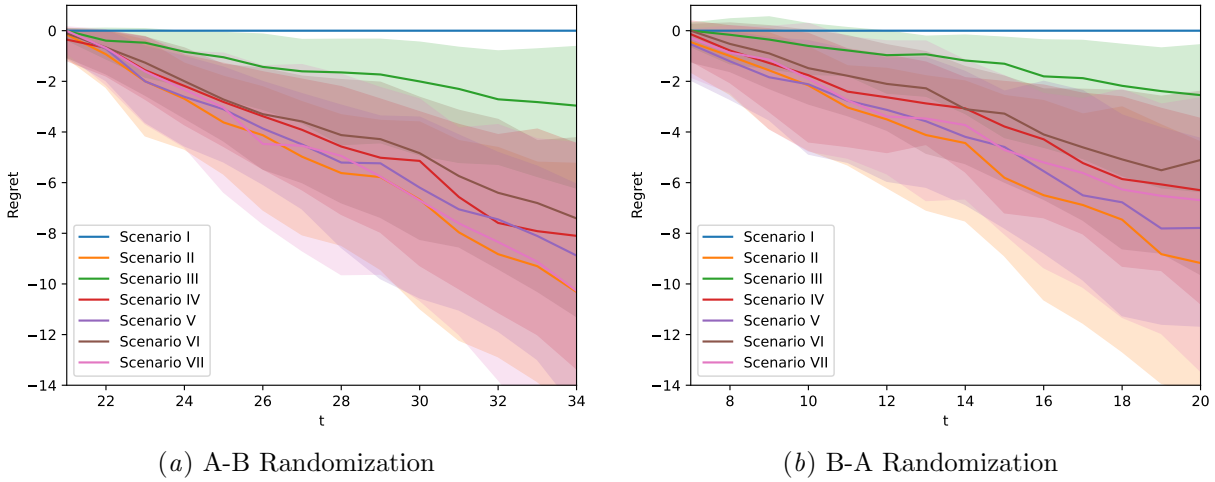


Figure 3: Results of simulation study showing the regret over time  $t$ . Lines show the median regret over time across the 100 simulated patients. Shaded areas show the 25% and 75% quantiles.

Scenario	Design	Regret		Entropy Type	Standard Deviation		Mean Probability	
		Mean	0.75 quantile		Duration	Intensity	Max	Min
I	A-B	0.00	0.00	1.10	0.25	0.30	0.29	0.05
	B-A	0.00	0.00	1.12	0.27	0.31	0.34	0.03
II	A-B	-11.76	-5.24	0.82	0.22	0.25	0.44	0.01
	B-A	-10.50	-4.40	0.90	0.24	0.27	0.46	0.01
III	A-B	-5.12	-0.63	0.93	0.23	0.26	0.36	0.03
	B-A	-4.40	-0.56	0.98	0.26	0.28	0.39	0.02
IV	A-B	-9.98	-4.46	0.85	0.22	0.25	0.43	0.01
	B-A	-8.32	-3.47	0.95	0.25	0.28	0.44	0.02
V	A-B	-10.74	-6.07	0.81	0.22	0.25	0.42	0.01
	B-A	-9.48	-4.26	0.90	0.24	0.27	0.44	0.01
VI	A-B	-8.76	-4.23	0.86	0.23	0.27	0.42	0.01
	B-A	-7.02	-2.40	0.96	0.25	0.28	0.43	0.02
VII	A-B	-11.74	-4.42	0.82	0.22	0.25	0.44	0.01
	B-A	-8.37	-2.66	0.97	0.25	0.28	0.43	0.02

Table 2: Results for evaluation in test bed, across 100 simulated patients.

2015. ISSN 1532-1770. doi: 10.1016/j.berh.2015.04.022. [e53a0a2978c28872a4505bdb51db06dc-Paper.pdf](#).
- M Armour, CC Ee, D Naidoo, Z Ayati, KJ Chalmers, KA Steel, MJ de Manincor, and E Delshad. Exercise for dysmenorrhoea. *Cochrane Database of Systematic Reviews*, (9), 2019a. ISSN 1465-1858. doi: 10.1002/14651858.CD004142.pub4. URL <https://doi.org/10.1002/14651858.CD004142.pub4>.
