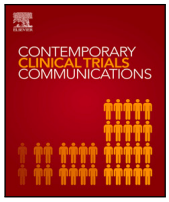
Contemporary Clinical Trials Communications 38 (2024) 101282

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc



Research paper

Analyzing population-level trials as N-of-1 trials: An application to gait

Lin Zhou [a], Juliana Schneider [b], Bert Arnrich [a], Stefan Konigorski [b,c,d,*]

[a] Digital Health - Connected Healthcare, Hasso Plattner Institute, University of Potsdam, Potsdam, Germany
[b] Digital Health & Machine Learning, Hasso Plattner Institute, University of Potsdam, Potsdam, Germany
[c] Department of Statistics, Harvard University, Cambridge, USA
[d] Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, NY, USA

ARTICLE INFO

Keywords:
Bayesian repeated-measures model
Gait
N-of-1 trials

ABSTRACT

Studying individual causal effects of health interventions is important whenever intervention effects are heterogeneous between study participants. Conducting N-of-1 trials, which are single-person randomized controlled trials, is the gold standard for their analysis. As an alternative method, we propose to re-analyze existing population-level studies as N-of-1 trials, and use gait as a use case for illustration. Gait data were collected from 16 young and healthy participants under fatigued and non-fatigued, as well as under single-task (only walking) and dual-task (walking while performing a cognitive task) conditions. As a reference to the N-of-1 trials approach, we first computed standard population-level ANOVA models to evaluate differences in gait parameters (stride length and stride time) across conditions. Then, we estimated the effect of the interventions on gait parameters on the individual level through Bayesian repeated-measures models, viewing each participant as their own trial, and compared the results. The results illustrated that while few overall population-level effects were visible, individual-level analyses revealed differences between participants. Baseline values of the gait parameters varied largely among all participants, and the effects of fatigue and cognitive task were also heterogeneous, with some individuals showing effects in opposite directions. These differences between population-level and individual-level analyses were more pronounced for the fatigue intervention compared to the cognitive task intervention. Following our empirical analysis, we discuss re-analyzing population studies through the lens of N-of-1 trials more generally and highlight important considerations and requirements. Our work encourages future studies to investigate individual effects using population-level data.

1. Introduction

In the majority of research studies, the focus lies on identifying average effects in a population of individuals, such as in large cohort studies or randomized controlled trials (RCTs). However, especially if there are heterogeneous individual effects, it can be of great interest to investigate associations on an individual level. Estimating and testing these individual effects pose challenges. One approach is to employ statistical or machine learning models to estimate individual effects from the population-level studies. To this end, different methods have been proposed in recent years [1–4]. As another approach, a new study can be designed with the specific aim of investigating individual-level effects. For this, the study design of so-called N-of-1 trials has been established as the gold standard [5]. In N-of-1 trials, the effect of one or more interventions is investigated in an individual by measuring the outcome of interest over time across alternating phases in which the interventions are applied. As such, N-of-1 trials are multi-crossover single-person RCTs [6,7]. As a third approach, which we propose in this study, population-level data can be re-analyzed through the lens of N-of-1 trials. This approach can be generally applied to experimental studies that use a repeated-measures design in which two or more conditions with interventions are measured repeatedly. We illustrate this for one specific study in the following.

In order to analyze individual data as N-of-1 trials, it is important that multiple measurements are available for each individual. This can be aided if sensors are used that capture data in short intervals. For example, heart rate variability can be calculated within time windows of several minutes, and continuous recording on a smartwatch can generate a large number of measurements for analysis [8]. Another example is gait analysis, which is the study of walking patterns. Devices such as instrumented walkways, or wearable sensors can measure hundreds of strides in minutes [9]. There are various other use cases

* Corresponding author.
E-mail addresses: lin.zhou@hpi.de (L. Zhou), stefan.konigorski@hpi.de (S. Konigorski).

https://doi.org/10.1016/j.conctc.2024.101282
Received 8 March 2023; Received in revised form 8 February 2024; Accepted 28 February 2024
Available online 29 February 2024

where our proposed analysis method may apply, for illustration, here we focus on an application to gait.

Gait can be quantified using spatiotemporal parameters, such as stride length, stride time, speed, or cadence. These gait parameters provide crucial insights into a person's health status. For example, gait speed and variability have been associated with life expectancy and risk for falls in older adults [10,11]. Although gait is typically regarded as an isolated and highly automatic task, evidence suggests that gait patterns differ when concurrently performing a secondary task (e.g., cognitive or motor interference task). Such dual-task situations, which closely mimic daily life walking [12,13], have been associated with slower gait speeds and increased stride times [14–17]. Consequently, studying gait under these conditions may provide clinically relevant insights into gait modulations in daily life.

A more comprehensive understanding of gait in real-life walking should also consider the aspect of physical fatigue because it knowingly affects gait kinematics and kinetics, and is linked to a higher risk of slip-induced falls as well as impaired movement controls in young and old healthy adults [18–20]. Existing studies investigating the effects of muscle fatigue on gait performance have reported heterogeneous outcomes. For example, for young healthy adults, Granacher et al. [21] observed statistically significant decreases in gait speed and stride length in a fatigue condition, while Parijat et al. [18] reported no statistically significant changes in gait speed. In older healthy adults, muscle fatigue only resulted in rather moderate changes in gait parameters [19]. Regarding stride length, some studies reported an increase [21–23], while others reported no changes [20,24,25].

One possible explanation for the aforementioned discrepancy in the observed effects of fatigue could be that the group-level analyses typically performed in gait studies do not capture the heterogeneous gait changes among individuals. The above-mentioned studies did not investigate heterogeneity among their participants, and none of the studies included covariates such as gender, body height or mass in their analysis. However, it is known that gait characteristics are highly individualized and persist for a long period. As reported by Horst et al. classification accuracy for identifying 46 healthy individuals using their gait patterns remained at 99% for one year [26]. Moreover, there is evidence that gait changes in response to interventions, such as athletic training or disease treatment, are also individualized [27–29].

The highly individualized nature of gait and gait modifications suggests that individual-level analyses could provide insights that are not evident from population-level analyses. In the context of the effects of fatigue and dual-task on gait, studying individual gait responses can allow assessing and monitoring an individual's risk for falls and mobility impairments. Typically, statistical models have accounted for inter-personal differences by including covariates, or by including random effects to better approximate each individual's own average gait response. In these cases, the estimand is still a population average and the individual responses are not characterized. Therefore, N-of-1 approaches are needed to adequately characterize the individual response. However, to the best of our knowledge, only one series of N-of-1 trials has been conducted on gait, in which Maguire et al. compared the effect of different walking aids on gait and balance for chronic stroke patients and revealed different responses across the participants [30].

Here, we investigate how existing data from population-level studies can be re-analyzed through the lens of N-of-1 trials to estimate individual-level effects. To this aim, we use data from a population-based study in a repeated-measures design that investigated the effects of physical fatigue and cognitive task on gait [31]. We estimate personalized gait parameters (stride length and stride time) from Bayesian repeated-measures models, and compare the results with a standard population-level ANOVA model. Finally, we discuss re-analyzing population studies through the lens of N-of-1 trials more generally and highlight important considerations and requirements.

