Objectively measured physical activity and symptoms of psychopathology in general population adolescents from the SIGMA cohort

Noëmi Hagemann[1,2*], Olivia J. Kirtley[1], Ginette Lafit[1,3], Martien Wampers[1], Robin Achterhof[1], Karlijn S.F.M. Hermans[1], Anu P. Hiekkaranta[1], Aleksandra Lecei[4], Davy Vancampfort[2] & Inez Myin-Germeys[1]

[1] *KU Leuven, Department of Neurosciences, Research Group Psychiatry, Center for Contextual Psychiatry, Leuven, Belgium*

[2] *KU Leuven, Department of Rehabilitation Sciences, Research Group Adapted Physical Activity and Psychomotor Rehabilitation, Leuven, Belgium*

[3]*KU Leuven, Department of Psychology, Research Group on Quantitative Psychology and Individual Differences, Leuven, Belgium*

[4]*KU Leuven, Department of Neurosciences, Research Group Psychiatry, Center for Clinical Psychiatry, Leuven, Belgium*

*Corresponding author: Noëmi Hagemann, Department of Neurosciences, Research Group Psychiatry, Center for Contextual Psychiatry, Kapucijnenvoer 33 bus 7001, 3000 Leuven, Belgium
E-mail: noemi.hagemann@kuleuven.be

CRediT statement

Conceptualization: [Noëmi Hagemann]; Methodology: [Ginette Lafit; Martien Wampers; Noëmi Hagemann], Formal analysis and investigation: [Ginette Lafit; Noëmi Hagemann]; Writing - original draft preparation: [Noëmi Hagemann]; Writing - review and editing: [Olivia J. Kirtley; Davy Vancampfort; Ginette Lafit; Inez Myin-Germeys; Martien Wampers; Robin Achterhof; Anu P. Hiekkaranta; Karlijn Hermans; Aleksandra Lecei]; Funding acquisition: [Noëmi Hagemann; Inez Myin-Germeys]; Supervision: [Olivia J. Kirtley; Davy Vancampfort; Inez Myin-Germeys].

Funding

Noëmi Hagemann, Ginette Lafit, Robin Achterhof, Karlijn Hermans, Anu Hiekkaranta, Martien Wampers, and Inez Myin-Germeys are supported by a Research Foundation Flanders Odysseus grant to Inez Myin-Germeys (FWO G0F8416N). Noëmi Hagemann is also supported by a doctoral grant by the German scholarship foundation Cusanuswerk e.V. Olivia Kirtley is supported by a Senior Postdoctoral Research Fellowship (FWO 1257821N) from Research Foundation Flanders. Aleksandra Lecei is supported by a Doctoral Grant (FWO 1104219N) from Research Foundation Flanders.

Conflicts of interest/Competing interests

The authors have no financial or non-financial interests to disclose.

Compliance with Ethical Standards

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Ethical approval for this study was obtained from the UZ/KU Leuven Medical Ethics Committee (Ref: S6 1395).

Consent to participate

Written informed consent was obtained from all individual participants included in the study. Additionally, written informed consent was obtained from the participants' parents/ care givers/ legal guardians if the participants were aged below 18 years.

Consent to publish

Written informed consent regarding publishing was obtained from all individual participants included in the study. Additionally, written informed consent regarding publishing was obtained from the participants' parents/ care givers/ legal guardians if the participants were aged below 18 years.

Acknowledgements

We would like to thank the research assistants Tessa Biesemans, Lore Depraetere and Jolien Bynens, data manager Martien Wampers, research coordinator Silke Apers, and administrative and financial coordinator Martine van Nierop for their invaluable work. We are extremely grateful for the assistance from master students Louise Dendoncker, Lore Dumolein, Romanie Kint, and Silke De Neve in the data extraction process. We would also like to thank the collaborating schools, teachers, and participants from the SIGMA study, who were incredibly supportive in the SIGMA study collaboration.

Abstract

Background: Less physical activity (PA) has been associated with the development of psychopathology in adolescence. Few studies, however, have focused on understanding the nature of the PA – psychopathology relationship and existing research relies mostly on self-report PA measures, which are less reliable. In this study, we investigated the nature of the relationships between objectively measured light physical activity (LPA) and moderate-vigorous physical activity (MVPA), and psychopathology symptoms in adolescents.

Methods: 934 adolescents (63% female) aged 11-19 years from the SIGMA cohort wore the Fitbit Charge 2 measuring LPA and MVPA during at least three days. Participants completed the Brief Symptom Inventory-53, assessing general symptoms of psychopathology, depression, anxiety, and psychoticism. Model selection was conducted in a subset of the dataset (n = 464) to determine the best fit. The optimal model was then fitted to the remainder of the dataset (n= 470).

Results: The results from the selected linear model indicated a non-significant relationship between LPA, MVPA, and symptoms of general psychopathology, depression, anxiety, and psychoticism.

Conclusions: This study does not provide evidence that higher durations of LPA and MVPA alone relate to reduced symptoms of general psychopathology, depression, anxiety, and psychoticism in the general adolescent population. A more integrative approach considering the interdependency of multiple lifestyle factors, as well as the domain, context, and individual experience of PA may add value to the understanding of the PA- psychopathology relationship.

Introduction

The transitional period of adolescence represents a phase of rapid growth and development, during which the individual grows towards more independence and adulthood (Esnaola et al., 2020; Farooq et al., 2020; Spear, 2013). In this challenging phase, adolescents are particularly vulnerable to developing symptoms of psychopathology (Kessler et al., 2007). One of the potential risk factors for the development of psychopathology is adolescents' physical inactivity (Schuch et al., 2018, 2019). Current physical activity (PA) guidelines recommend that adolescents should spend on average at least 60 minutes in moderate-vigorous physical activity (MVPA) on a daily basis to receive health benefits, such as reducing symptoms of depression (World Health Organization, 2020). Yet more than 80% of the adolescent population worldwide does not meet these recommendations (Guthold et al., 2020).

Recent literature suggests that more PA is associated with lower psychopathology symptoms among children and adolescents, independently of other lifestyle factors (e.g. sedentary behaviour). There is, for example, supportive evidence from longitudinal studies indicating that more PA is associated with a lower risk for depression (Schuch et al., 2018), anxiety (Schuch et al., 2019) and psychosis (Koivukangas et al., 2010; Suetani et al., 2017). Potential underlying mechanisms for the mental health benefits of PA in adolescents include psychosocial factors (e.g., improved self-esteem, better social connectedness), behavioural factors (e.g., improved sleep volume, better coping strategies), and neurobiological mechanisms (e.g., higher neuroplasticity, decreased inflammation) (Lubans et al., 2016).