- Mike Armour, Justin Sinclair, K. Jane Chalmers, and Caroline A. Smith. Self-management strategies amongst australian women with endometriosis: a national online survey. *BMC complementary and alternative medicine*, 19(1):17, Jan 2019b. ISSN 1472-6882. doi: 10.1186/s12906-019-2431-x.
- Nikolaos Athanasiou, Gregory C. Bogdanis, and George Mastorakos. Endocrine responses of the stress system to different types of exercise. *Reviews in Endocrine & Metabolic Disorders*, 24(2): 251–266, Apr 2023. ISSN 1573-2606. doi: 10.1007/s11154-022-09758-1.
- Daniel L. Belavy, Jessica Van Oosterwijck, Matthew Clarkson, Evy Dhondt, Niamh L. Mundell, Clint T. Miller, and Patrick J. Owen. Pain sensitivity is reduced by exercise training: Evidence from a systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*, 120:100–108, Jan 2021. ISSN 1873-7528. doi: 10.1016/j.neubiorev.2020.11.012.
- Camila M Bonocher, Mary L Montenegro, Julio C Rosa e Silva, Rui A Ferriani, and Juliana Meola. Endometriosis and physical exercises: a systematic review. *Reproductive Biology and Endocrinology: RB&E*, 12:4, Jan 2014. ISSN 1477-7827. doi: 10.1186/1477-7827-12-4.
- Maria Elena Bote, Juan Jose Garcia, Maria Dolores Hinchado, and Eduardo Ortega. Fibromyalgia: anti-inflammatory and stress responses after acute moderate exercise. *PloS One*, 8(9):e74524, 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0074524.
- Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011. URL [https://proceedings.neurips.cc/paper\\_files/paper/2011/file/](https://proceedings.neurips.cc/paper_files/paper/2011/file/e53a0a2978c28872a4505bdb51db06dc-Paper.pdf)
- Daniel W. Cramer and Stacey A. Missmer. The epidemiology of endometriosis. *Annals of the New York Academy of Sciences*, 955:11–22; discussion 34–36, 396–406, Mar 2002. ISSN 0077-8923. doi: 10.1111/j.1749-6632.2002.tb02761.x.
- Nediyana Daskalova, Jina Yoon, Yibing Wang, Cintia Araujo, Guillermo Beltran, Nicole Nugent, John McGeary, Joseph Jay Williams, and Jeff Huang. Sleepbandits: Guided flexible self-experiments for sleep. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–13, New York, NY, USA, Apr 2020. Association for Computing Machinery. ISBN 978-1-4503-6708-0. doi: 10.1145/3313831.3376584. URL <https://dl.acm.org/doi/10.1145/3313831.3376584>.
- Ipek Ensari, Adrienne Pichon, Sharon Lipsky-Gorman, Suzanne Bakken, and Noémie Elhadad. Augmenting the clinical data sources for enigmatic diseases: A cross-sectional study of self-tracking data and clinical documentation in endometriosis. *Applied Clinical Informatics*, 11(5): 769–784, Oct 2020. ISSN 1869-0327. doi: 10.1055/s-0040-1718755.
- Ipek Ensari, Sharon Lipsky-Gorman, Emma N. Horan, Suzanne Bakken, and Noémie Elhadad. Associations between physical exercise patterns and pain symptoms in individuals with endometriosis: a cross-sectional mhealth-based investigation. *BMJ Open*, 12(7), 2022. ISSN 2044-6055. doi: 10.1136/bmjopen-2021-059280. URL <https://bmjopen.bmj.com/content/12/7/e059280>.
- Caroline A Figueroa, Adrian Aguilera, Bibhas Chakraborty, Arghavan Modiri, Jai Aggarwal, Nina Deliu, Urmimala Sarkar, Joseph Jay Williams, and Courtney R Lyles. Adaptive learning algorithms to optimize mobile applications for behavioral health: guidelines for design decisions. *Journal of the American Medical Informatics Association*, 28(6):1225–1234, June 2021. ISSN 1527-974X. doi: 10.1093/jamia/ocab001.
- Jessica Fourquet, Diego E. Zavala, Stacey Missmer, Nabal Bracero, Josefina Romaguera, and Idhaliz Flores. Disparities in healthcare services in women

- with endometriosis with public vs private health insurance. *American Journal of Obstetrics and Gynecology*, 221(6):623.e1–623.e11, Dec 2019. ISSN 1097-6868. doi: 10.1016/j.ajog.2019.06.020.