2. Materials and methods

2.1. Overview of the gait study

Sixteen young healthy participants (eight males, eight females) were enrolled in the study. Eligibility for the study was determined using the Physical Activity Readiness Questionnaire (PAR-Q) and only participants without medical restrictions for performing physical activities (i.e. with all negative responses) were allowed to take part in the study. At the first visit (see study design below), personal characteristics were assessed, including the physical activity levels of the participants on a scale of 1 (low) - 3 (high) using the International Physical Activity Questionnaire (IPAQ) [32].

Fig. 1 shows an overview of the study design. The study consisted of two visits, referred to as visits A and B in the following, which were seven days apart. The order of A and B was randomized among participants, and at each of the two visits, the participants performed two walking assessments, separated by a fatigue protocol. During visit A, participants first completed a 6-minute walking assessment in a corridor with a distance of 35 meters in one direction. Then, the participants performed a repeated sit-to-stand all-out fatigue protocol to induce muscular fatigue in the lower limbs. Immediately following the fatigue protocol, participants repeated another 6-minute walking assessment. During visit B, the experimental procedure was the same as in visit A, except that the participants had to count numbers while walking (i.e., the dual-task condition). Details of the fatigue protocol and the cognitive task are described in Supplementary Text A.

Our primary focus of the measurements obtained in the walking sessions was on two gait parameters: stride length and stride time. These parameters were captured using inertial measurement units (IMUs) attached to the participants' shoes, which measured tri-axial acceleration and angular velocity of the foot movement. Details of the IMU gait analysis methods are described in Supplementary Text A. In total, gait parameters from four walking sessions were collected for each participant: single-task control (ST-Control), single-task fatigue (ST-Fatigue), dual-task control (DT-Control) and dual-task fatigue (DT-Fatigue). Hence, each participant had either the intervention sequence ST-Control – ST-Fatigue – DT-Control – DT-Fatigue, or the sequence DT-Control – DT-Fatigue – ST-Control – ST-Fatigue. Hundreds of gait measurements were made during each block.

The study was approved by the ethics committee of the University of Potsdam (number 63/2020), and all experiments were conducted according to the latest revision of the Declaration of Helsinki. All participants provided written consent prior to data collection.

2.2. Statistical analyses

Descriptive statistics and population-level analyses

As a first step, we computed descriptive statistics of age, body mass, height, and the IPAQ physical activity level of all participants. Next, we performed a population-level analysis using a two-way repeated measures ANOVA to serve as a reference for the comparison to the N-of-1 trial analyses. In ANOVA, we used stride length and stride time as outcomes and tested for the effect of physical fatigue and cognitive tasks, which were included as fixed factors. In addition, age, body height and body mass were included as covariates in the model. The model is described in detail in Supplementary Text B.

N-of-1 trials analysis using Bayesian repeated-measures models

In our main analysis, we analyzed the data through the lens of N-of-1 trials. For each participant, we estimated the individual marginal effects of the physical fatigue intervention and cognitive intervention on the gait parameters stride length and stride time. In contrast to typical N-of-1 trials with multiple crossovers, the data from our study consists of four blocks of repeated measurements of the outcome gait

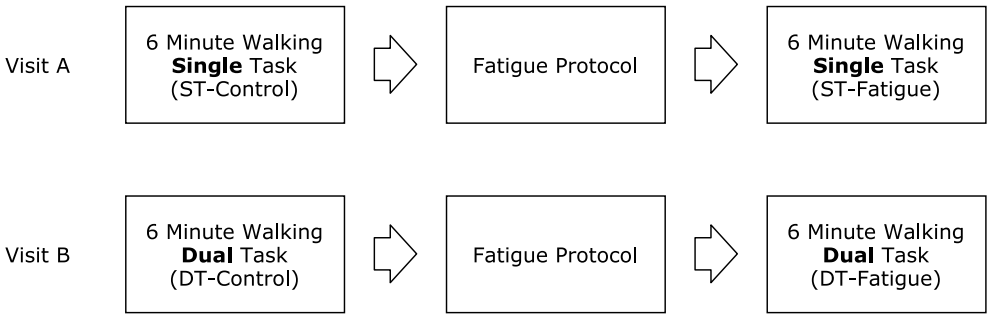


Fig. 1. Study design. The study consisted of two visits which were seven days apart. The order of visits A and B was randomized among participants. During each visit, the participants performed two 6-minute walking assessments, separated by a fatigue protocol. During visit B, the participants performed a number-counting task while walking. ST: Single-Task (only walking), DT: Dual-Task (walking while counting numbers).

parameters for each participant. This 2 × 2 experimental design is typical in population-level studies [21].

We used Bayesian linear repeated-measures models to fit probabilistic models of the data distribution to the gait time series data, separately for each participant, and separately for stride length and stride time. Such Bayesian models provide a probabilistic description of the data for interpretation [33] and allow for the incorporation of prior knowledge, which is not as easily or transparently done in conventional frequentist analysis. A model with a first-order autoregressive (AR1) error structure was used, which acknowledges that (for the same person) the covariance between errors from the observations may not be equal, and decreases towards zero with increasing lag. The AR1 error structure has been recommended in the literature for single-subject time series data in the context of Bayesian hypothesis testing [34].

For a more detailed description, let $y_i, i = 1, \dots n$, denote the ith observation (i.e., stride time or stride length of the ith stride) of a participant in the study. It is worth noting that the total number of observations n varies for each participant. For ease of notation, we use n in the following in our description of the linear model that is fit for each individual separately:

$$y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i2} \cdot X_{i3} + \epsilon_i = \boldsymbol{X_i \beta} + \epsilon_i, \tag{1}$$

where $\boldsymbol{X} = (X_1, X_2, X_3, X_2 \cdot X_3)$ is the $(n \times 4)$ design matrix. $X_1 = \mathbf{1_n}$ is a vector of ones of length n that represents the intercept. X_2 denotes whether the individual was in the cognitive task condition (dual task, $X_{i2} = 1$) or not (single task, $X_{i2} = 0$), and X_3 denotes whether the individual was in the fatigue condition ($X_{i3} = 1$) or in the control condition ($X_{i3} = 0$). Further, $X_2 \cdot X_3$ denotes the interaction between cognitive task and fatigue. $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)^T$ is a vector of fixed effects. Hence, the mean gait parameters under the four walking conditions were estimated using the following combinations of $\boldsymbol{\beta}$ coefficients. ST-Control: β_1, DT-Control: $\beta_1 + \beta_2$, ST-Fatigue: $\beta_1 + \beta_3$, DT-Fatigue: $\beta_1 + \beta_2 + \beta_3 + \beta_4$. $\boldsymbol{\epsilon}$ represents the error drawn from a multivariate normal distribution:

$$\boldsymbol{\epsilon} \sim MVN(\mathbf{0}, \sigma^2 \boldsymbol{\Psi}) \tag{2}$$

where σ^2 is the error variance, and $\boldsymbol{\Psi}$ is a variance–covariance matrix determined by the AR1 process [34], as in Eq. (3):

$$\boldsymbol{\Psi} = \frac{1}{1-\phi^2} \begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^{n-1} \\ \phi & 1 & \phi & \dots & \phi^{n-2} \\ \phi^2 & \phi & 1 & \dots & \phi^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \dots & 1 \end{bmatrix} \tag{3}$$

Here, ϕ is the correlation coefficient between two successive time points. The exponent of ϕ reflects the decline of the correlation between time points according to their distance from one another.

