Data management and statistical analysis plan

The Short Term Efficacy of Reducing Screen-based Media Use on Physical Activity, Sleep, and Physiological Stress and Well-Being in Families With Children - A Randomized Controlled Trial

Clinicaltrials.gov ID: NCT04098913

Date: February 21, 2020

Authors

Martin Gillies Banke Rasmussen*[1], Jesper Pedersen[1], Line Grønholt Olesen[1], Søren Brage[1,2], Heidi Klakk[1,3], Peter Lund Kristensen[1], Jan Christian Brønd[1] and Anders Grøntved[1]

[1] Research Unit for Exercise Epidemiology, Centre of Research in Childhood Health, Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, 5230 Odense, Denmark

[2] MRC Epidemiology Unit, Cambridge School of Clinical Medicine, Institute of Metabolic Science, University of Cambridge, Box 285, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK.

[3] Department of Physiotherapy and Research Center for Health Science, University College Lillebælt, Odense, Denmark

*Corresponding author: mgrasmussen@health.sdu.dk

Contents

Data management

Data flow and data storage

Outcome data from physical measurements will be collected by subjects themselves during everyday life activities. When measurements are completed the data is stored on the devices and the devices must be kept in the household. Shortly after data has been collected, during scheduled meetings, research staff will transport the stored data back to the Department of Sports Science and Clinical Biomechanics at the University of Southern Denmark. At the university the equipment is stored only shortly, in a locked safe, behind two doors. The former door can only be opened using a staff key and the latter electronically only by staff card. Student helpers will extract the data from the devices and store it on a PF-folder, on a University server. A PF folder is a type of folder which meets the requirements of the General Data Protection Regulation (including logging of user folder activity) for storage of sensitive data. After the data has been extracted and stored on the servers, it will be deleted from the devices, and the equipment will be ready for re-use.

Television measurement devices will be installed on televisions in the household to quantify tv-usage. Flow of data from these instruments is identical to that of the physical outcome measurements described above. We will also monitor computer and portable device (smartphone and tablets) activity using SDU-device tracker. SDU device tracker is non-commercial software which we have developed solely for research purposes. Data is continuously transmitted directly to university servers via the device's online connection. It will then be exported from this server to the PF-folder, when we are ready for data analysis.

Adult subjects will collect saliva samples. After each sample is taken the subject will store it in the household freezer. At the final meeting, the samples will be transported in a freezer box to a freezer at the Department of Sports Science and Clinical Biomechanics at University of Southern Denmark. When enough samples have been transported to the Department (no more than 500 samples), these will be transported in a freezer box from the University to the Clinical Biochemical Department at Slagelse Hospital (~ 70 km drive) for analysis for cortisol and cortisone content. The research staff will have continued dialogue with biochemists at the hospital to assure that protocols for correct storage and analysis are met. Following laboratory analysis, the samples will be destroyed. The data from the samples will be sent encrypted to MGR, who will store the data on the trial PF-folder.

We will create a folder where we will store all raw data from physical measurements. A separate folder will be created, where a copy of the original raw data (the back-up) will be saved. This new file will in its filename include a "_copy" suffix. Only the copied version will be used for data handling and analysis.

During the study, several diaries will be filled out by hand by the subjects. These diaries will be transported back to the Department upon completion of the trial. Here, student helpers will via single-entry write the dairy content into an excel sheet, which is saved on the PF-folder.

Survey data will be collected and stored using REDCap (a secure application for building and managing online surveys and databases), which is managed by the researcher service organization Odense Patient Data Exploratory Network (OPEN) in the Region of Southern Denmark. This mode of storage is in accordance with the General Data Protection Regulation (GDPR) for data handling. When the survey data is to be analyzed, it is exported from this online system in Stata format directly to the PF-folder.

All data will be stored in its raw form until completion of the study. No data handling or analyses will be conducted before completion of data collection. Therefore, researchers will have no knowledge of the results of any analyses, until after the last subjects have completed the trial. The last subject is expected to have completed the trial in December of 2020.

Data will be deleted no later than 01-01-2023. An application may be filed to the Danish Data Protection Agency to extend the date of deletion, if necessary.