- Barry A. Franklin. Evolution of the acsm guidelines: Historical perspectives, new insights, and practical implications. *ACSM’s Health & Fitness Journal*, 25(2):26, Apr 2021. ISSN 1091-5397. doi: 10.1249/FIT.0000000000000645.
- Andrea Vasconcelos Gonçalves, Nelson Filice Barros, and Luis Bahamondes. The practice of hatha yoga for the treatment of pain associated with endometriosis. *Journal of Alternative and Complementary Medicine (New York, N.Y.)*, 23(1):45–52, Jan 2017. ISSN 1557-7708. doi: 10.1089/acm.2015.0343.
- Ghufuran Jaleel, Muhammad Abu Shaphe, Abdur Raheem Khan, Deepak Malhotra, Huma Khan, Sana Parveen, Mohammed Qasheesh, Rashid Ali Beg, Aksh Chahal, Fuzail Ahmad, and Md Faruque Ahmad. Effect of exercises on central and endocrine system for pain modulation in primary dysmenorrhea. *Journal of Lifestyle Medicine*, 12(1):15–25, Jan 2022. ISSN 2234-8549. doi: 10.15280/jlm.2022.12.1.15.
- John Jarrell, Ruchi Mohindra, Sue Ross, Paul Taenzer, and Rollin Brant. Laparoscopy and reported pain among patients with endometriosis. *Journal of obstetrics and gynaecology Canada: JOGC = Journal d’obstetrique et gynecologie du Canada: JOGC*, 27(5):477–485, May 2005. ISSN 1701-2163. doi: 10.1016/s1701-2163(16)30531-x.
- Erin E. Krebs, Amy Gravely, Sean Nugent, Agnes C. Jensen, Beth DeRonne, Elizabeth S. Goldsmith, Kurt Kroenke, Matthew J. Bair, and Siamak Noorbaloochi. Effect of opioid vs nonopioid medications on pain-related function in patients with chronic back pain or hip or knee osteoarthritis pain: The space randomized clinical trial. *JAMA*, 319(9):872–882, Mar 2018. ISSN 1538-3598. doi: 10.1001/jama.2018.0899.
- Peng Liao, Kristjan Greenewald, Predrag Klasnja, and Susan Murphy. Personalized heartsteps: A reinforcement learning algorithm for optimizing physical activity. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 4(1):18, Mar 2020. ISSN 2474-9567. doi: 10.1145/3381007.
- Elizabeth O. Lillie, Bradley Patay, Joel Diamant, Brian Issell, Eric J. Topol, and Nicholas J. Schork. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Personalized Medicine*, 8(2):161–173, Mar 2011. ISSN 1741-0541. doi: 10.2217/pme.11.7.
- Adriaan Louw, Kory Zimney, Emilio J. Puentedura, and Ina Diener. The efficacy of pain neuroscience education on musculoskeletal pain: A systematic review of the literature. *Physiotherapy Theory and Practice*, 32(5):332–355, Jul 2016. ISSN 1532-5040. doi: 10.1080/09593985.2016.1194646.
- Ivana Malenica, Aurelien Bibaut, and Mark J. van der Laan. Adaptive sequential design for a single time-series, 2021. URL <https://arxiv.org/abs/2102.00102>.
- Ronald Melzack. The mcgill pain questionnaire: major properties and scoring methods. *Pain*, 1(3):277–299, Sep 1975. ISSN 0304-3959. doi: 10.1016/0304-3959(75)90044-5.
- Jan Middelkamp, Maaïke van Rooijen, Peter Wolfhagen, and Bert Steenbergen. The effects of a self-efficacy intervention on exercise behavior of fitness club members in 52 weeks and long-term relationships of transtheoretical model constructs. *Journal of Sports Science & Medicine*, 16(2):163–171, Jun 2017. ISSN 1303-2968.
- Kathleen Mikkelsen, Lily Stojanovska, Momir Polenakovic, Marijan Bosevski, and Vasso Apostolopoulos. Exercise and mental health. *Maturitas*, 106:48–56, Dec 2017. ISSN 1873-4111. doi: 10.1016/j.maturitas.2017.09.003.