Despite this seemingly convincing evidence for an association between PA and psychopathology, the existing literature exploring associations between PA and psychopathology is hampered by shortcomings. First, the majority of studies rely on single-timepoint self-report surveys. Subjective measures are less reliable, as adolescents underestimate durations of low physical activity (LPA) (Sullivan et al., 2012), and overestimate durations of MVPA (LeBlanc & Janssen, 2010; Manios et al., 2013), as compared to objective measures (e.g. accelerometery). Consequently, objective measures (e.g. Fitbit Charge 2) may provide more accurate and higher quality evidence. Fitbit Charge 2 devices do not only overcome the limitations of subjective measures by continuously capturing PA as it occurs in daily life, they are also less cost-intensive than research-grade devices (e.g. Actigrapgh) - which could allow greater sample size - while achieving comparable measurement validity (Brewer et al., 2017).

Second, studies investigating associations between objectively measured PA duration and psychopathology have mainly focused on the effects of higher intensity PA, i.e. MVPA. Yet, lower intensity PA, (LPA), may also relate to symptoms psychopathology independently of MVPA (Kandola et al., 2020). Given that adolescents spent at least 50% of their daily activity in LPA (Hoos et al., 2004), and that LPA may provide a lower threshold for being active than MVPA to reach activity goals, it is important to better understand the associations between LPA and the development of psychopathology symptoms. Further, previous prospective studies on objectively measured PA have focused on effects of PA on symptoms of depression in adolescents (Kandola et al., 2020; Toseeb et al., 2014; Wiles et al., 2012), while associations with anxiety, psychoticism, and general psychopathology need further exploration.

Third, knowledge regarding the *nature* of the PA - psychopathology relationship between objectively measured LPA and MVPA and psychopathology, is missing in adolescents. It is unclear if there is a consistent, linear relationship, where more LPA and MVPA relates to fewer symptoms of psychopathology, or if the associations are non-linear, and only emerge after a certain threshold of LPA and MVPA has been reached. In a study in adults (n=8150), for example, objectively measured LPA was only associated with symptoms of psychopathology after a threshold of 400 min/day of LPA had been reached (Bernard et al., 2018). Therefore, understanding the nature of the relationship more clearly in adolescents may be relevant for providing evidence- based, psychopathology-informed PA guidelines in the long-term.

Considering these shortcomings in the literature, the current study aims to examine, in adolescents, (1) what kind of model (e.g. linear or non-linear) represents the nature of the objectively PA- general psychopathology relationship best, and (2) if a relationship exists between objectively measured LPA, MVPA and general psychopathology symptoms. The registered confirmatory analyses investigated the relationship between LPA, MVPA, and general psychopathology. We hypothesise that more objectively measured LPA and MVPA are linearly associated with fewer symptoms of general psychopathology. Further exploratory analyses were conducted to examine the relationships between

LPA, MVPA, depression, anxiety, and psychoticism, subscales of the general psychopathology measure.

Methods

Sample and recruitment

Data from Wave 1 of the SIGMA study, a large-scale, ongoing, longitudinal study of adolescent mental health in Flanders, Belgium, are used in this study. In total, 1913 adolescents were included in Wave 1 (age range 11-19 at enrolment). Full details of the methods, measures and sample have been described elsewhere (Kirtley et al., in preparation). The participants in the SIGMA study were recruited from the general population via schools, with an opt-in consent procedure. Students at each school were invited to take part in the study, regardless of sociodemographic or psychological factors, e.g. known presence/absence of psychological disorder. The inclusion criteria were: (a) being a current student in the first, third or fifth year at a participating school; and (b) being able to read and write in Dutch. Ethical approval for this study was obtained from the UZ/KU Leuven Medical Ethics Committee (Ref: S6 1395). This study was post-registered on the Open Science Framework (a form of pre-registration that occurs following data collection, but before conducting the analyses; (Benning et al., 2019)) available via https://osf.io/jt8zb/?view_only=0b80007da3af473bbfca83a22f1a27ac. Deviations from the post-registration are reported in ***Supplement 1***.

Procedure

Self-reported symptoms of psychopathology were assessed in an initial testing session in the classroom. Following completion, students were instructed to wear a Fitbit Charge 2 device on their wrist, during seven consecutive days and 24 hours per day. The Fitbit Charge 2 continuously collects data on the participant's light (LPA), moderate (MPA), and vigorous physical activity (VPA). Participants were asked to remove the device whenever they would be in contact with water (e.g. when swimming, showering) since the device is not water resistant. After the 7-day wear period, they returned the Fitbit at their school. Participants received a 10-euro voucher for participation in the entire SIGMA study.

Symptoms of psychopathology

The Brief Symptom Inventory-53 (BSI-53; Derogatis, 1993) was used to assess symptoms of psychopathology. It includes ten different subscales on somatization, compulsiveness, social insecurity, depression, anxiety, aggression, phobia, paranoia, psychoticism, and additional items (e.g. suicidal thoughts, sleep quality). Example items are feeling 'unworthy', 'nervous', or 'distrustful to most people' from the depression, anxiety, and psychosis subscales, respectively. For all 53 items, participants were asked to report how much they had experienced these thoughts or feelings during the last week, including the day of assessment, on a scale ranging from 0 ('not at all') to 4 ('very much'). There was also a response option 'I do not want to answer this question'.[1]

As outcome variables, we used the Global Severity Index (GSI), which is a mean score of all 53 items, and mean values of the subscales of depression, anxiety, and psychosis. The Dutch version of the BSI-53 is sufficiently valid and reliable (Beurs et al., 2006). In the current study, overall reliability was good: McDonald's Omega $\omega = .96$, as was the reliability of the subscales depression ($\omega = .90$), anxiety ($\omega =.87$), and psychoticism ($\omega = .75$). GSI scores were calculated in accordance with the BSI-53 guidelines. Therefore, participants with BSI- 53 data containing more than 3 missing values in total or more than 1 missing value per subscale were excluded.

Physical activity

We measured LPA, MPA, and VPA with the consumer wearable device Fitbit Charge 2. To create average daily minutes of LPA, MPA, and VPA, the sum of activity minutes over all valid days was divided by the number of valid measurement days. LPA, MPA, and VPA were based on an algorithm from Fitbit using metabolic equivalents (METs). The algorithm counts MPA and VPA minutes from an intensity level of 3 METs or higher and from a duration of 10 minutes or longer. Previous research demonstrates acceptable concurrent validity ($r = .658$) between the Fitbit Charge 2 and research-grade wearables such as Actigraph when assessing LPA, MPA, and VPA over a 7-day period (Brewer et al., 2017).

[1] This option was introduced to all questionnaires at the request of schools (see Kirtley et al., 2021 for further information).

The choice of the Fitbit Charge 2 device was motivated by good assessment practice of finding a balance between validity and feasibility of measurement instruments (Corder et al., 2008). Compared to adults, adolescents would be less willing to wear a device that clashes with their idea of 'social wearability', which is defined by the degree to which adolescents perceive the visual appearance of the device as acceptable in a social setting (Corder et al., 2008; Dunne et al., 2014).