The Markov Chain Monte Carlo (MCMC) method with Gibbs sampling was used. Among all parameters of the model, the parameter of primary interest is the vector $\boldsymbol{\beta}$, since combinations of its elements make up the mean gait parameter distributions for the four walking conditions (i.e., ST-Control, ST-Fatigue, DT-Control and DT-Fatigue). While an informative prior distribution for the parameter β_1 can be directly inferred from studies on normal gait parameters of young healthy adults [35], there is not enough information available to assume priors for the other parameters. As a result, we chose to use non-informative priors in the main analyses. We assumed a half-Cauchy distribution for the residual variance σ (cf. Eq. (2)) in the Bayesian model as described in [36], with default priors recommended by Gelman et al. [37]. In sensitivity analyses we tested different informative priors, the details are described in the "Sensitivity analyses" section below. Table 1 provides an overview of the priors.

In the sampling procedure, we used 2 chains, 5000 burn-in steps, 1 thinning step (i.e. no thinning), and 10,000 iterations. To reduce the amount of computation, data used for the AR1 model were taken only from the left foot, and down-sampled by a factor of five (i.e., selecting every fifth sample sequentially). We confirmed with visual inspection that down-sampling did not change the overall distribution of the data. In addition, there was no obvious change in the gait parameters over time with each walking condition for the same person. Example plots of the data can be found in Supplementary Text A.

From the Bayesian models and conditional parameters, our aim was to estimate the marginal effects of physical fatigue or cognitive task, which we expect to vary among the participants. To estimate the probability of the marginal effect being clinically meaningful, we compared the effect to a pre-defined threshold [38]. More concretely, the marginal effects of fatigue or cognitive task were derived as a function of the $\boldsymbol{\beta}$ coefficients of the AR1 model (see details in Supplementary Text B). Then, the posterior probability distribution of the marginal effects was estimated along with other model parameters in the AR1 model. Subsequently, we defined the probability of the marginal effect being meaningful as the posterior probability of the estimated marginal effect being greater than a given threshold (i.e., the area under the curve of the posterior distribution bounded below by this threshold, following Stunnenberg et al. [38]). Throughout the text, we refer to this as the posterior probability of a meaningful effect or change (e.g., as in Fig. 3). For our example dataset, the thresholds for meaningful change were determined based on previous studies on the effects of fatigue in a similar population. We defined clinically meaningful effects on stride length and stride time to be 3% and 2%, respectively [21,22]. This can be generalized to a variety of settings to investigate the probability of meaningful effects for an individual. Future studies adopting this method should define the meaningful effect threshold based on the specific domain context.

The convergence of the MCMC chain was confirmed with potential scale reduction factor (PSRF) and trace plots. MCMC chain resolution was evaluated using the effective sample size (ESS), which measures the efficiency of Monte Carlo methods such as MCMC [39]. More details on the MCMC diagnostics and on their results can be found in Supplementary Text C. To confirm that the posterior estimates accurately represent

the observed data, a posterior predictive check was performed by comparing the posterior distributions with the distribution of the observed samples. More specifically, the posterior distribution of the intercept and effects were used to reconstruct the modeled distributions of gait parameters under the four conditions. These modeled distributions are then compared with the observed sample distributions using boxplots. More details are described in Supplementary Text C. The Bayesian analysis was performed using JAGS version 4.3.0, run from R version 4.1.1 (R Project for Statistical Computing). Formal specifications of the JAGS models can be found in Supplementary Text B. The data and R scripts used for running the analysis can be found at https://github.com/HIAlab/gait_nof1trials.

Sensitivity analyses

To test how well alternative Bayesian models can estimate the posterior distribution, we implemented two additional models, a simple basic model and a time covariate model. In contrast to the AR1 model introduced in Section 2.2, both these models assumed that the errors are independent and identically distributed. As a basic model, we implemented a simple Bayesian fixed effects model with the two fixed factors of fatigue exposure and cognitive task type. Similar to the AR1 model, we assumed a linear relationship and normally-distributed errors, but we assumed here that each data point, namely, each stride from the same recording session, is independent of each other. The model structure is identical to that described in Eq. (1), except that the error term follows a normal distribution with a diagonal covariance matrix $\boldsymbol{I}$: $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I})$.

As a second alternative model based on the basic model, we included a linear time trend by appending an incremental integer array to the design matrix:

$$y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i2} \cdot X_{i3} + \beta_5 C_i + \epsilon_i = \boldsymbol{X}_i' \boldsymbol{\beta} + \epsilon_i, \quad (4)$$

where $C = (1, 2, \ldots, n_1, 1, 2, \ldots, n_2, 1, 2, \ldots, n_3, 1, 2, \ldots, n_4)^T$ is a vector of incremental integers with length n. The integers re-start from one for each of the four walking conditions. Concretely, $n = n_1 + n_2 + n_3 + n_4$, where n_1, n_2, n_3, n_4 are the number of samples under the four walking conditions, respectively. Apart from the linear time trend covariate, the model structure was identical to the basic model. It is worth noting that we estimate one coefficient, namely, β_5, for the added time trend column C. This approach assumes the outcome has the same time trend regardless of walking condition. However, the outcome time trend may in fact differ across the four walking conditions, and this coefficient represents the mean of the time trend under all of these conditions. Future models can accommodate this assumption by including interaction terms between C and each walking condition.

In further sensitivity checks, we compared models based on non-informative and informative priors for all three above-mentioned models. The investigated priors are summarized in Table 1. The distribution of informative priors was based on the corresponding gait parameter values reported for young healthy adults which included a mean stride length of 1.36 m with a standard deviation of 0.08 m, and mean stride time (estimated as doubled step time) of 1.05 s with a standard deviation of 0.06 s [35].

3. Results

3.1. Characteristics of study participants

In total, data from sixteen participants (eight males, eight females) were collected for the four walking conditions (ST-Control, ST-Fatigue, DT-Control, and DT-Fatigue). The dataset consisted of 3117 strides pooled across all participants. Stride length and stride time from each stride were used as outcome variables in the analyses. The observations were balanced across all walking conditions and participants, with 788 strides from ST-Control (49.3 ± 3.7 strides per person), 792 strides from ST-Fatigue (49.5 ± 3.7 strides per person), 766 strides from DT-Control (47.9 ± 3.2 strides per person) and 771 strides from DT-Fatigue (48.2 ± 3.5 strides per person). Table 2 summarizes the participant characteristics, and Table 3 summarizes the gait parameters.

3.2. Population-level analysis

Next, we performed baseline analyses to investigate the population-level effects of physical fatigue and cognitive task on gait. Two-way repeated measures ANOVA indicated very small effects induced by physical fatigue. The main effects of physical fatigue on stride length and stride time had a generalized eta-squared effect size of 0.01 or less (stride length: $F(1,15) = 5.86$, $p = 0.03$, $\eta^2 = 0.01$; stride time: $F(1,15) = 2.56$, $p = 0.13$, $\eta^2 = 8.5 \times 10^{-3}$). The main effects of cognitive task were moderate, with generalized eta-squared effect sizes 0.15 and 0.20 for stride length and stride time, respectively (stride length: $F(1,15) = 18.46$, $p = 6.4 \times 10^{-4}$, $\eta^2 = 0.16$; stride time: $F(1,15) = 21.14$, $p = 3.5 \times 10^{-4}$, $\eta^2 = 0.22$). No statistically significant interaction effects were found ($p = 0.77$ for stride length, and $p = 0.99$ for stride time). Table 3 summarizes the ANOVA results.

