Statistical Analysis Plan (SAP) for "Predictors of physical activity levels in children and adolescents with cerebral palsy: clinical cohort study protocol"

1. Administrative information

Title	**SAP - Predictors of physical activity levels in children and adolescents with cerebral palsy: clinical cohort study protocol**
Protocol article	doi: 10.1136/bmjopen-2020-047522
Trial registration number	ClinicalTrials.gov number NCT04614207
SAP authors	Christina Esmann Fonvig[12], Jens Troelsen[3], Sören Möller[2], Anders Holsgaard Larsen[12] [1]Department of Orthopaedics and Traumatology, Odense University Hospital, Odense, Denmark [2]Department of Clinical Research, University of Southern Denmark, Odense, Denmark [3]Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark
SAP version	002
Date of SAP version	2022-06-02
Date of approval	2022-06-02

Alterations to the SAP

See 'Protocol Deviations' on page 5.

Table of contents

2. Introduction

Background and rational

Children and adolescents with cerebral palsy may be trapped in a vicious circle of low physical fitness, resulting in deconditioning that causes a further decrease in physical activity (PA), a lower quality of life and an increased risk of developing non-communicable diseases. Therefore, establishing a healthy and active lifestyle during childhood is even more important for individuals with a disability. However, the factors that influence habitual PA in children and adolescents with cerebral palsy remain unknown.

In Scandinavia, healthcare professionals offer standardized clinical examinations throughout childhood using the Cerebral Palsy Follow-Up Program (CPUP). Identifying objective predictors of PA from this national clinical quality database, as well as from proxy-reported questionnaires concerning quality of life, functional status, and personal factors by means of a conceptual statistical framework (e.g. the International Classification of Functioning, Disability and Health (ICF)) may be used for an elaborated knowledge on habitual PA in children with CP. Specifically, the results will allow for early detection and potential improved interventions.

Objectives/hypothesis

The objective of this study is to identify and investigate potential predictors of habitual physical activity in children and adolescents with cerebral palsy. We hypothesize that potential predictors of habitual physical activity can be identified through the CPUP Program database, as well as in proxy-reported questionnaires, using the ICF as a conceptual statistical framework.

3. Methods

A protocol article for the study has been published (1).

Study design

The present study is a prospective clinical cohort study. Variables from historical registry data (from the Cerebral Palsy Follow-Up Program (CPUP)), along with proxy-reported questionnaires, will be used as predictive variables. Accelerometry data on physical activity will be used as outcome. For more details on the included variables, see the study protocol article (1).

Study size

The expected study size of 300-400 children, as mentioned in the protocol article, was obtained for consent given to participate. However, due to the COVID-19 pandemic, and consequences thereof, accelerometer data was only obtained for 123 children. Furthermore, due to the incompleteness of registered children in the CPUP database, only 78 of these children are eligible for analysis with CPUP data.

Timeline for the analysis

All analyses are to be performed after the SAP has been published on ClinicalTrials.gov (NCT04614207).

4. Statistical principals

Confidence intervals and p-values

The significance level for statistically significant differences will be set at an $\alpha = 0.05$ level (two-tailed testing) and will be presented as means and 95% confidence interval unless stated otherwise.

Protocol deviations

Due to the COVID-19 pandemic and related shutdown of schools and after school activities, data collection was hindered and thus resulted in a reduced number of expected participants. Thus, deviations to protocol (as published (1), and initially registered at ClinicalTrials.gov November 3, 2020) are as follows:

- Data collection ended January 31st 2022
- It is not a requirement for the children to be registered in the CPUP
- Primary analysis: Due to a relatively large amount of included children not being registered in the CPUP, additional analyses will be performed on all children with accelerometry data (Model 1b):
    - Model 1
        a) Multiple linear regression analysis between accelerometer counts (response variable) and all CPUP variables within each ICF component (predictive variables), using the latest CPUP registration prior to accelerometer data collection.
        b) Multiple linear regression analysis between accelerometer counts (response variable) and all questionnaire variables within each ICF component (predictive variables), using the questionnaire data.
- Secondary analysis. Due to the limited amount of children and adolescents included in the study, the protocol's model 2a (backward stepwise regression with accelerometer counts as the response variable and all included CPUP variables as predictive variables) will be performed for both model 1a and 1b using penalized regression. Furthermore, the initial model 2b (Multiple linear regression analysis between accelerometer counts (response variable) and all included variables as predictive variables) will be eliminated:
    - Model 2
        a) Penalized regression (lasso) with accelerometer counts as the response variable and covariates as chosen by lasso to obtain an optimal fit from model 1a as predictive variables to determine which variables to retain in the model.
        b) Penalized regression (lasso) with accelerometer counts as the response variable and covariates as chosen by lasso to obtain an optimal fit from model 1b as predictive variables to determine which variables to retain in the model.
- The expected study size of 300-400 children was obtained for consent given to participate. However, due to the COVID-19 pandemic, and consequences thereof, accelerometry data was only obtained for 123 children. Furthermore, due to the incompleteness of registered children in the CPUP database, only 78 of these children are eligible for analysis with CPUP data.
- Due to the reduced number of expected participants, the model will not be testing the significance of CPUP data collection periods relative to the time in months from the measure of PA, for differences in prediction analyses between the following time periods: 0≤12 months, 13≤24 months and 25+ months