- Inbal Nahum-Shani, Shawna N. Smith, Bonnie J. Spring, Linda M. Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A. Murphy. Just-in-time adaptive interventions (jitais) in mobile health: Key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine*, 52(6):446–462, May 2018. ISSN 1532-4796. doi: 10.1007/s12160-016-9830-8.
- National Institutes of Health (NIH), 2023. Estimates of funding for various research, condition, and disease categories, 2023. URL <https://report.nih.gov/funding/categorical-spending#/>. Accessed: 2023-09-06.

- Kelly M. Naugle, Roger B. Fillingim, and Joseph L. Riley. A meta-analytic review of the hypoalgesic effects of exercise. *The Journal of Pain*, 13(12): 1139–1150, Dec 2012. ISSN 1528-8447. doi: 10.1016/j.jpain.2012.09.006.
- S. Nolen-Hoeksema, J. Morrow, and B. L. Fredrickson. Response styles and the duration of episodes of depressed mood. *Journal of Abnormal Psychology*, 102(1):20–28, Feb 1993. ISSN 0021-843X. doi: 10.1037//0021-843x.102.1.20.
- Philip Pallmann, Alun W. Bedding, Babak Choodari-Oskoei, Munyaradzi Dimairo, Laura Flight, Lisa V. Hampson, Jane Holmes, Adrian P. Mander, Lang’o Odondi, Matthew R. Sydes, Sofía S. Villar, James M. S. Wason, Christopher J. Weir, Graham M. Wheeler, Christina Yap, and Thomas Jaki. Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC Medicine*, 16(1):29, Feb 2018. ISSN 1741-7015. doi: 10.1186/s12916-018-1017-7.
- Elena Ricci, Paola Viganò, Sonia Cipriani, Francesca Chiaffarino, Stefano Bianchi, Giorgia Rebonato, and Fabio Parazzini. Physical activity and endometriosis risk in women with infertility or pain: Systematic review and meta-analysis. *Medicine*, 95(40):e4957, Oct 2016. ISSN 1536-5964. doi: 10.1097/MD.0000000000004957.
- David Rice, Jo Nijs, Eva Kosek, Timothy Wideman, Monika I. Hasenbring, Kelli Koltyn, Thomas Graven-Nielsen, and Andrea Polli. Exercise-induced hypoalgesia in pain-free and chronic pain populations: State of the art and future directions. *The Journal of Pain*, 20(11):1249–1266, Nov 2019. ISSN 1528-8447. doi: 10.1016/j.jpain.2019.03.005.
- Peter A. W. Rogers, Thomas M. D’Hooghe, Asgerally Fazleabas, Caroline E. Gargett, Linda C. Giudice, Grant W. Montgomery, Luk Rombauts, Lois A. Salamonsen, and Krina T. Zondervan. Priorities for endometriosis research: recommendations from an international consensus workshop. *Reproductive Sciences (Thousand Oaks, Calif.)*, 16(4):335–346, Apr 2009. ISSN 1933-7205. doi: 10.1177/1933719108330568.
- John Salvatier, Thomas V. Wiecki, and Christopher Fonnesebeck. Probabilistic programming in python using pymc3. *PeerJ Computer Science*, 2:e55, Apr 2016. ISSN 2376-5992. doi: 10.7717/peerj-cs.55.
- K. C. Schliep, S. L. Mumford, C. M. Peterson, Z. Chen, E. B. Johnstone, H. T. Sharp, J. B. Stanford, A. O. Hammoud, L. Sun, and G. M. Buck Louis. Pain typology and incident endometriosis. *Human Reproduction (Oxford, England)*, 30(10):2427–2438, Oct 2015. ISSN 1460-2350. doi: 10.1093/humrep/dev147.
- Siththara Gedara J. Senarathne, Antony M. Overstall, and James M. McGree. Bayesian adaptive n-of-1 trials for estimating population and individual treatment effects. *Statistics in Medicine*, 39(29): 4499–4518, Dec 2020. ISSN 0277-6715, 1097-0258. doi: 10.1002/sim.8737.
- C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. doi: 10.1002/j.1538-7305.1948.tb01338.x.
- Sama Shrestha and Sonia Jain. A bayesian-bandit adaptive design for n-of-1 clinical trials. *Statistics in Medicine*, 40(7):1825–1844, Mar 2021. ISSN 0277-6715, 1097-0258. doi: 10.1002/sim.8873.