Coding

MGR, JP, PLK, JCB and AG will be responsible for data handling, including coding, of all data (physical activity monitoring, sleep measurements, heart rate variability, cortisol and cortisone from saliva samples, diaries as well as data from the tv, computer and portable devices). Survey data from REDCap will be exported to Stata, where much coding is already included, based on how the survey was built in REDCap. Data handling will take place using Stata IC 16 (Statacorp) and R version 3.6.2 for heart rate variability analysis (RHRV software package version 4.2.5).

Subject identification and confidentiality

A list of randomly selected adults from households with at least one child between 6-10 years of age on Fune is gathered from the Danish Civil Registry. The list is stored on the trial PF-folder. Each adult will receive a survey addressed to him or herself, as well as one randomly selected child between 6-10 years, in the household. Each unit of adult and child is designated a number. From this stage and onwards, pseudo-anonymity is established, based on this number. Based on the survey, subjects may be recruited for a trial. Those recruited will, in most cases, be included along with the remaining members of the household. The unit identifier will follow the subjects into the trial. The adult randomly selected for the survey will for the trial be identified by an "01" suffix in addition to the original identifier. Remaining adults in the household will be designated a "02" and "03" suffix, respectively. Children will be identified using an "04", "05" or "06" suffix, respectively. In addition, adults will also be designated a separate number between 1 and the number of adult participants (currently unknown, estimated 186) who will collect saliva samples. This number is used by laboratory staff who will analyze the saliva samples for concentration of cortisol and cortisone. When the hormone data is gathered from the saliva samples, we will cross-reference the saliva id with the personal identifier described above.

For analysis data from different sources (physical measurement, screen media use data and diaries) will be merged based on the original and individual identifier. Therefore, data will be analyzed using pseudo-anonymity. However, MGR, JP, LGO, PLK, JCB, AG and a pre-graduate student and two student helpers have been granted access to the data (on the PF-folder) by the Danish Data Protection Agency and will be able to identify the identity of the subjects. No individuals beyond those listed above have been granted access to the personal data. Therefore, no one beyond the researchers involved, will be able to identify participants of the trial. Participation in the study will therefore be confidential.

Statistical analysis

Descriptive statistics

Descriptive statistics of covariates for the two groups and for all participants will be calculated for baseline data. For interval/ratio scaled covariates, which follow a normal distribution, a mean and standard deviation will be calculated. For skewed interval/ratio scaled data, a median and interquartile range will be calculated. Other covariate data types will be presented as proportions.

Time annotation of exposure and outcome data

At baseline (week one) and follow-up (week three, second week of intervention or control period) participants must note in daily checklist time of awakening, arrival and when leaving work, as well as bedtime. Based on this information we will time annotate the time series data (accelerometry and heart rate variability measurements), as well as screen media use data, into leisure and school or work periods, as well as awake versus sleep time. If schedule information is missing, we will impute based on the available data. For time of awakening, data from the closest day will be imputed (done separately for weekdays and weekend days). Missing data on time of arrival to and time of leaving from work will be imputed based on data from the closest day with non-missing data. Lastly, imputation of sleep time will be performed in three steps; first, imputation of sleep time on days with missing data from Monday to Thursday, will be performed using the closest day with non-missing data, between Monday and Thursday. For Fridays, sleep time from a Friday with non-missing data, will be imputed. For weekends, the closest weekend day with non-missing data on sleep time will be imputed. The algorithm will start with the previous day before moving towards the following day, one day at a time, to identify the closest day with a non-missing time specification. The amount of missing data and number of imputations performed will be reported.

Participants will not be asked to note their schedule during week two (the first week of the intervention or control period). For this week, we will impute the time which the participants arrive and leave work, based on the arithmetic mean of the reported times (including those imputed above), for the same day of the week, at week 1 and week 3. We will use this information to time annotate the screen media use data only, as there during this time are no outcome measurements.

Analyses of outcomes

The primary outcome is the mean between-group difference in non-sedentary time (min/day) during leisure, in children. The primary endpoint is the primary outcome at follow-up assessment (only one follow-up).