Definition of the study population to be analyzed

Analysis will be performed on children and adolescents that have all completed the a-priori defined minimum number of valid wear days of the accelerometer (four, including one weekend day) (1).

5. Study population

Screening data

Screening data to describe representativeness of the study population will be presented (age, gender, region of residence, CP type and subtype, GMFCS level) for children and adolescents participating in the study as well as for all additional children and adolescents registered in the CPUP database.

Inclusion criteria

- Children and adolescents diagnosed with cerebral palsy,
- at a Gross Motor Function Classification System level I-III (independent gait function with or without assistive devices),
- born in the period 2003-2013,
- alive and living in Denmark.

Recruitment

Social security number (CPR-number) and names of parents of a child diagnosed with cerebral palsy, retrieved from the Danish National Patient Register, was used to invite potential families to participate via secure digital mail (e-Boks) on November 3, 2020.

Information on recruitment will be included in the flow diagram (Figure 1).

Withdrawal/exclusion

Potential participants (parents) received an invitation to participate via secure electronic mail (e-Boks). For those who did not respond to the invitation, a reminder to accept/decline consent to participate in the study was sent an additional two times. Participants who had given consent to participate, but had not stated an address to which the accelerometer could be sent, received a mail stating the importance of this information and requesting the address be stated. Families who returned an accelerometer and stated that they did not use it, were contacted via mail to ask whether they could be interested in receiving a new accelerometer.

Reasons for exclusion will be presented in the manuscript as well as Figure 1.

Participant characteristics

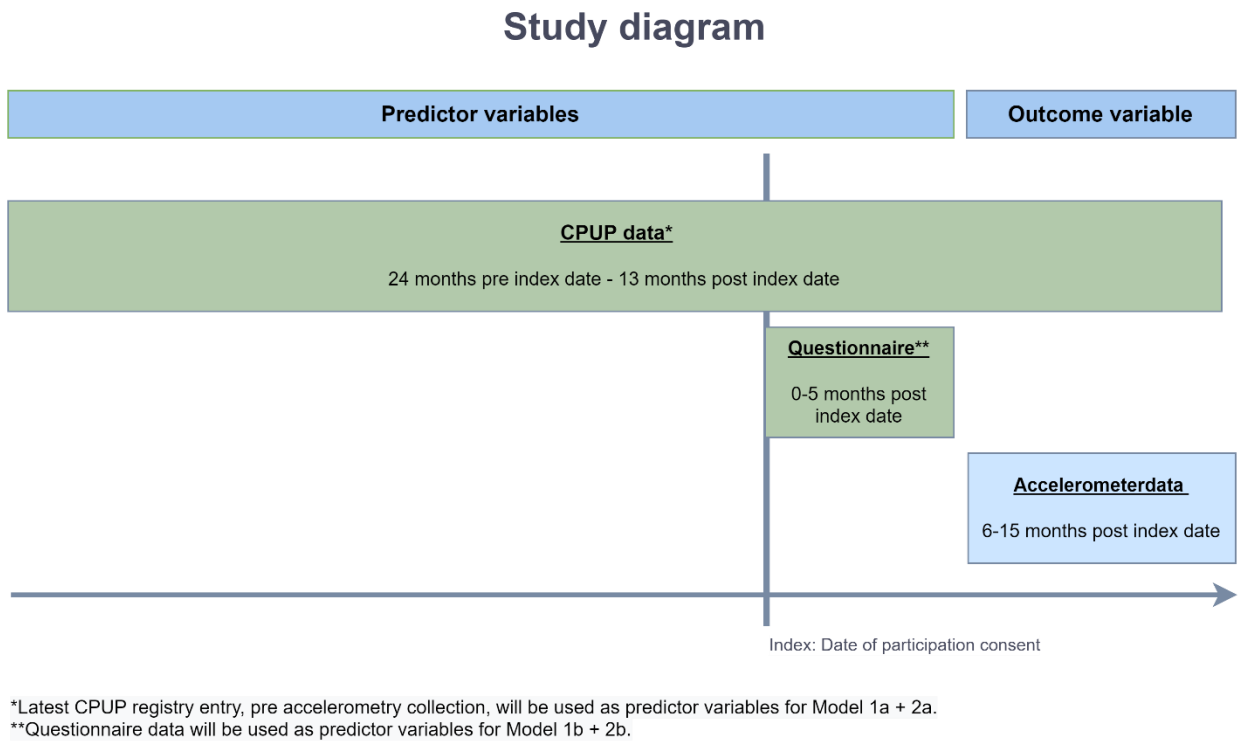
Characteristics of participants will be outlined in Table 1.