The screen of the device was disabled to provide feedback on the participants' activity, yet the distance (km) they covered daily was displayed as this could not be turned off. We decided a priori to create a variable of MVPA, a sum of MPA and VPA. We originally planned to also include sedentary behaviour in this study, but in a deviation from our post-registration, we decided not to impute the missing SB data or proceed with analysis of sedentary behaviour due to high amounts of missing data (~58%), (see ***Supplement 2*** for more information).

Inclusion and exclusion of data into the final dataset

We used a threshold of at least ≥3 valid measurement days as an inclusion criterion in line with previous research (Kandola et al., 2020). Since a minimum of 4 valid days has been suggested as well (Migueles et al., 2017), we conducted an additional sensitivity analysis (see ***Supplement 3***). Sufficient wear time usually includes 8-10 hours in a 24 hr day (Migueles et al., 2017). Common proxies for non-wear time (e.g. heart rate, aerometry count) were not available for this Fitbit day-level data sample. Thus, a valid day was alternatively defined as having a measurement day with: (a) LPA for ≥60 minutes and (b) having started measurement day 1 before or the latest at 12noon, and (c) having a sleep period of more than 200 minutes and with a sleep onset after 1pm. If on all seven measurement days the participant had zero values for all measured variables of the device (LPA, MPA, VPA, minutes asleep), we assumed that the device was not worn, and data were excluded. For more details on defining missing data, non-wear time, and the rationale for in- and exclusion criteria we refer to ***Supplement 2***.

Covariates

Longitudinal evidence suggests that adolescents are becoming less moderate-vigorously physically active with age, and this has been observed in both genders, although particularly in girls (Farooq et al., 2020). Further, a cross-sectional study from Germany with comparable seasonal conditions shows that adolescents are less light to moderately physically active during winter and on rainy and short sunlight days (Quante et al., 2019). Therefore, we included age (years), gender (male/female), and season (spring/summer/autumn/winter) as covariates into our analyses.

Statistical analysis

The aims of the statistical analysis were threefold. Given that there is no established ‘gold standard’ for handling missingness in Fitbit data, first, we wanted to explore the nature of missing data patterns and find a suitable method to deal with missing data (e.g. imputation). Second, since the nature of the relationship between PA and psychopathology is currently unclear (e.g. linear, non-linear), we wanted to investigate if the relationship between PA and psychopathology is explained better by a linear or non-linear model. In order to address these two aims, we used a random subsample, comprising 50% of the total sample, as a training set. Within the training dataset, we first explored the missing data pattern and developed a suitable technique for handling missing data. Then, we conducted a model selection procedure with four models. The lowest Bayesian information Criterion (BIC) indicated the best fitting model. Third and finally, we fitted the selected model to the remaining data from the total sample, i.e. the testing dataset.

Missing data exploration and imputation

We used a random 50% subsample (n=587) in order to explore the amount and pattern of missing data in the PA variables and select the best fitting model for our data. After having excluded participants who did not wear the device at all during the 7-day measurement period (N=23), and participants who had less than 3 days of valid data (N=100), the training subsample consisted of 464 participants. Based on our criteria for missing data (**see *Supplement 2***), 17.87% of LPA, MPA, and VPA data, and 48.64% of sedentary behaviour data, were missing after applying our criteria for in- and exclusion. We used Little’s Test for Missing Completely At Random (MCAR) (Little, 1988) within Stata Version 15.1 (StataCorp, 2017) to explore the pattern of the missingness in the dataset.

The results indicated that the missing data within the training dataset were not missing completely at random (X^2= 179.96; df= 31; p= ≤.001).

There is a lack of information regarding best practices for dealing with missing Fitbit data (e.g. Feehan et al., 2018), therefore we used a multiple imputation method to impute missing data in multilevel models (van Ginkel et al., 2020). This approach is preferred since it considered that there a repeated measurements per participant. The data within this study have a multilevel structure, where accelerometery data are nested within participants, which are nested within schools. In general, a recommended approach is then to perform a multilevel multiple imputation (van Ginkel et al., 2020). We followed this general recommendation, but did not empirically evaluate this approach, or compare approaches, within our sample.

To perform a multiple imputation technique for multilevel models, we used the 'mitml'-package (Grund et al., 2019) in R (R Core Team, 2019), which imputed the missing values for LPA, MPA, and VPA and generated ten imputed datasets. In general, three to ten imputations are considered to be sufficient (Rubin, 1987; Von Hippel, 2018). We chose 10 imputations to maximise the efficiency of estimates and replicability (Rubin, 1987). The same multiple imputation technique for multilevel model was applied within all ten datasets. After data were fully cleaned and all inclusion and exclusion criteria were applied as described above, the missing data were imputed.

We did not investigate if the estimates differed significantly between imputed datasets. However, previous work has illustrated that the imputation of point estimates has good replicability even if missingness is very high (e.g. 75%), with point estimates differing only by 0.1% between the first and second imputation (Von Hippel, 2018). Finally, we created a mean value for LPA, MPA, and VPA across the number of observations per participants in every imputed dataset. MPA and VPA were summed to create the MVPA variable, which we used in our analysis.

Model selection with the training subsample (n=464)

After imputing the missing data, we estimated for each imputed dataset the following models: (1) a linear regression, (2) a linear mixed effects model considering school differences, (3) a generalized additive model, and (4) a generalized additive mixed effect model considering school differences. Model 1 and 2 assume a linear relationship between PA and psychopathology, while model 3 and 4 assume non-linear one, to test the ‘threshold’ effect. Mixed effects models were included in the model selection procedure to investigate if accounting for school differences would improve the model fit. All models were estimated using the mgcv package (Wood, 2017). The linear model was estimated with the gls function (1), the linear mixed effects model by the lme function (2), and the generalized additive models (3, 4) with the gam function. To allow for comparability, the number of knots for the generalized additive models were fixed (k=3), and Maximum Likelihood (ML) Estimation was chosen for all models. For each of the four different models we wanted to compare, we computed the Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), and mean squared prediction error (MSPE).

The MSPE represents a summary of differences between actual and predicted response values and indicates how well a model predicts in the future and is calculated by k-fold cross-validation (Hastie et al., 2008). In this approach, the sample is split into k subsets. Using all subsets except the kth subset, the model is estimated for each fold, whereby the kth subset serves a validation sample. This process is repeated until every subset has served as the validation sample. Then, the computed MSPE values are averaged across the folds (Hastie et al., 2008). In this study, we used 10-fold cross-validation in each imputed dataset to compute the MSPE value, which was later pooled across the M imputed datasets (Hastie et al., 2008). The best fit of the model was determined by the lowest AIC, BIC, and MSPE value across all imputed datasets.