3.3. N-of-1 trials using Bayesian repeated-measures models

The posterior distributions for stride length and stride time are illustrated in Fig. 2. A complete summary of the posterior distributions of parameters can be found at https://github.com/HIAlab/gait_nof1trials/wiki/Data. Distributions of the gait parameters under the four conditions are derived from combinations of the elements in the $\boldsymbol{\beta}$ vector in the linear model, as described in Section 2.2. The results showed that the baseline values of the gait parameters (under the ST-Control condition) varied largely among all participants, and the gait changes under the four walking conditions were also highly heterogeneous among all participants. Fig. 2 also shows the aggregated posterior distributions from all participants for reference, which further indicates that the aggregated population-level summary was not a good representation of the highly heterogeneous individual gait effects.

For stride length, there was a consistent trend among participants that the values from DT conditions were smaller than those from ST conditions, as seen in the population-level ANOVA main effect of the cognitive task described above. Nevertheless, inter-person variation could be observed, for example, as illustrated in Fig. 2, study participant 10 exhibited almost no response under the DT condition compared to the ST condition (mean values changed from 1.34 under ST to 1.33 under DT), whereas participant 2 largely reduced the stride length (mean values changed from 1.62 under ST to 1.35 under DT).

In contrast, the effects of physical fatigue on stride lengths were smaller on average but more complex on the individual level compared to those induced by the cognitive task, and opposite effects could be observed for different individuals. Especially under DT condition, stride length increased from non-fatigue to fatigue condition for participant 6 (from 1.44 to 1.48), participant 14 (from 1.41 to 1.45), and participant 15 (from 1.45 to 1.48) but remained unchanged or decreased for the other participants.

Moreover, the effects of fatigue were larger under DT condition compared to under ST condition for all participants. Similar trends could be observed for stride time, where the DT condition generally induced an increase for all participants, but the individual posterior distributions were heterogeneous. The cognitive task appeared to have increased the variance as well as the effects of fatigue for many participants. It is worth noting that the posterior estimates for participants 7 and 13 had unusually large variations compared to those for all other participants. Quality control analyses revealed that the MCMC chains did not converge for these two participants. More details are presented in section MCMC Chain Convergence in Supplementary Text C.

To further quantify the marginal effects of fatigue and cognitive task on gait, the posterior probabilities of meaningful effects were estimated. The population-level analysis only showed small effects of fatigue on stride length and stride time. However, as illustrated in Fig. 3, the posterior estimates for individual participants revealed highly heterogeneous gait effects among the participants, and the posterior probability of meaningful effect on the gait parameters were

Table 1
Non-informative and informative prior distributions of the Bayesian models.

Model	Non-informative priors	Informative priors (SL)	Informative priors (ST)
Basic & Time Covariate	$\beta_{i\in\{1,2,3,4\}} \sim N(0, 10^{-3})$ $\sigma \sim U(0, 100)$	$\beta_1 \sim N(1.36, 0.08)$ $\beta_{i\in\{2,3,4\}} \sim N(0, 10^{-3})$ $\sigma \sim U(0, 100)$	$\beta_1 \sim N(1.05, 0.06)$ $\beta_{i\in\{2,3,4\}} \sim N(0, 10^{-3})$ $\sigma \sim U(0, 100)$
AR1	$\beta_{i\in\{1,2,3,4\}} \sim N(0, 10^{-3})$ $\sigma \sim half\text{-}Cauchy(2.5)$ $\phi \sim U(-1, 1)$	$\beta_1 \sim N(1.36, 0.08)$ $\beta_{i\in\{2,3,4\}} \sim N(0, 10^{-3})$ $\sigma \sim half\text{-}Cauchy(2.5)$ $\phi \sim U(-1, 1)$	$\beta_1 \sim N(1.05, 0.06)$ $\beta_{i\in\{2,3,4\}} \sim N(0, 10^{-3})$ $\sigma \sim half\text{-}Cauchy(2.5)$ $\phi \sim U(-1, 1)$

SL: stride length, ST: stride time.
The default priors are recommended by Gelman et al. [37], and the informative priors for stride lengths and stride time are based on the corresponding gait parameter values reported for young healthy adults [35].

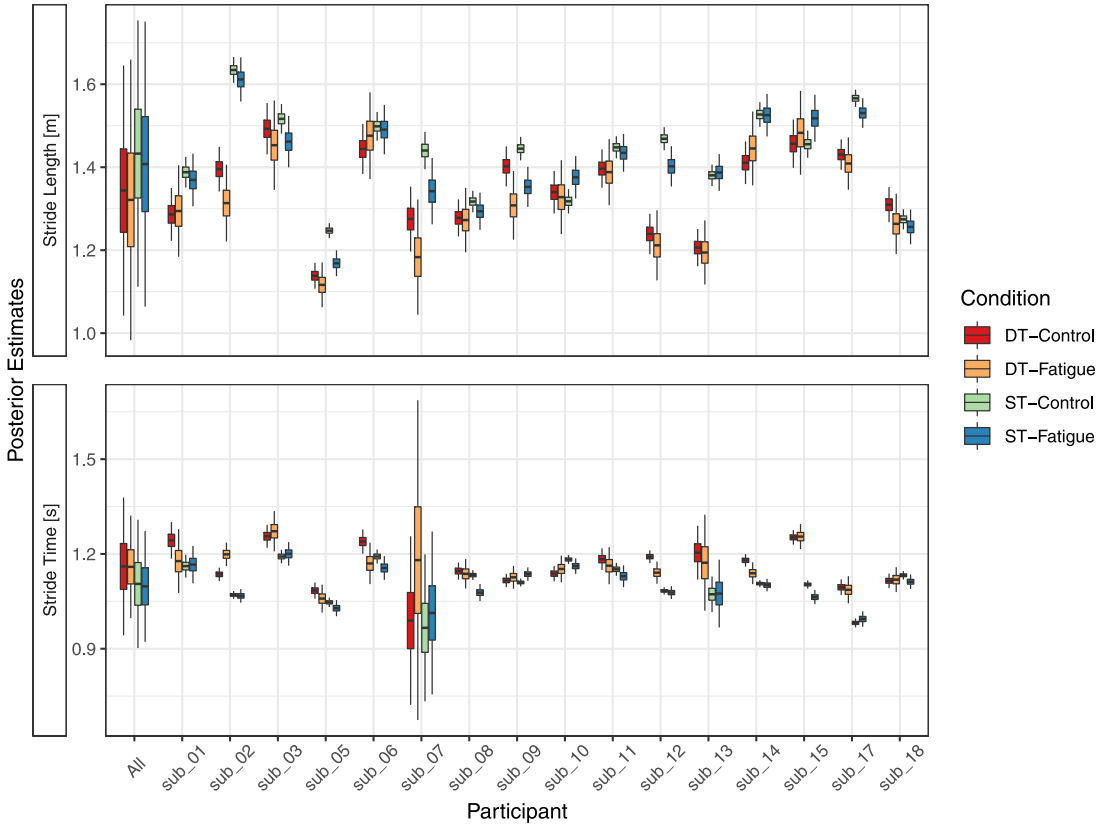


Fig. 2. Posterior estimates from the AR1 model for stride length and stride time. Heterogeneous gait changes could be observed among the participants. Top: Posteriors for stride length; bottom: posteriors for stride times. Distribution of the gait parameters under the four conditions are derived from combinations of the coefficients in the β vector in the linear model.