6. Variable description

Outcome

Daily average accelerometer counts will be used as outcome variable. The child/adolescent is encouraged to wear the belt with the Axivity AX3 accelerometer for seven consecutive days, but must have worn the accelerometer for a minimum of four days, including one weekend day, to be included in this study. The variable, daily average accelerometer counts, is weighted 5/7 and 2/7 for weekdays and weekend days, respectively.

Accelerometry data is collected 6-15 months post index date (see study diagram below).



*Latest CPUP registry entry, pre accelerometry collection, will be used as predictor variables for Model 1a + 2a.
**Questionnaire data will be used as predictor variables for Model 1b + 2b.

Exposure and adjustment variables

Questionnaire variables and CPUP variables within each ICF component will be used as predictors for predicting the outcome. See the protocol article for a list of included predictors (1).

Gender, region of residence, GMFCS level, and cerebral palsy subtype will be used as covariates and will be extracted from the CPUP database. Age will be calculated as age in years at index date. For non-participants, index date will be set at November 3, 2020.

For children and adolescents without CPUP data, gender, region of residence and GMFCS level will be extracted from parent-reported questionnaires. Age will be calculated as age in years at index date.

7. Analysis

Method of analysis

Descriptive statistics – participant characteristics

Gaussian distribution of numerical variables will be inspected by normal quantile plots (qq-plots). If variables are normally distributed, these will be presented as means with standard deviations (SD) and compared by two-sample t-tests. If variables are not normally distributed, they will be presented as medians and quartiles and compared by non-parametric equality-of-medians test.

Categorical variables will be presented as counts and proportions and compared with chi-squared test (or Fisher's exact test in case of counts below 5).

We will perform two descriptive comparisons: Participants with CPUP data versus participants without CPUP data, as well as, participants with CPUP data versus non-participants included in CPUP.

Regression models

Models will be checked using graphic inspection of diagnostic plots. Splines will be used to account for non-linear effects, and interactions will be included in the model based on relevant subject-matter knowledge. The coefficient of determination, adjusted R-squared, will illustrate the percentage of variance in PA that is explained by the predictive variables. The higher the coefficient, the stronger the relationship. The root mean squared error of the estimate will indicate the accuracy of the predictions.

The following models will be applied:

- Model 1
    a) Multiple linear regression analysis between accelerometer counts (response variable) and all CPUP variables within each ICF component (predictive variables), using the latest CPUP registration prior to accelerometer data collection.
    b) Multiple linear regression analysis between accelerometer counts (response variable) and all questionnaire variables within each ICF component (predictive variables), using the questionnaire data.
- Model 2
    a) Penalized regression (lasso) with accelerometer counts as the response variable and covariates as chosen by lasso to obtain an optimal fit from model 1a as predictive variables to determine which variables to retain in the model.
    b) Penalized regression (lasso) with accelerometer counts as the response variable and covariates as chosen by lasso to obtain an optimal fit from model 1b as predictive variables to determine which variables to retain in the model.

Missing data

Missing outcome data (accelerometer data) will exclude the child from the study population. Missing data on predictors (CPUP and questionnaire data, including personal factors) will be handled by multiple imputation by multivariate imputation using chained equation (MICE) including all predictors and 100 imputation samples.

In model 2a and 2b the imputation will be performed before applying the lasso, and afterwards the predictors will be ordered by proportion of imputation sets in which they are included. From this prioritized list, we will construct a final prediction model including all predictors included in at least 50% of the

imputation sets, but other cutoffs than 50% will be evaluated by an elbow plot, and if a different cutoff appears convincing from the plot, this cutoff will be applied.

Adverse events

Measuring habitual PA by accelerometry is a non-invasive method commonly used in research and has no known risks or side effects. However, should there be any reports of adverse events then these will be reported.