Estimation of the selected model in the testing subsample (n=464)

After selecting the model with the better fit, we estimated the model to assess (a) our confirmatory hypothesis on the association between LPA and MVPA and general symptoms of psychopathology, and (b) exploratory hypothesis on the association between LPA and MVPA on symptoms of depression, anxiety, and psychoticism. In preparation of the analysis of the testing sample, same steps were taken as in the training sample: After the data of the testing sample (n=586) had been cleaned, 10 different imputed datasets were created. We used the same multiple imputation method for 2-level data as in the training subsample. In every imputed dataset, we created a mean value for LPA, MPA, and VPA across the number of observations per participants. MPA and VPA were summed to create an additional value; MVPA.

We used Rubin's rules for pooling effect estimates and standard errors (Rubin, 1987). The Wald test was calculated using the pooled regression coefficient and standard error. The degrees of freedom (df) and the p-value for the pooled estimate were derived from the adjusted formula version to calculate the df (Van Buuren, 2018) using the MICE package in R (van Buuren & Groothuis-Oudshoorn, 2011). In a deviation from our post-registration, we performed a sensitivity power analysis using a fixed multiple linear regression modelling R^2 increase within the testing subsample (n=464) using G*Power version 3.1.9.4. (Faul et al., 2007). Testing the hypothesis that R^2 increase is zero, the minimum effect size to achieve a power larger than 80% is $f^2 = .02$. Within our OLS regression, we achieved a partial R^2 value of .08. Please see ***Supplement 4*** for more detailed information on the power analysis calculations.

Results

Characteristics of the training and testing subsample

Table 1 shows the means, standard deviations, and percentages for the variables age, gender, season, LPA, MVPA, BSI-GSI, and the subscales depressiveness, anxiety, and psychoticism for both the training and the testing subset with and without data imputation. **Table 2** shows the minimum, maximum, and quartile values for light and moderate-vigorous physical activity within the unimputed and imputed training and testing subsample.

Table 1

Means, standard deviations, and percentages within the unimputed and imputed training (n=464) and testing subsample (n=470)

	Unimputed training subset (n=464)	Unimputed testing subset (n=470)	Imputed training subset (n=464)	Imputed testing subset (n=470)
Age (years)	13.27 (1.74)	13.20 (1.59)	13.27 (1.74)	13.20 (1.59)
Gender (% female)	65	60	65	60
Season: Spring (%)	45	40	45	40
Season: Autumn (%)	15	49	15	49
Season: Winter (%)	40	1	40	1
LPA (minutes)	255.19 (50.83)	248.45 (55.80)	253.43 (58.11)	249.52 (51.06)
MVPA (minutes)	47.51 (32.83)	50.59 (38.41)	47.25 (34.26)	50.83 (35.84)
BSI-GSI	.90 (.64)	.83 (.60)	.90 (.64)	.83 (.60)
Anxiety	.92 (.64)	.83 (.67)	.92 (.64)	.83 (.67)
Depressiveness	.85 (.64)	.72 (.80)	.85 (.64)	.72 (.80)
Psychoticism	.69 (.64)	.63 (.67)	.69 (.64)	.63 (.67)

Note: LPA= light physical activity, MVPA= moderate - vigorous physical activity, BSI-GSI= General Severity Index of the Brief Symptom Inventory-53

Table 2

Minimum, maximum, quartile, and skewness-index values for light (LPA) and moderate-vigorous physical activity (MVPA) within the unimputed and imputed training (n=464) and testing subsample (n=470)

Sample	Variable	Min.	Average physical activity per quartile			Max.	Skewness-Index
			1st Quartile	2nd Quartile	3rd Quartile		
Unimputed	LPA	88.25	211.96	250.33	296.58	422.00	.03
training	MVPA	0	23.94	38.45	61.67	197.00	1.58
Imputed	LPA	92.00	218.57	252.42	292.16	399.13	.05
training	MVPA	-8.91	25.10	40.64	62.03	196.37	1.49
Unimputed	LPA	66.0	208.8	273.0	320.0	501.0	-.14
testing	MVPA	0	17.00	46.50	96.75	305.00	1.34
Imputed	LPA	92.21	216.62	248.54	281.36	412.00	.04
testing	MVPA	-13.62	26.48	43.33	67.38	233.97	1.55

Note: LPA= light physical activity, MVPA= moderate - vigorous physical activity

Missing data and model selection in training subsample (n=464)

We determined the BIC, AIC, and MSPE values from a linear regression (LM), a linear mixed effect model (LME), a generalized additive model (GAM), and a generalized additive mixed effect model (GAMM) across 10 imputed datasets. Considering all values, we found that the linear regression model had the lowest BIC and MSPE values, while there was an equal number of datasets presenting the lowest AIC value in either a linear or a generalized additive model. Overall, we concluded that the linear model provided the best fit (**Table 3** and **Table 4**).

Table 3

BIC and AIC values for a linear model (LM), linear mixed effects model (LME), generalized additive model (GAM), and generalized additive mixed effects model (GAMM) among ten imputed datasets of the training sample (n=464).

Dataset	1	2	3	4	5	6	7	8	9	10
BIC										
LM	719.57	718.98	719.52	718.87	717.74	718.05	718.51	718.93	718.35	719.42
LME	725.48	724.89	725.42	724.93	723.64	723.95	724.42	724.81	724.26	725.33
GAM	727.33	726.44	729.78	725.92	725.95	726.32	725.63	725.20	726.34	727.35
GAMM	727.33	726.44	729.78	725.92	725.95	726.32	725.63	725.20	726.34	727.35
AIC										
LM	688.32	687.73	688.28	687.63	686.49	686.80	687.27	687.66	687.11	688.18
LME	690.32	689.73	690.28	689.63	688.49	688.80	689.27	689.66	689.11	690.18
GAM	688.28	687.38	690.73	686.87	686.89	687.26	686.57	686.14	687.29	688.30
GAMM	688.28	687.38	690.73	686.87	686.89	687.26	686.57	686.14	687.29	688.30

Note: BIC = Bayesian Information Criterion; AIC = Akaike Information Criterion; LM = linear model; LME = linear mixed effects model; GAM = generalized additive model; GAMM = generalized additive mixed effects model

Table 4

Mean Squared Prediction Error (MSPE) values for a linear model (LM), linear mixed effects model (LME), generalized additive model (GAM), and generalized additive mixed effects model (GAMM) among ten imputed datasets of the training sample (n=464).

Dataset	1	2	3	4	5	6	7	8	9	10
MSPE										
LM	.3751	.3810	.3743	.3800	.3907	.3905	.3843	.3866	.3775	.3821
LME	.3751	.3810	.3743	.3800	.3907	.3905	.3843	.3866	.3775	.3821
GAM	.3738	.3801	.3789	.3828	.3935	.3938	.3879	.3888	.3838	.3781
GAMM	.3738	.3801	.3789	.3828	.3935	.3938	.3879	.3888	.3838	.3781

Note: MSPE= Mean Squared Prediction Error; LM = linear model; LME = linear mixed effects model; GAM = generalized additive model; GAMM = generalized additive mixed effects model; Standard error values of the MSPE values are reported in ***Supplement 7****.*

Model fit values for the associations between light and moderate-vigorous physical activity and symptoms of general psychopathology in the testing sample (n=470)

Based on the training set results, we conducted a confirmatory analysis using a linear regression model in the testing subsample. The pooled results across the 10 imputed datasets can be found in **Table 5**. There were no significant associations between LPA and MVPA, and symptoms of general psychopathology. These findings remained unchanged when conducting a sensitivity analysis applying multilevel multiple imputation to the BSI variable.