Table 2
Participant characteristics.

Variable	Mean ± SD	Min	Max
Age	27.1 ± 3.8	21	35
Body Mass (kg)	71.2 ± 12.2	54	103
Height (cm)	173.8 ± 8.6	158	190
Activity Level[a]	2	1	3

[a] 1, 2, 3 means low, medium, and high activity levels in IPAQ, respectively. The median is reported instead of Mean ± SD since data contain ordinal values.

considerably high for some individuals (e.g., stride lengths for participants 5, 7 and 9, stride times for participants 2, 6, 7 and 8) compared to the others. Moreover, the pattern of probability for the two gait parameters also differed from person to person. The posterior probability of meaningful effect remained consistent between the two gait parameters for some participants (e.g., participants 7, 10 and 17), while differed largely for some other participants (e.g., participants 1, 6 and 8). Similar patterns could also be observed for the effects of cognitive task.

To provide a qualitative overview of the heterogeneous gait changes under the four walking conditions for each participant, we computed the difference between each pair of conditions using mean values under the four walking conditions obtained from the posterior distributions. As illustrated in Fig. 4, for the great majority of the condition pairs, the gait changes varied in both magnitude and direction among all participants.

In the sensitivity checks, we investigated how the results might change when different regression models (AR1, basic, time covariate) or different priors (informative and non-informative) were used. Overall, the AR1 and basic model had similar posterior distributions of parameters, and the time covariate model had a slightly shifted distribution. The difference between ST-Control and DT-Control for stride length, as represented by β_2, was similar in the AR1 model, the basic model and observed data, and larger compared to the posterior estimate from the time covariate model (−0.09 from observed data, the

Table 3
Descriptive statistics and ANOVA results of the gait parameters from all participants.

	Stride length avg (m)	Stride time avg (s)
ST-Control (n=788)	1.43 ± 0.12	1.10 ± 0.07
ST-Fatigue (n=792)	1.41 ± 0.13	1.09 ± 0.06
DT-Control (n=766)	1.34 ± 0.11	1.16 ± 0.07
DT-Fatigue (n=771)	1.32 ± 0.12	1.15 ± 0.06
Main Effect Control-Fatigue	$F(1,15) = 5.86$, $p = 0.03$, $\eta^2 = 0.01$	$F(1,15) = 2.56$, $p = 0.13$, $\eta^2 = 8.5 \times 10^{-3}$
Main Effect ST-DT	$F(1,15) = 18.46$, $p = 6.4 \times 10^{-4}$, $\eta^2 = 0.16$	$F(1,15) = 21.14$, $p = 3.5 \times 10^{-4}$, $\eta^2 = 0.22$

ST = Single Task, DT = Dual Task, Avg = Average. n: total number of strides. F = F-value of ANOVA, p = p-value of ANOVA, η^2 = generalized eta-squared (effect size). Summary of gait parameters are expressed as mean ± standard deviation.

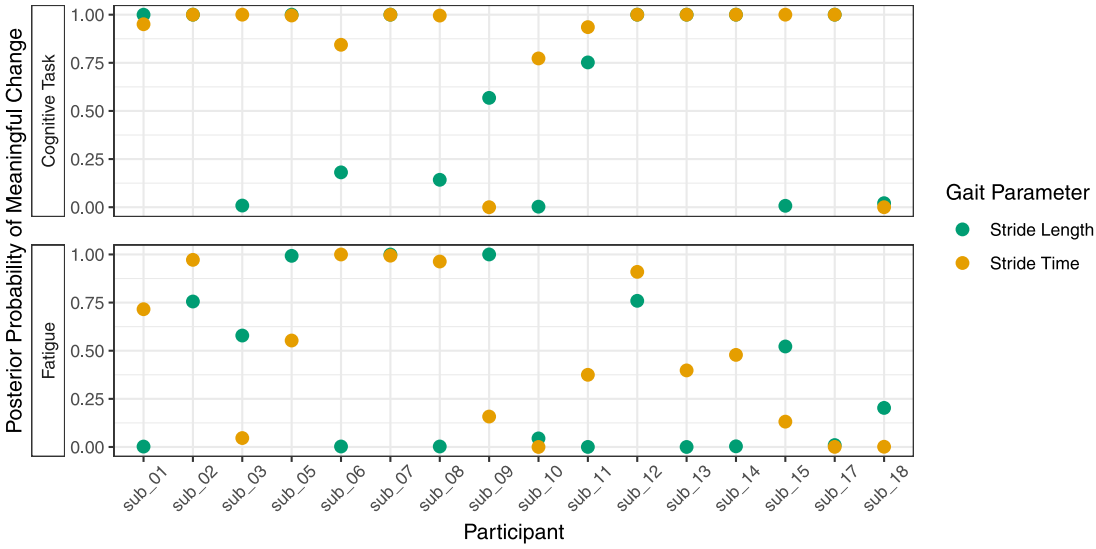


Fig. 3. Posterior probabilities of meaningful effect for each participant and gait parameter. The probabilities of meaningful effect are heterogeneous among participants and between the two gait parameters.

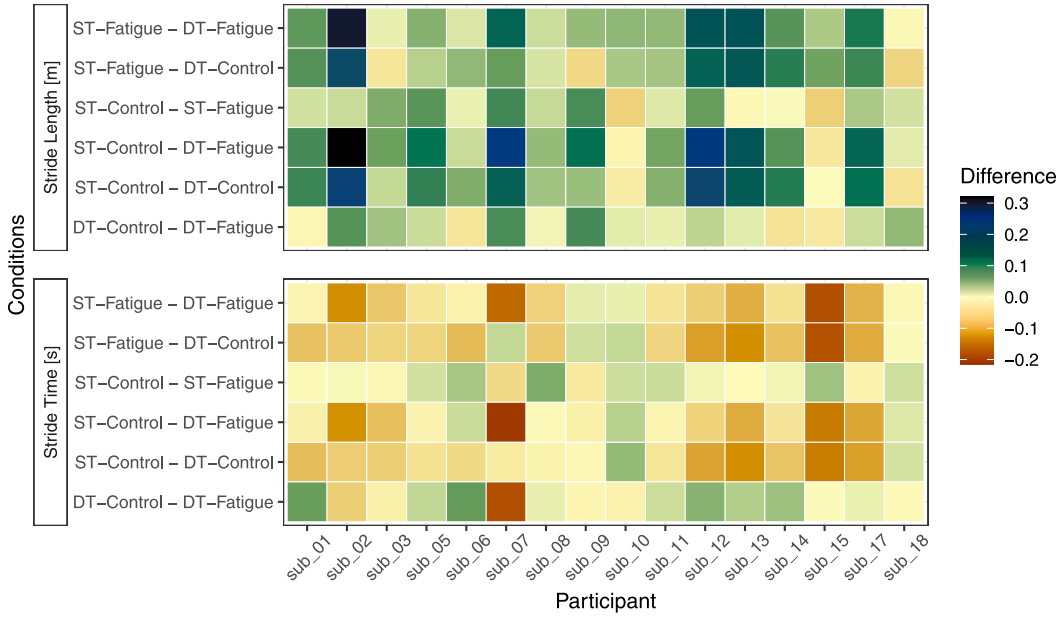


Fig. 4. Heatmap of gait parameter differences between all combinations of condition pairs based on Posterior estimates of coefficients from the AR1 model. Each row represents the difference of the gait parameters between one pair of conditions. Gait changes between condition pairs are heterogeneous among participants.