To analyze the effect of the screen media use restriction on the outcomes of interest, we will examine between-group mean differences from baseline to follow-up. We will do so using a linear mixed-effect model, where the random effect of siblings will be included in the model when children are the unit of analyses and random effect of partner or spouse, when adults are the unit of analysis. Those variables whose mean, median or proportional value at baseline are statistically different between those allocated to screen media restriction and those allocated to control, which also are expected to independently predict the outcome, will be included as covariates in the fixed part of the model. We will include the baseline value of the outcome as a covariate (fixed effect) regardless of outcome differences between intervention and control group at baseline. We will also report the results from a crude analysis, only adjusting for the baseline level of the outcome.

Subgroup analyses

To the extent that it is possible given the data we will perform the following pre-defined sub-group analyses for the primary outcome, which are outlined in the table below.

Table 1. Pre-planned subgroup analyses for the primary outcome

Stratification variable	Justification for subgroup analysis
Sex (male/female)	It is well-established that males and females have different health behaviors. Thus, it is important to investigate potential sex differences in the effect of screen restriction on non-sedentary time during leisure.
Age: 4-7 years versus ≥8 years	As biological age, including level of maturity, is an important determinant of physical activity behavior, differences in effect between age groups will be investigated.
Baseline amount of leisure non-sedentary time	It is important to investigate whether there are differences in the effect (i.e. change), depending on the level (above/below median) of the outcome at baseline.
Baseline amount of leisure screen time	Similarly, as a above, but for the level (above/below median) of screen time at baseline.
Baseline amount of leisure time	As the baseline amount of leisure time most likely will influence the potential for change in non-sedentary behavior during leisure time, we will also assess whether there is an interaction in the effect based on being above or below the median for hours of available leisure time.
Baseline amount of shared leisure time within in the family	The potential for change in non-sedentary time may be partly understood by the social environment in the family. We will quantify the amount (above/below median) of overlapping (potentially shared) leisure time between the parents (either parent) and siblings with each child (unit of analysis).
Siblings of similar age	Another social factor which may influence change in non-sedentary time is whether a child has siblings of similar age whom she or he may interact with. We will categorize children according to whether they have siblings that are no more than three years younger or older than themselves.
Season of intervention period	The SCREENS trial spans all seasons of the year and because screen media use in the family most likely is influenced by seasonal variation, we will assess whether season modifies the effect. Seasons (Northern Hemisphere) are defined as Spring; 1st of March – 31st of May, Summer: 1st of June – 31st of August, Autumn; 1st of September – 30th of November, Winter; 1st of December – 28th of February. If the data does not allow for assessment based on all four seasons, we will dichotomize seasons into spring and summer versus autumn and winter.

The table above gives a brief overview of the different pre-planned subgroup analyses along with a justification for planning to conduct said analyses.

We will test for interaction using the likelihood ratio test comparing a model with an interaction term between the grouping variable and one of the variables above as well as the main effects, with a model including the main effects only.

We will check all the linear mixed effect models for the basic statistical assumptions, which underlie the integrity of the models. We will visually assess normality of residuals, linearity of each independent variable to the dependent variable, homogeneity of variance of the residuals, as well as check that means from the fixed effect of the cluster groups are not correlated with the effects associated with the cluster variable groups.

The main analyses will be conducted using multivariate chained equation imputation (MICE) approach including all covariates to impute missing values on covariates under the assumption that the data is missing at random. Here, we will investigate the possibility of selection bias arising from missing data, e.g. due to loss to follow-up (not completing follow-up measurements) or potentially for other reasons. Our analyses will be based on 20 imputed datasets. Patterns of missing data will be assessed a priori. We will also conduct a secondary analysis; a complete case analysis based on listwise deletion. Both analyses will be conducted on an intention-to-treat basis.

A per-protocol analysis will also be carried out comparing participants based on whether they were sufficiently compliant to the trial protocol. For this analysis we will define sufficiently compliant as having no more than seven hours/week/person of leisure screen time for entertainment purposes during the intervention period. This corresponds to a level of leisure screen time for entertainment purposes of 233% of the compliance threshold (three hours/week/person).

The statistical analyses will be conducted using Stata IC 16 (Statacorp) with an $\alpha=0.05$ (two-sided testing).