Table 5

Associations between light (LPA) and moderate-vigorous physical activity (MVPA) and psychopathology (GSI-BSI) estimated with a linear regression in the testing subsample (n=470).

	Dependent variable			
Independent variables	GSI-BSI			
	β (SE)	p	b (SE)	p
LPA	.00009 (.0006)	.84	.0001 (.0007)	.81
MVPA	-.0008 (.0009)	.54	-.001 (.0009)	.15
Age	.04 (.02)	.07	**.05 (.02)**	**.02**
Gender	**.26 (.06)**	**<.001**	**-**	**-**
Season (Spring)	-.48 (.58)	.41	-.61 (.21)	.30
Season (Autumn)	-.34 (.58)	.56	-.49 (.60)	.41

Note: in bold = alpha < 0.05; β= unstandardized regression coefficient; b= standardized regression coefficient; SE= standard error, LPA= light physical activity, MVPA= moderate - vigorous physical activity

Model fit values for the associations between light (LPA) and moderate-vigorous physical activity (MVPA) and symptoms of depression, anxiety, and psychoticism in the testing subsample (n=470)

We conducted exploratory analyses using a linear regression model. The pooled results across the 10 imputed datasets can be found in **Table 6**. There are no significant associations between LPA and MVPA with symptoms of depression, anxiety, and psychoticism. These findings remained unchanged when conducting a sensitivity analysis applying multilevel multiple imputation to the BSI variable.

Table 6

Associations between light (LPA) and moderate-vigorous physical activity (MVPA) and general psychopathology (GSI-BSI-53) estimated with a linear regression (n=470).

	Dependent variables					
Independent variables	Depression		Anxiety		Psychoticism	
	β (SE)	p	β (SE)	p	β (SE)	p
LPA	-.0002 (.0008)	.81	.0004 (.0007)	.54	-.000007 (.0007)	.99
MVPA	-.0004 (.001)	.72	-.001 (.001)	.18	-.0006 (.001)	.55
Age	**.08 (.03)**	**.008**	.007 (.02)	.75	**.03 (.02)**	**.02**
Gender	**.30 (.08)**	**<.001**	**.24 (.07)**	**<.001**	**.22 (.07)**	**.001**
Season (Spring)	**-1.54 (.79)**	**.05**	-.22 (.65)	.74	-.45 (.66)	.49
Season (Autumn)	-1.37 (.79)	.08	-.09 (.65)	.88	-.27 (.66)	.69

Note: in bold = alpha < 0.05; β= unstandardized regression coefficient; SE= standard error, LPA= light physical activity, MVPA= moderate - vigorous physical activity. Standardized coefficients are reported in ***Supplement 5****.*

Discussion

To the best of our knowledge, this is the first study investigating the nature of the relationship between objectively measured PA and symptoms of general psychopathology, depression, anxiety, and psychoticism in adolescents. We hypothesized that a linear relationship exists, where more minutes of LPA and MVPA would be negatively associated with fewer symptoms of psychopathology in adolescents. Yet, our results showed that the (standardized) effect sizes between LPA and MVPA and symptoms of general psychopathology, depression, anxiety, and psychoticism, were negligible to small, and that the associations were non-significant. Therefore, our findings do not provide evidence that more LPA and MVPA relate to fewer symptoms of psychopathology in the general adolescent population.

Our results are not in line with previous self-report data from the general adolescent population, showing significant associations between more PA and lower levels of depression and anxiety symptoms (Korczak et al., 2017; McMahon et al., 2017). However, these previous findings need to be interpreted cautiously, as self-reported PA is less reliable and prone to social desirability (Adams et al., 2005) and recall biases, meaning adolescents might either overreport MVPA (LeBlanc & Janssen, 2010) or underreport LPA (Sullivan et al., 2012).

In contrast, our observation that MVPA durations were not associated with depression is consistent with findings from other accelerometery studies in adolescents. For example, in the ROOTS study (n= 1238; 54% girls; mean age = 13.8 years) and the Avon Longitudinal Study of Parents and Children (ALSPAC) study (n= 3298, 53% girls mean age= 14.5 years) (Kandola et al., 2020; Toseeb et al., 2014; Wiles et al., 2012), objectively measured MVPA was also not associated with symptoms of depression, and reported small, (unstandardized), effect sizes of weekday MVPA (.02).

Our findings do not provide evidence that more LPA and MVPA are associated with fewer symptoms of psychopathology in the general adolescent population. Recent prospective evidence suggests that the relationship between lifestyle behaviours and psychopathology is interdependent, where multiple lifestyle factors, such as sleep, sedentary behaviour, and PA, have an integrative effect on psychopathology, as opposed to individual lifestyle factors alone (Brown et al., 2021). These findings suggest that a more inclusive, broader approach may be needed for future research and intervention. Additionally, individuals' experience of PA, the type, domain, and context of PA need to be considered for psychopathology-informed PA recommendations (Teychenne et al., 2020).

Preliminary evidence suggests that PA is associated with fewer psychopathology symptoms when PA occurs during leisure time (White et al., 2017), in the company of others (Kleppang et al., 2017), in nature (Coon et al., 2011), and with a focus on the mind-body relationship, such as yoga (Galantino, Galbavy, & Quinn, 2008; Lubans, Plotnikoff, & Lubans, 2012). In contrast, PA during transport, household chores, and physical education classes is not associated with fewer symptoms of psychopathology (White et al., 2017). There is also evidence that PA carried out alone is associated with higher odds of developing depression compared to PA performed in a social setting (e.g. in a sports clubs) (Kleppang et al., 2018). Therefore, further research may investigate interactions of PA with domain, context, type, and experience of PA, in relation to symptoms of psychopathology in adolescents, using the Experience Sampling Method, for example.

Our finding that LPA duration and depression levels were not significantly associated is in contrast to previous research using objectively measured LPA from the ALSPAC cohort study (Kandola et al., 2020). Currently, little is known about psychopathology-related associations with LPA. Cross-sectional evidence in adults suggests that associations with LPA are threshold-dependent, showing that associations emerge only after a certain minimum duration of LPA (~400 min/day) (Bernard et al., 2018). Potentially, a similar threshold-dependent relationship exists in adolescents, where the minimum amount to affect psychopathology is reached with at least ~320 minutes of LPA per day (Kandola et al., 2020), whereas shorter durations, as in our study (~250 minutes/ day) do not. Further exploration showed that the majority of LPA values (i.e. 75%) were below an average of 320 minutes. Following a reviewer's suggestion, because our threshold of LPA inclusion was low (60 minutes), we conducted exploratory analyses with an increased threshold (300 min) to estimate the associations of higher LPA (mean 360.28, SD 25.98) with psychopathology symptoms in a linear and non-linear model in a sample of N=332. Our results indicated that in both models the linear and non-linear relationship between both LPA and MVPA, and psychopathology, remained non-significant (see **Supplement 6**).