Statistical software

All statistical analyses will be performed using Stata/IC 17 or later for Windows (Stata-Corp LLC, College Station, TX, USA). Random seeds will be specified for imputation.

8. Supplementary information

SAP tables and figures

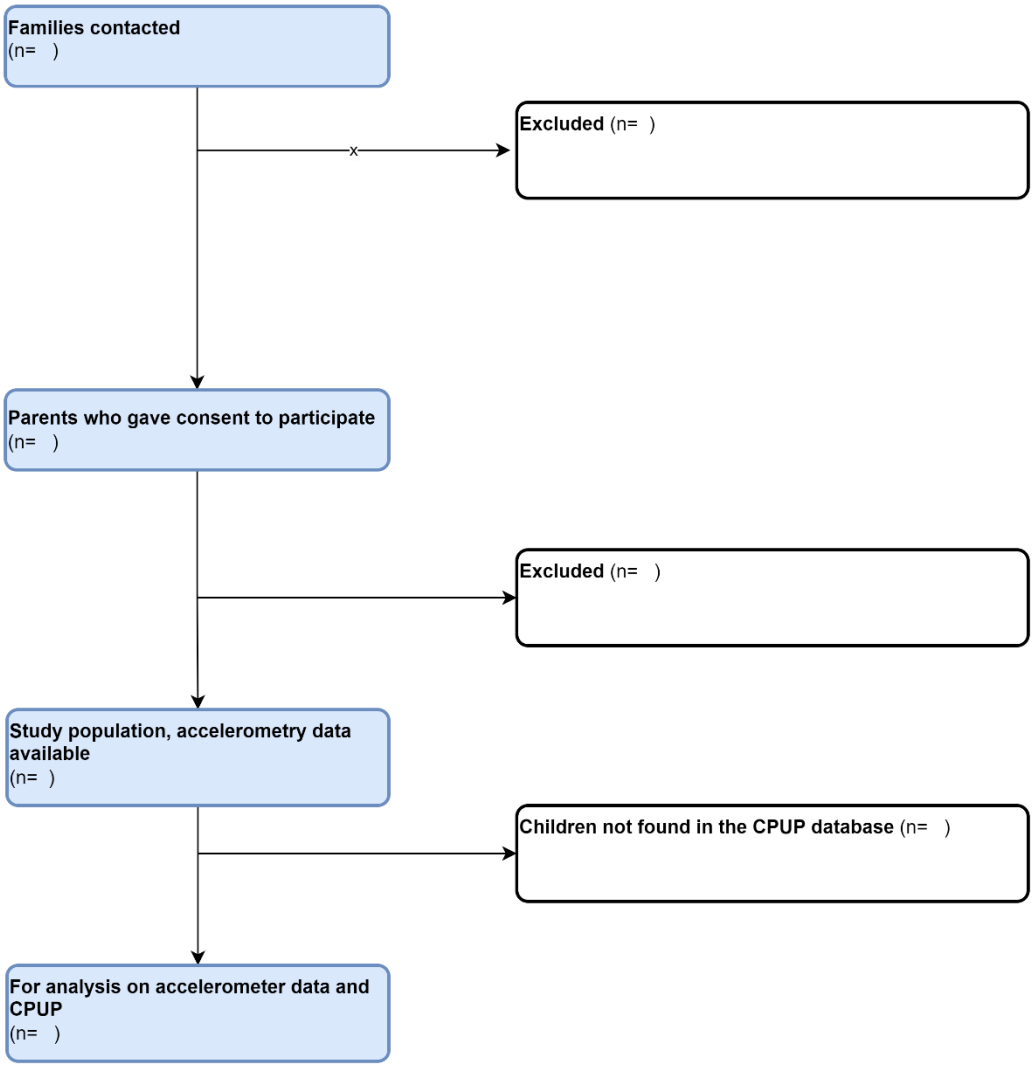


Figure 1. Flowchart of participants

Table 1. Participant characteristics

	Participants with CPUP data (n=)	Participants without CPUP data (n=)	Non-participants from CPUP (n=)
Age in years *(mean (SD))*			
At index date	xx (xx)	xx (xx)	xx (xx)*
Time in months from index date (mean (SD))			
To collection of questionnaire data	xx (xx)	xx (xx)	-
To collection of accelerometer data	xx (xx)	xx (xx)	-
To collection of CPUP data	xx (xx)	-	xx (xx)
Gender *(n (%))*			
Boy	xx (%)	xx (%)	xx (%)
Girl	xx (%)	xx (%)	xx (%)
Region of residence *(n (%))*			
Capital Region of Denmark	xx (%)	xx (%)	xx (%)
Central Denmark Region	xx (%)	xx (%)	xx (%)
Region of Northern Denmark	xx (%)	xx (%)	xx (%)
Region Zealand	xx (%)	xx (%)	xx (%)
Region of Southern Denmark	xx (%)	xx (%)	xx (%)
Cerebral palsy subtype *(n (%))*			
Spastic	xx (%)	-	xx (%)
Ataxic	xx (%)	-	xx (%)
Dyskinesia	xx (%)	-	xx (%)
Mixed/not classified	xx (%)	-	xx (%)
Missing data	xx (%)	-	xx (%)
GMFCS level *(n (%))*			
I	xx (%)	xx (%)**	xx (%)
II	xx (%)	*xx (%)***	*xx (%)*
III	xx (%)	*xx (%)***	*xx (%)*
Accelerometer count (mean (SD))			
	xx (xx)	xx (xx)	-

Abbreviations: GMFCS: Gross Motor Function Classification System, CPUP: Cerebral Palsy Follow-Up Program
Index date= date on which consent was given for participation, *= the time of mass release of invitation to the study (November 3, 2020), **=as reported by parents

Table 2. Prediction of physical activity using CPUP data predictive variables (Model 1a)

ICF constructs	Variables	Coef. [95% CI]	p-value	Adj.R^2
Body Function & Structure				xx
	ROM	xx [xx,xx]	xx	
	Pain	xx [xx,xx]	xx	
	Muscle tone (modified Ashworth Scale)	xx [xx,xx]	xx	
	GMFCS level	xx [xx,xx]	xx	
Activities				xx
	FMS	xx [xx,xx]	xx	
	GMFM-66 score	xx [xx,xx]	xx	
	Ability to climb stairs	xx [xx,xx]	xx	
	Bikes	xx [xx,xx]	xx	
Participation				xx
	Participation in PT at school	xx [xx,xx]	xx	