AR1 model and the basic model, −0.06 from time covariate model). The difference between ST-Control and ST-Fatigue, as represented by β_3, was negative in observed data and in the AR1 and basic model, but positive when estimated by the time covariate model (−0.03 from observed data, from the AR1 model and from the basic model, 0.03 from posterior estimate). Posterior estimates for stride times exhibited similar trends, in that the β_2 and β_3 estimates from the time covariate model were slightly different from those from the AR1 and the basic models. A comparison of models with different priors indicated that they had no meaningful influence on the sampling results. A more

detailed illustration of the effects of models and priors can be found in Supplementary Text C. Since no statistically significant differences in the posterior estimates using non-informative and informative priors were found, we focused on reporting and discussing results from models using non-informative priors.

4. Discussions

In this study, we re-analyzed data from a gait study, which was originally designed as a population-based trial, as single N-of-1 trials. Our observations from the posterior estimates demonstrate that heterogeneous responses do exist in our cohort. Especially for the effects of fatigue on gait, analyses based on N-of-1 trials are necessary to enable in-depth investigation of the individual effects that are not visible in the group-based ANOVA results. This is consistent with our prior knowledge that gait and gait responses to interventions are highly individual [26–28], and the posteriors for each individual allow further investigation into causes underlying these individual effects.

Our study introduced personalized analysis methods using example data from a healthy population. Nonetheless, spatiotemporal gait parameters such as stride length and stride time are commonly used to objectively and quantitatively measure gait performance in neurological diseases. In the example of Parkinson's disease, gait impairment is often associated with a decrease in stride length and an increased risk of falling [40]. Falls can result in injuries, fear of falling, and activity restriction, which are associated with increased institutionalization, reduced independence, and higher mortality rates. Previous studies have shown that by using various cueing strategies (e.g., visual or auditory cueing), the stride length could significantly increase up to that of control levels [41,42]. A personalized approach could enable the identification of effective interventions for an individual to improve gait performance, and thus mitigate adverse consequences of gait impairment. Moreover, our method can also be generalized to investigate individual outcomes in a variety of settings.

When comparing group-based analyses with N-of-1 trials approaches, it is worth noting that the covariate effects are treated differently. Typical group-level models statistically control for the effects of covariates related to the traits of participants, such as age, body height and mass. In contrast, the N-of-1 approach automatically adjusts for all baseline covariates, since each person always has the same baseline covariate values.

The dataset used in our study was unconventional for N-of-1 trials methods in the sense that it was obtained from a study originally designed for population-level analyses. In contrast to the multiple crossovers for typical N-of-1 trials, each participant was measured only in one session for the baseline, underwent the intervention once, and the intervention effects were measured in a subsequent session. Hundreds of data points (gait cycles) were measured under both conditions, which are sufficient for statistical analysis on a single person. Nevertheless, in order to enable the analysis of the data through the lens of N-of-1 trials, several assumptions were made with respect to time-dependent effects, carryover effects, and effects of task/treatment order.

In the following, we discuss these assumptions and the ramifications of not having additional crossovers in our study design. Regarding time-dependent effects, we assumed that the one-week break between the ST and DT visits did not induce any effect on the gait characteristics of an individual. Only in this case, the effect observed from DT condition could be attributed to cognitive task and not with time as a confounder. We based our assumption on evidence that an individual's gait characteristics are persistent over a long period of time [26]. For other types of outcomes that fluctuate over time or are more sensitive to uncontrolled factors, the effect of time between visits should not be neglected.

The order of ST and DT visits was randomized among all participants. However, with only one crossover, the order was fixed for the same person. In addition, the order of the control and fatigue conditions during both visits was fixed by the experimental design. This lack of within-person randomization may become problematic when two types of carryover effects exist: (1) carryover from the first visit to the second visit, and (2) carryover within one visit, from the first walking session to the second walking session. Regarding the first scenario, the one-week break between two visits can be considered a washout period, where the effects of fatigue exercise from the previous visit are sufficiently removed. Before the second visit, we confirmed with each participant that they did not feel any effects of the fatigue exercise. Regarding the second scenario, it is only reasonable to keep the order of first non-fatigued walk, then fatigued walk during the same visit, and there were no obvious carryover effects from the non-fatigued state to the fatigued state. When generalizing our methods to other data from population-based trials, it is important to examine all possible carryover effects, and evaluate to which extent these effects will affect the analysis outcome.

In the case of prominent carryover effects, the single-crossover design does not isolate the carryover effects from the intervention effects for the individual. As one approach, the carryover effects could be modeled in the analysis to still allow efficient and unbiased estimation of the effects [43]. In future work, additional crossovers between the fatigue and dual-task conditions can be added to the study design, in order to introduce randomization within one person.

In addition to the time-dependent effect and carryover effect, the task/treatment order may have an influence on motivation and habituation to the task. Furthermore, events that alter a participant's gait, such as injury, may occur between visits. Although no evidence of such an effect was observed in our example study, future studies should consider these factors when analyzing trials with fixed treatment orders.

In our analyses, informative and non-informative priors did not have an effect on posteriors for the same model, as illustrated in Supplementary Figure 10. One possible reason could be that the values of the non-informative priors were similar to those of the informative priors. More concretely, only prior knowledge of the mean and standard deviation for the baseline (ST-Control) was introduced, and these values (centered around 1 with very small standard deviations) were close to the non-informative priors (centered around 0 with a standard deviation of 10^{-3}). It is therefore worth emphasizing that for use cases where the informative priors differ largely from non-informative default priors, an informed choice of priors might affect the posterior estimates to a larger degree and cannot be ignored. Visualizations similar to our Supplementary Figure 10 are helpful for qualitative comparison between different priors.

Posterior estimates of the model parameters from the AR1 and basic models matched the distributions of the observed values, whereas slight deviations could be observed for posterior estimates of the time covariate model. Moreover, for four participants (#2, 9, 12, 18), the MCMC chains did not converge for the model parameter β_5 which was associated with the linear representation of time in the design matrix. In our opinion, these observations indicate that the assumption of a linear time effect with the time covariate model does not accurately represent the data. As discussed by Heckenstenden et al. [44], repeated measurements during a single uninterrupted intervention period could be used as a surrogate for repeated interventions, however, it is reasonable to assume autocorrelation between measurements, and non-linear adaptation may occur during the measurement period. Our study indicates that in such settings, the effects of time are more appropriately modeled with the temporal autocorrelation described by the AR1 model. With the assumption of stationarity, our AR1 model describes a simple pattern of correlation that declines linearly according to the time lag, but is independent of the actual time of the observation. In this regard, samples from all four walking conditions were concatenated and modeled with the same correlation structure, which yielded a good model fit. Nonetheless, it is worth noticing that the MCMC chains failed to converge for the AR1 models for two participants (#7, 13)

for stride length. During data collection, we observed that the general gait patterns of these two participants were particularly affected by the interventions, and initial data exploration revealed a large variability in gait parameters. We assume that the true data distributions of their gait parameters are different from those of the other participants, and the true relationship between the variables is non-linear. In this case, a more flexible model with a non-linear structure could be better suited for analysis.