Overall, given that no associations were found between general psychopathology, depression, anxiety, and psychoticism and LPA/MVPA duration, these findings also suggest that factors other than PA may be more important in relation to psychopathology. Further research should elucidate the role of social risk factors, for example low family cohesion and adverse life events in relation to PA in adolescents (Cuffe et al., 2005)

Since the results of the linear model investigating the PA-psychopathology relationship was non-significant, this study does not enable us to conclude whether a linear or a non-linear model describes the relationship best. In the previous literature, the nature of the PA-psychopathology relationship has been much debated. Other studies in adults, using self-report data, are inconsistent in their findings, as they have been suggesting various natures of relationships: Linear (Galper et al., 2006), quasi-linear (Stephens, 1988), s-shaped (Mummery et al., 2004), and U-shaped relationships (Kim et al., 2012). The only accelerometery study to date investigating the nature of the PA-psychopathology relationship shows a non-linear (i.e. U-shaped) association in adults (Bernard et al., 2018).

Our study has several methodological strengths. First, we aimed for maximal transparency by post-registering our study on the Open Science Framework. This approach may increase replicability of future work by minimising 'researcher degrees of freedom' and counteracting questionable research practices such as p-hacking (i.e. conducting multiple analyses until a significant result is achieved) and HARKing, standing for 'Hypothesising After Results are Known, can increase the risk of type 1-errors and reduce replicability of findings. Second, we conducted a sensitivity power analyses to increase the interpretability of our findings and guide future work in sample size decisions. Third, we addressed the risk of bias related to missingness in accelerometery data by using multiple imputation methods for multilevel data (van Ginkel et al., 2020). Given the many misconceptions about multiple imputation, listwise deletion still remains a conventional method in dealing with missing data, which may bias study conclusions (van Ginkel et al., 2020). Given the absence of a "gold standard" for handling missing Fitbit data (Feehan et al., 2018) our methods may also help to increase the rigor of future research using Fitbit devices.

Our results also need to be interpreted in light of some limitations. First, common proxies of non-wear time (e.g. heart rate; accelerometery count) were unavailable in this data sample using Fitbit

Charge 2. Therefore, we cannot definitively know if the included data represents sufficient wear time, which usually includes at least 8 hours in each 24-hour day. Future research using Fitbit devices may include minute-level heart rate measurement as a proxy for non-wear time. Second, since the device is not water resistant, inconvenient to wear during contact sports (e.g. volleyball), and MVPA durations shorter than 10 consecutive minutes are not captured by the Fitbit Charge 2, full PA may not have been captured in the data. Options for participants manually logging additional non-recorded activities could be explored. Third, while including at least three valid days is common practice in accelerometery research (Kandola et al., 2020), current recommendations suggest to include at least four valid days (Migueles et al., 2017). However, within our sensitivity analysis, in which we estimated a linear regression including at least 4 valid days, results were similar to our analyses using three valid days. Fourth, our study design did not account for the influence of additional lifestyle aspects, such as sleep and sedentary behaviour on adolescent psychopathology symptoms, although previous research suggests an interdependency and combined association of these behaviours on adolescent mental health (Brown et al., 2021). This means that the single assessment of only one lifestyle aspect (i.e. physical activity) in this study may represent an incomplete picture of the relationship (Tremblay, 2020).

To conclude, our results show that objectively measured LPA and MVPA durations are not significantly associated with symptoms of general psychopathology, depression, anxiety, and psychoticism in adolescents, and that the relationships between PA and psychopathology are best represented by a linear relationship. Overall, this study does not provide evidence that higher durations of LPA and MVPA alone and regardless of the context relate to reduced symptoms of psychopathology in the general adolescent population. A more integrative approach considering the interdependency of multiple lifestyle factors, as well as the domain, context, and individual experience of PA may add value to the understanding of the PA- psychopathology relationship.

References

Adams, S. A., Matthews, C. E., Ebbeling, C. B., Moore, C. G., Cunningham, J. E., Fulton, J., & Hebert, J. R. (2005). The effect of social desirability and social approval on self-reports of physical activity. *American Journal of Epidemiology*, *161*(4), 389–398. https://doi.org/10.1093/aje/kwi054

Benning, S. D., Bachrach, R. L., Freeman, E., & Wright, A. (2019). The Registration Continuum in Clinical Science: A Guide toward Transparent Practices. *J Abnorm Psychol*, *128*(6), 528–540. https://doi.org/doi:10.1037/abn0000451

Bernard, P., Doré, I., Romain, A. J., Hains-Monfette, G., Kingsbury, C., & Sabiston, C. (2018). Dose response association of objective physical activity with mental health in a representative national sample of adults: A cross-sectional study. *PLoS ONE*, *13*(10). https://doi.org/10.1371/journal.pone.0204682

Beurs, E. D. E., Zitman, F. G., & Inventory, F. G. D. B. S. (2006). *De Brief Symptom Inventory (BSI)*. November 2015*, 120–141.

Brewer, W., Swanson, B. T., & Ortiz, A. (2017). Validity of Fitbit's active minutes as compared with a research-grade accelerometer and self-reported measures. *BMJ Open Sport and Exercise Medicine*, *3*(1), 1–7. https://doi.org/10.1136/bmjsem-2017-000254

Brown, D. M. Y., Kwan, M. Y., Arbour-Nicitopoulos, K. P., & Cairney, J. (2021). Identifying patterns of movement behaviours in relation to depressive symptoms during adolescence: A latent profile analysis approach. *Preventive Medicine*, *143*, 106352. https://doi.org/10.1016/j.ypmed.2020.106352

Coon, J. T., Boddy, K., Stein, K., Whear, R., Barton, J., & Depledge, M. H. (2011). Does participating in physical activity in outdoor natural environments have a greater effect on physical and mental wellbeing than physical activity indoors? A systematic review. *Environmental Science and Technology*, *45*(5), 1761–1772. https://doi.org/10.1021/es102947t

Corder, K., Ekelund, U., Steele, R. M., Wareham, N. J., & Brage, S. (2008). Assessment of physical activity in youth. *Journal of Applied Physiology*, *105*(3), 977–987. https://doi.org/10.1152/japplphysiol.00094.2008Cuffe, S. P., Mckeown, R. E., Addy, C. L., & Garrison, C. Z. (2005). Family and psychosocial risk factors in a longitudinal epidemiological study of adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*(2), 121–129. https://doi.org/10.1097/00004583-200502000-00004

Derogatis, L. R. (1993). *BSI Brief Symptoms Inventory: Administration, Scoring, and Procedure Manual (4th Ed.)*. Pearson, NCS.