	Participation in recreational activities	xx [xx,xx]	xx	
Personal Factors				xx
	Age	xx [xx,xx]	xx	
	Gender	xx [xx,xx]	xx	
	CP classification	xx [xx,xx]	xx	
Environmental Factors				xx
	Residence region	xx [xx,xx]	xx	
	Use of orthosis	xx [xx,xx]	xx	
	Use of wheelchair	xx [xx,xx]	xx	

Abbreviations: ROM: Range of motion; FMS: Functional Mobility Scale, PODCI: Pediatric Outcomes Data Collection Instrument; PT: physical training;

Table 3. Prediction of physical activity using questionnaires as predictive variables (Model 1b)

ICF constructs	Variables	Coef. [95% CI]	p-value	Adj.R^2
Body Function & Structure				xx
	ROM	xx [xx,xx]	xx	
	Number of hours of sleep per night	xx [xx,xx]	xx	
	GMFCS level	xx [xx,xx]	xx	
	PODCI – pain/comfort scale	xx [xx,xx]	xx	
Activities				xx
	FMS (5m, 50m 500m)	xx [xx,xx]	xx	
	Means of transport to school	xx [xx,xx]	xx	
	Hours of screen time per day	xx [xx,xx]	xx	
Participation				xx
	Participation in PT at school	xx [xx,xx]	xx	
	Participation in recreational activities	xx [xx,xx]	xx	
	PODCI - Global Function score	xx [xx,xx]	xx	
	PODCI - Sports and Physical Functioning score			
Personal Factors				xx
	Age	xx [xx,xx]	xx	
	Gender	xx [xx,xx]	xx	
	Parent's educational level	xx [xx,xx]	xx	
Environmental Factors				xx
	Residence region	xx [xx,xx]	xx	
Quality of Life				xx
	PedsQL	xx [xx,xx]	xx	

Abbreviations: ROM: Range of motion; FMS: Functional Mobility Scale, PODCI: Pediatric Outcomes Data Collection Instrument; PT: physical training; PedsQL: Pediatric Quality of Life Inventory, CP Module

Table 4. Prediction model using "the best" CPUP variables (penalized regression, lasso) (model 2a)

(Like table 2, only including the variables that the lasso regression has selected)

Table 5. Prediction model using "the best" questionnaire variables (penalized regression, lasso) (model 2b)

(Like table 3, only including the variables that the lasso regression has selected)

References

1. Fonvig CE, Troelsen J, Dunkhase-Heinl U, Lauritsen JM, Holsgaard-Larsen A. Predictors of physical activity levels in children and adolescents with cerebral palsy: clinical cohort study protocol. BMJ Open. 2021;11(9):e047522.