As future work, the effect sizes of interventions can be estimated and further investigated based on the posteriors obtained from the Bayesian analyses [45]. Aggregated N-of-1 trials analyses can be performed to investigate the underlying causes of the personalized responses to intervention [43]. In our study, the different responses could potentially be associated with the participants' pre-existing health conditions, anthropometric features or stable lifestyle habits (e.g., as measured by the IPAQ questionnaire), or a combination of all these factors. In future studies, larger and more heterogeneous cohorts will provide additional features for analysis. Based on these findings, personalized advice or interventions could reduce the risk of falls or injury for vulnerable individuals.

5. Conclusion

Our study provides an example of how to initiate an in-depth investigation of treatment effects on an individual level using data from population-level studies. We demonstrate the use of Bayesian models to study individual-level effects of interventions, and point out aspects to consider for future studies.

CRediT authorship contribution statement

Lin Zhou: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Visualization, Writing – original draft, Writing – review & editing. **Juliana Schneider:** Formal analysis, Investigation, Methodology, Validation, Writing – review & editing. **Bert Arnrich:** Conceptualization, Funding acquisition, Investigation, Supervision, Writing – review & editing. **Stefan Konigorski:** Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Links to the data and code is shared in the manuscript.

Acknowledgments

The authors would like to thank all participants in this study. We would also like to thank Urs Granacher and Clemens Markus Brahms for their support in developing the study. This study received funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 491466077, and has been partly funded by the Federal Ministry of Education and Research of Germany in the framework of KI-LAB-ITSE (project number 01IS19066).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.conctc.2024.101282.

References

[1] Ioana Bica, Ahmed M. Alaa, Craig Lambert, Mihaela Van Der Schaar, From real-world patient data to individualized treatment effects using machine learning: current and future methods to address underlying challenges, Clin. Pharmacol. Ther. 109 (1) (2021) 87–100.

[2] Ahmed M. Alaa, Mihaela Van Der Schaar, Bayesian inference of individualized treatment effects using multi-task gaussian processes, Adv. Neural Inf. Process. Syst. 30 (2017).

[3] Uri Shalit, Fredrik D. Johansson, David Sontag, Estimating individual treatment effect: generalization bounds and algorithms, in: International Conference on Machine Learning, PMLR, 2017, pp. 3076–3085.

[4] Changhee Lee, Nicholas Mastronarde, Mihaela van der Schaar, Estimation of individual treatment effect in latent confounder models via adversarial learning, 2018, arXiv preprint arXiv:1811.08943.

[5] J. Nikles, G. Mitchell, The Essential Guide to N-of-1 Trials in Health, Springer, Dordrecht, 2015, pp. 135–153.

[6] Elizabeth O. Lillie, Bradley Patay, Joel Diamant, Brian Issell, Eric J. Topol, Nicholas J. Schork, The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? Pers. Med. 8 (2) (2011) 161–173.

[7] R.D. Mirza, S. Punja, S. Vohra, G. Guyatt, The history and development of N-of-1 trials, J. R. Soc. Med. 110 (8) (2017) 330–340.

[8] Syem Ishaque, Naimul Khan, Sri Krishnan, Trends in heart-rate variability signal analysis, Front. Digit. Health 3 (2021) 639444.

[9] Justin Trautmann, Lin Zhou, Clemens Markus Brahms, Can Tunca, Cem Ersoy, Urs Granacher, Bert Arnrich, Tripod—A treadmill walking dataset with imu, pressure-distribution and photoelectric data for gait analysis, Data 6 (9) (2021) 95.

[10] Fiona F. Stanaway, Danijela Gnjidic, Fiona M. Blyth, David G. Le Couteur, Vasi Naganathan, Louise Waite, Markus J. Seibel, David J. Handelsman, Philip N. Sambrook, Robert G. Cumming, How fast does the Grim Reaper walk? Receiver operating characteristics curve analysis in healthy men aged 70 and over, Bmj 343 (2011).

[11] Jeffrey M. Hausdorff, Dean A. Rios, Helen K. Edelberg, Gait variability and fall risk in community-living older adults: A 1-year prospective study, Arch. Phys. Med. Rehabil. 82 (8) (2001) 1050–1056.

[12] Inbar Hillel, Eran Gazit, Alice Nieuwboer, Laura Avanzino, Lynn Rochester, Andrea Cereatti, Ugo Della Croce, Marcel Olde Rikkert, Bastiaan R. Bloem, Elisa Pelosin, et al., Is every-day walking in older adults more analogous to dual-task walking or to usual walking? Elucidating the gaps between gait performance in the lab and during 24/7 monitoring, Eur. Rev. Aging Phys. Activity 16 (1) (2019) 1–12.

[13] Madli Bayot, Kathy Dujardin, Céline Tard, Luc Defebvre, Cédrick T. Bonnet, Etienne Allart, Arnaud Delval, The interaction between cognition and motor control: A theoretical framework for dual-task interference effects on posture, gait initiation, gait and turning, Neurophysiol. Clin. 48 (6) (2018) 361–375, XXVe congrès SOFPEL.

[14] Georg Ebersbach, Milan R. Dimitrijevic, Werner Poewe, Influence of concurrent tasks on gait: A dual-task approach, Percept. Motor Skills 81 (1) (1995) 107–113.

[15] Denisa Nohelova, Lucia Bizovska, Nicolas Vuillerme, Zdenek Svoboda, Gait variability and complexity during single and dual-task walking on different surfaces in outdoor environment, Sensors 21 (14) (2021).

[16] Erin Smith, Tara Cusack, Catherine Blake, The effect of a dual task on gait speed in community dwelling older adults: A systematic review and meta-analysis, Gait Posture 44 (2016) 250–258.

[17] Manuel Montero-Odasso, Susan W. Muir, Mark Speechley, Dual-task complexity affects gait in people with mild cognitive impairment: The interplay between gait variability, dual tasking, and risk of falls, Arch. Phys. Med. Rehabil. 93 (2) (2012) 293–299.

[18] Prakriti Parijat, Thurmon E. Lockhart, Effects of lower extremity muscle fatigue on the outcomes of slip-induced falls, Ergonomics 51 (12) (2008) 1873–1884.

[19] Paulo Cezar Rocha dos Santos, Fabio Augusto Barbieri, Inge Zijdewind, Lilian Teresa Bucken Gobbi, Claudine Lamoth, Tibor Hortobágyi, Effects of experimentally induced fatigue on healthy older adults' gait: A systematic review, PLoS One 14 (12) (2019) e0226939.