Dunne, L., Profita, H., & Zeagler, C. (2014). Social Aspects of Wearability and Interaction. In *Wearable Sensors: Fundamentals, Implementation and Applications*. Elsevier Inc. https://doi.org/10.1016/B978-0-12-418662-0.00026-XEsnaola, I., Sesé, A., Antonio-Agirre, I., & Azpiazu, L. (2020). The Development of Multiple Self-Concept Dimensions During Adolescence. *Journal of Research on Adolescence*, *30*(S1), 100–114. https://doi.org/10.1111/jora.12451

Farooq, A., Martin, A., Janssen, X., Wilson, M. G., Gibson, A. M., Hughes, A., & Reilly, J. J. (2020). Longitudinal changes in moderate-to-vigorous-intensity physical activity in children and adolescents: A systematic review and meta-analysis. In *Obesity Reviews* (Vol. 21, Issue 1). Blackwell Publishing Ltd. https://doi.org/10.1111/obr.12953

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Feehan, L. M., Geldman, J., Sayre, E. C., Park, C., Ezzat, A. M., Young Yoo, J., Hamilton, C. B., & Li, L. C. (2018). Accuracy of fitbit devices: Systematic review and narrative syntheses of quantitative data. *JMIR MHealth and UHealth*, *6*(8). https://doi.org/10.2196/10527

Galantino, M. Lou, Galbavy, R., & Quinn, L. (2008). Therapeutic Effects of Yoga for Children: A Systematic Review of the Literature. *Pediatric Physical Therapy*, *20*(1), 66–80. https://doi.org/10.1097/PEP.0b013e31815f1208

Galper, D. I., Trivedi, M. H., Barlow, C. E., Dunn, A. L., & Kampert, J. B. (2006). Inverse association between physical inactivity and mental health in men and women. *Medicine and Science in Sports and Exercise*, *38*(1), 173–178. https://doi.org/10.1249/01.mss.0000180883.32116.28

Grund, S., Robitzsch, A., & Luedtke, O. (2019). *mitml: Tools for Multiple Imputation in Multilevel Modeling. R package version 0.3-7.* https://cran.r-project.org/package=mitml

Guthold, R., Stevens, G. A., Riley, L. M., & Bull, F. C. (2020). Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1·6 million participants. *The Lancet Child and Adolescent Health*, *4*(1), 23–35. https://doi.org/10.1016/S2352-4642(19)30323-2

Hastie, T., Tibshirani, R., & Friedman, J. (2008). Model assessment and selection. In *The Elements of Statistical Learning. Data Minining, Inference, and Prediction* (Second edi, p. 745). Springer.

Hoos, M. B., Kuipers, H., Gerver, W. J. M., & Westerterp, K. R. (2004). Physical activity pattern of children assessed by triaxial accelerometry. *European Journal of Clinical Nutrition*, *58*(10), 1425–1428. https://doi.org/10.1038/sj.ejcn.1601991

Kandola, A., Lewis, G., Osborn, D. P. J., Stubbs, B., & Hayes, J. F. (2020). Depressive symptoms and objectively measured physical activity and sedentary behaviour throughout adolescence: a prospective cohort study. *The Lancet Psychiatry*, *7*(3), 262–271. https://doi.org/10.1016/S2215-0366(20)30034-1

Kessler, R. C., Angermeyer, M., Anthony, J. C., De Graaf, R., Demyttenaere, K., Gasquet, I., Giovanni, De Girolamo, G., Gluzman, S., Gureje, O., Haro, J. M., Kawakami, N., Karam, A., Levinson, D., Mora, M. M. E., Aguilar-Gaxiola, S., Alonso, J., Lee, S., Heeringa, S., … for the WHO World Mental Health Survey Consortium. (2007). Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry*, *6*(October), 168–176.

Kim, Y. S., Park, Y. S., Allegrante, J. P., Marks, R., Ok, H., Ok Cho, K., & Garber, C. E. (2012). Relationship between physical activity and general mental health. *Preventive Medicine*, *55*(5), 458–463. https://doi.org/10.1016/j.ypmed.2012.08.021

Kirtley, O. J., Achterhof, R., Hagemann, N., Hermans, K. S. F. M., Hiekkaranta, A. P., Lecei, A., …

Myin-Germeys, I. (2021, April 2). Initial cohort characteristics and protocol for SIGMA: An accelerated longitudinal study of environmental factors, inter- and intrapersonal processes, and mental health in adolescence. https://doi.org/10.31234/osf.io/jp2fk

Kleppang, A. L., Hartz, I., Thurston, M., & Hagquist, C. (2018). The association between physical activity and symptoms of depression in different contexts - A cross-sectional study of Norwegian adolescents. *BMC Public Health*, *18*(1). https://doi.org/10.1186/s12889-018-6257-0

Kleppang, A. L., Thurston, M., Hartz, I., & Hagquist, C. (2017). Psychological distress among Norwegian adolescents: Changes between 2001 and 2009 and associations with leisure time physical activity and screen-based sedentary behaviour. *Scandinavian Journal of Public Health*, *January*, 140349481771637. https://doi.org/10.1177/1403494817716374

Koivukangas, J., Tammelin, T., Kaakinen, M., Mäki, P., Moilanen, I., Taanila, A., & Veijola, J. (2010). Physical activity and fitness in adolescents at risk for psychosis within the Northern Finland 1986 Birth Cohort. *Schizophrenia Research*, *116*(2–3), 152–158. https://doi.org/10.1016/j.schres.2009.10.022

Korczak, D. J., Madigan, S., & Colasanto, M. (2017). Children’s physical activity and depression: A meta-analysis. In *Pediatrics* (Vol. 139, Issue 4). https://doi.org/10.1542/peds.2016-2266

LeBlanc, A. G. W., & Janssen, I. (2010). Difference between self-reported and accelerometer measured moderate-to-vigorous physical activity in youth. *Pediatric Exercise Science*, *22*(4), 523–534. https://doi.org/10.1123/pes.22.4.523

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202.

Lubans, D. R., Plotnikoff, R. C., & Lubans, N. J. (2012). Review: A systematic review of the impact of physical activity programmes on social and emotional well-being in at-risk youth. *Child and Adolescent Mental Health*, *17*(1), 2–13. https://doi.org/10.1111/j.1475-3588.2011.00623.x

Lubans, D., Richards, J., Hillman, C., Faulkner, G., Beauchamp, M., Nilsson, M., Kelly, P., Smith, J., Raine, L., & Biddle, S. (2016). Physical Activity for Cognitive and Mental Health in Youth: A Systematic Review of Mechanisms. *Pediatrics*, *138*(3), e20161642–e20161642. https://doi.org/10.1542/peds.2016-1642

Manios, Y., Androutsos, O., Moschonis, G., Birbilis, M., Maragkopoulou, K. Giannopulou, A., Argyri, E., & Kourlaba, G. (2013). Criterion validity of the Physical Activity Questionnaire for Schoolchildren (PAQ-S) in assessing physical activity levels: The Healthy Growth Study. *The Journal of Sports Medicine and Physical Fitness*, *October*, 501–508.