- Raymond Sinatra. Causes and consequences of inadequate management of acute pain. *Pain Medicine (Malden, Mass.)*, 11(12):1859–1871, Dec 2010. ISSN 1526-4637. doi: 10.1111/j.1526-4637.2010.00983.x.
- Kathleen A. Sluka, Laura Frey-Law, and Marie Hoeger Bement. Exercise-induced pain and analgesia? underlying mechanisms and clinical translation. *Pain*, 159 Suppl 1(Suppl 1):S91–S97, Sep 2018. ISSN 1872-6623. doi: 10.1097/j.pain.0000000000001235.
- Pamela Stratton, Ninet Sinaii, James Segars, Deloris Koziol, Robert Wesley, Carolyn Zimmer, Craig Winkel, and Lynnette K. Nieman. Return of chronic pelvic pain from endometriosis after raloxifene treatment: a randomized controlled trial. *Obstetrics and Gynecology*, 111(1):88–96, Jan 2008. ISSN 0029-7844. doi: 10.1097/01.AOG.0000297307.35024.b5.
- John S. Tamareisis, Juan C. Irwin, Gabriel A. Goldfien, Joseph T. Rabban, Richard O. Burney, Camran Nezhat, Louis V. DePaolo, and Linda C. Giudice. Molecular classification of endometriosis and disease stage using high-dimensional genomic data. *Endocrinology*, 155(12):4986–4999, Dec 2014. ISSN 1945-7170. doi: 10.1210/en.2014-1490.

- Sabina Tomkins, Peng Liao, Predrag Klasnja, and Susan Murphy. Intelligentpooling: Practical thompson sampling for mhealth. (arXiv:2008.01571), Dec 2020. URL <http://arxiv.org/abs/2008.01571>. arXiv:2008.01571 [cs, stat].
- Anna L. Trella, Kelly W. Zhang, Inbal Nahum-Shani, Vivek Shetty, Finale Doshi-Velez, and Susan A. Murphy. Designing reinforcement learning algorithms for digital interventions: Pre-implementation guidelines. *Algorithms*, 15(88): 255, Aug 2022. ISSN 1999-4893. doi: 10.3390/a15080255.
- U.S. Dept. of Health and Human Services, 2018. U.s. department of health and human services. physical activity guidelines for americans 2018, 2018. URL <https://health.gov/our-work/nutrition-physical-activity/physical-activity-guidelines/current-guidelines>. Accessed: 2023-09-06.
- Hironobu Uzawa, Kazuya Akiyama, Hiroto Furuyama, Shinta Takeuchi, and Yusuke Nishida. Autonomic responses to aerobic and resistance exercise in patients with chronic musculoskeletal pain: A systematic review. *PloS One*, 18(8):e0290061, 2023. ISSN 1932-6203. doi: 10.1371/journal.pone.0290061.
- Chantal Villemure and Catherine M. Bushnell. Cognitive modulation of pain: how do attention and emotion influence pain processing? *Pain*, 95(3):195–199, Feb 2002. ISSN 0304-3959. doi: 10.1016/S0304-3959(02)00007-6.
- Daniel W. D. West and Stuart M. Phillips. Associations of exercise-induced hormone profiles and gains in strength and hypertrophy in a large cohort after weight training. *European Journal of Applied Physiology*, 112(7):2693–2702, Jul 2012. ISSN 1439-6327. doi: 10.1007/s00421-011-2246-z.
- Elad Yom-Tov, Guy Feraru, Mark Kozdoba, Shie Mannor, Moshe Tennenholtz, and Irit Hochberg. Encouraging physical activity in patients with diabetes: Intervention using a reinforcement learning system. *Journal of Medical Internet Research*, 19(10):e338, Oct 2017. ISSN 1438-8871. doi: 10.2196/jmir.7994.
- D. R. Zucker, C. H. Schmid, M. W. McIntosh, R. B. D’Agostino, H. P. Selker, and J. Lau. Combin-
- ing single patient (n-of-1) trials to estimate population treatment effects and to evaluate individual patient responses to treatment. *Journal of Clinical Epidemiology*, 50(4):401–410, Apr 1997. ISSN 0895-4356. doi: 10.1016/s0895-4356(96)00429-5.