[20] Anna L. Hatton, Jasmine C. Menant, Stephen R. Lord, Joanne C.M. Lo, Daina L. Sturnieks, The effect of lower limb muscle fatigue on obstacle negotiation during walking in older adults, Gait Posture 37 (4) (2013) 506–510.

[21] Urs Granacher, Irene Wolf, Anja Wehrle, Stephanie Bridenbaugh, Reto W. Kressig, Effects of muscle fatigue on gait characteristics under single and dual-task conditions in young and older adults, J. Neuroeng. Rehabil. 7 (1) (2010) 1–12.

[22] Fabio Augusto Barbieri, Paulo Cezar Rocha dos Santos, Lucas Simieli, Diego Orcioli-Silva, Jaap H. van Dieën, Lilian Teresa Bucken Gobbi, Interactions of age and leg muscle fatigue on unobstructed walking and obstacle crossing, Gait Posture 39 (3) (2014) 985–990.

[23] Steven Morrison, Sheri R. Colberg, Henri K. Parson, Serina Neumann, Richard Handel, Etta J. Vinik, James Paulson, Arthur I. Vinik, Walking-induced fatigue leads to increased falls risk in older adults, J. Am. Med. Dir. Assoc. 17 (5) (2016) 402–409.

[24] Marcel J.P. Toebes, Marco J.M. Hoozemans, Joost Dekker, Jaap H. van Dieën, Effects of unilateral leg muscle fatigue on balance control in perturbed and unperturbed gait in healthy elderly, Gait Posture 40 (1) (2014) 215–219.
[25] Jorunn L. Helbostad, Sara Leirfall, Rolf Moe-Nilssen, Olav Sletvold, Physical fatigue affects gait characteristics in older persons, J. Gerontol A 62 (9) (2007) 1010–1015.
[26] F. Horst, M. Mildner, W.I. Schöllhorn, One-year persistence of individual gait patterns identified in a follow-up study – A call for individualised diagnose and therapy, Gait Posture 58 (2017) 476–480.
[27] Zoe Y.S. Chan, Janet H. Zhang, Reed Ferber, Gary Shum, Roy T.H. Cheung, The effects of midfoot strike gait retraining on impact loading and joint stiffness, Phys. Ther. Sport 42 (2020) 139–145.
[28] Franz Marxreiter, Heiko Gaßner, Olga Borozdina, Jens Barth, Zacharias Kohl, Johannes Schlachetzki, Caroline Thun-Hohenstein, Dieter Volc, Bjoern M. Eskofier, Jürgen Winkler, et al., Sensor-based gait analysis of individualized improvement during apomorphine titration in Parkinson's disease, J. Neurol. 265 (11) (2018) 2656–2665.
[29] Jorik Nonnekes, Alice Nieuwboer, Towards personalized rehabilitation for gait impairments in Parkinson's disease, J. Park. Dis. 8 (s1) (2018) S101–S106.
[30] Clare C. Maguire, Judith M. Sieben, Nathanael Lutz, Gisela van der Wijden, Heike Scheidhauer, RobertA de Bie, Replacing canes with an elasticated orthotic-garment in chronic stroke patients–The influence on gait and balance. A series of N-of-1 trials, J. Bodyw. Mov. Ther. 24 (4) (2020) 203–214.
[31] Lin Zhou, Eric Fischer, Clemens Markus Brahms, Urs Granacher, Bert Arnrich, Duo-gait: A gait dataset for walking under dual-task and fatigue conditions with inertial measurement units, Sci. Data 10 (1) (2023) 543.
[32] Cora L. Craig, Alison L. Marshall, Michael Sjöström, Adrian E. Bauman, Michael L. Booth, Barbara E. Ainsworth, Michael Pratt, U.L.F. Ekelund, Agneta Yngve, James F. Sallis, et al., International physical activity questionnaire: 12-country reliability and validity, Med. Sci. Sports Exerc. 35 (8) (2003) 1381–1395.
[33] Dominique Makowski, Mattan S. Ben-Shachar, S.H. Annabel Chen, Daniel Lüdecke, Indices of effect existence and significance in the Bayesian framework, Front. Psychol. 10 (2019) 2767.
[34] Rivka M. de Vries, Richard D. Morey, Bayesian hypothesis testing for single-subject designs, Psychol. Methods 18 (2) (2013) 165.
[35] Antonio Gomez Bernal, Ricardo Becerro-de Bengoa-Vallejo, Marta Elena Losa-Iglesias, Reliability of the OptoGait portable photoelectric cell system for the quantification of spatial-temporal parameters of gait in young adults, Gait Posture 50 (2016) 196–200.
[36] Andrew Gelman, Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper), Bayesian Anal. 1 (3) (2006) 515–534.
[37] Andrew Gelman, Aleks Jakulin, Maria Grazia Pittau, Yu-Sung Su, A weakly informative default prior distribution for logistic and other regression models, Ann. Appl. Stat. 2 (4) (2008) 1360–1383.
[38] Bas C. Stunnenberg, Joost Raaphorst, Hans M. Groenewoud, Jeffrey M. Statland, Robert C. Griggs, Willem Woertman, Dick F. Stegeman, Janneke Timmermans, Jaya Trivedi, Emma Matthews, et al., Effect of mexiletine on muscle stiffness in patients with nondystrophic myotonia evaluated using aggregated N-of-1 trials, JAMA 320 (22) (2018) 2344–2353.
[39] Luca Martino, Víctor Elvira, Francisco Louzada, Effective sample size for importance sampling based on discrepancy measures, Signal Process. 131 (2017) 386–401.
[40] Nicolas Bayle, Amar S. Patel, Diana Crisan, Lanjun J. Guo, Emilie Hutin, Donald J. Weisz, Steven T. Moore, Jean-Michel Gracies, Contribution of step length to increase walking and turning speed as a marker of Parkinson's disease progression, PLoS One 11 (4) (2016) e0152469.
[41] Cornelia Schlick, Alina Ernst, Kai Bötzel, Annika Plate, Olena Pelykh, Josef Ilmberger, Visual cues combined with treadmill training to improve gait performance in Parkinson's disease: a pilot randomized controlled trial, Clin. Rehabil. 30 (5) (2016) 463–471.
[42] Massimiliano Pau, Federica Corona, Roberta Pili, Carlo Casula, Fabrizio Sors, Tiziano Agostini, Giovanni Cossu, Marco Guicciardi, Mauro Murgia, Effects of physical rehabilitation integrated with rhythmic auditory stimulation on spatio-temporal and kinematic parameters of gait in Parkinson's disease, Front. Neurol. 7 (2016) 126.
[43] Thomas Gärtner, Juliana Schneider, Bert Arnrich, Stefan Konigorski, Comparison of bayesian networks, g-estimation and linear models to estimate causal treatment effects in aggregated n-of-1 trials with carry-over effects, BMC Medical Research Methodology 23 (1) (2023) 191.
[44] Anne Hecksteden, Jochen Kraushaar, Friederike Scharhag-Rosenberger, Daniel Theisen, Stephen Senn, Tim Meyer, Individual response to exercise training-a statistical perspective, J. Appl. Physiol. 118 (12) (2015) 1450–1459.
[45] Riko Kelter, Analysis of Bayesian posterior significance and effect size indices for the two-sample t-test to support reproducible medical research, BMC Med. Res. Methodol. 20 (1) (2020) 1–18.