McMahon, E. M., Corcoran, P., O’Regan, G., Keeley, H., Cannon, M., Carli, V., Wasserman, C., Hadlaczky, G., Sarchiapone, M., Apter, A., Balazs, J., Balint, M., Bobes, J., Brunner, R., Cozman, D., Haring, C., Iosue, M., Kaess, M., Kahn, J. P., … Wasserman, D. (2017). Physical activity in European adolescents and associations with anxiety, depression and well-being. *European Child and Adolescent Psychiatry*, *26*(1), 111–122. https://doi.org/10.1007/s00787-016-0875-9

Migueles, J. H., Cadenas-Sanchez, C., Ekelund, U., Delisle Nyström, C., Mora-Gonzalez, J., Löf, M., Labayen, I., Ruiz, J. R., & Ortega, F. B. (2017). Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical

Considerations. *Sports Medicine*, *47*(9), 1821–1845. https://doi.org/10.1007/s40279-017-0716-0

Mummery, K., Schofield, G., & Caperchione, C. (2004). Physical activity dose-response effects on mental health status in older adults. *Australian and New Zealand Journal of Public Health*, *28*(2), 188–192. https://doi.org/10.1111/j.1467-842X.2004.tb00934.x

Quante, M., Wang, R., Weng, J., Kaplan, E. R., Rueschman, M., Taveras, E. M., Rifas-Shiman, S. L., Gillman, M. W., & Redline, S. (2019). Seasonal and weather variation of sleep and physical activity in 12–14-year-old children. *Behavioral Sleep Medicine*, *17*(4), 398–410. https://doi.org/10.1080/15402002.2017.1376206

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org

Rubin, D. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley.

Schuch, F. B., Stubbs, B., Meyer, J., Heissel, A., Zech, P., Vancampfort, D., Rosenbaum, S., Deenik, J., Firth, J., Ward, P. B., Carvalho, A. F., & Hiles, S. A. (2019). Physical activity protects from incident anxiety: A meta-analysis of prospective cohort studies. *Depression and Anxiety*. https://doi.org/10.1002/da.22915

Schuch, F. B., Vancampfort, D., Firth, J., Rosenbaum, S., Ward, P. B., Silva, E. S., Hallgren, M., De Leon, A. P., Dunn, A. L., Deslandes, A. C., Fleck, M. P., Carvalho, A. F., & Stubbs, B. (2018). Physical activity and incident depression: A meta-analysis of prospective cohort studies. *American Journal of Psychiatry*, *175*(7), 631–648. https://doi.org/10.1176/appi.ajp.2018.17111194

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632Spear, L. P. (2013). Adolescent neurodevelopment. *Journal of Adolescent Health*, *52*(2 SUPPL.2), S7–S13. https://doi.org/10.1016/j.jadohealth.2012.05.006

StataCorp. (2017). *Stata Statistical Software: Release 2015*. StataCorp LLC.

Stephens, T. (1988). Physical activity and mental health in the United States and Canada: Evidence from four population surveys. *Preventive Medicine*, *17*(1), 35–47. https://doi.org/10.1016/0091-7435(88)90070-9

Suetani, S., Mamun, A., Williams, G. M., Najman, J. M., McGrath, J. J., & Scott, J. G. (2017). Longitudinal association between physical activity engagement during adolescence and mental health outcomes in young adults: A 21-year birth cohort study. *Journal of Psychiatric Research*, *94*, 116–123. https://doi.org/10.1016/j.jpsychires.2017.06.013

Sullivan, R., Kinra, S., Ekelund, U., Vaz, M., Kurpad, A., Collier, T., Srinath Reddy, K., Prabhakaran, D., Ebrahim, S., & Kuper, H. (2012). Evaluation of the Indian Migration Study Physical Activity Questionnaire (IMS-PAQ): a cross-sectional study. In *International Journal of Behavioral Nutrition and Physical Activity* (Vol. 9). https://doi.org/10.1186/1479-5868-9-13

Teychenne, M., White, R. L., Richards, J., Schuch, F. B., Rosenbaum, S., & Bennie, J. A. (2020). Do we need physical activity guidelines for mental health: What does the evidence tell us? *Mental Health and Physical Activity*, *18*(August 2019). https://doi.org/10.1016/j.mhpa.2019.100315

Toseeb, U., Brage, S., Corder, K., Dunn, V. J., Jones, P. B., Owens, M., St Clair, M. C., Van Sluijs, E. M. F., & Goodyer, I. M. (2014). Exercise and depressive symptoms in adolescents: A longitudinal cohort study. *JAMA Pediatrics*, *168*(12), 1093–1100. https://doi.org/10.1001/jamapediatrics.2014.1794

Van Buuren, S. (2018). *Flexible Imputation of Missing Data* (Second). Chapman & Hall/ CRC.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://www.jstatsoft.org/v45/i03/

van Ginkel, J. R., Linting, M., Rippe, R. C. A., & van der Voort, A. (2020). Rebutting Existing Misconceptions About Multiple Imputation as a Method for Handling Missing Data. *Journal of Personality Assessment*, *102*(3), 297–308. https://doi.org/10.1080/00223891.2018.1530680

Von Hippel, P. T. (2018). How many imputations do you need? A two-stage calculation using a quadratic rule. In *Sociological Methods and Research*.

White, R. L., Babic, M. J., Parker, P., Lubans, D. R., Astell-Burt, T., & Lonsdale, C. (2017). Domain-Specific Physical Activity and Mental Health: A Meta-analysis. *American Journal of Preventive Medicine*, *52*(5). https://doi.org/https://doi.org/10.1016/j.amepre.2016.12.008

Wiles, N. J., Haase, A. M., Lawlor, D. A., Ness, A., & Lewis, G. (2012). Physical activity and depression in adolescents: Cross-sectional findings from the ALSPAC cohort. *Social Psychiatry and Psychiatric Epidemiology*, *47*(7), 1023–1033. https://doi.org/10.1007/s00127-011-0422-4

World Health Organization. (2020). WHO Guidelines on physical activity, sedentary behaviour. *World Health Organization*, 104. https://apps.who.int/iris/bitstream/handle/10665/325147/WHO-NMH-PND-2019.4-eng.pdf?sequence=1&isAllowed=y%0Ahttp://www.who.int/iris/handle/10665/311664%0Ahttps://apps.who.int/iris/handle/10665/325147