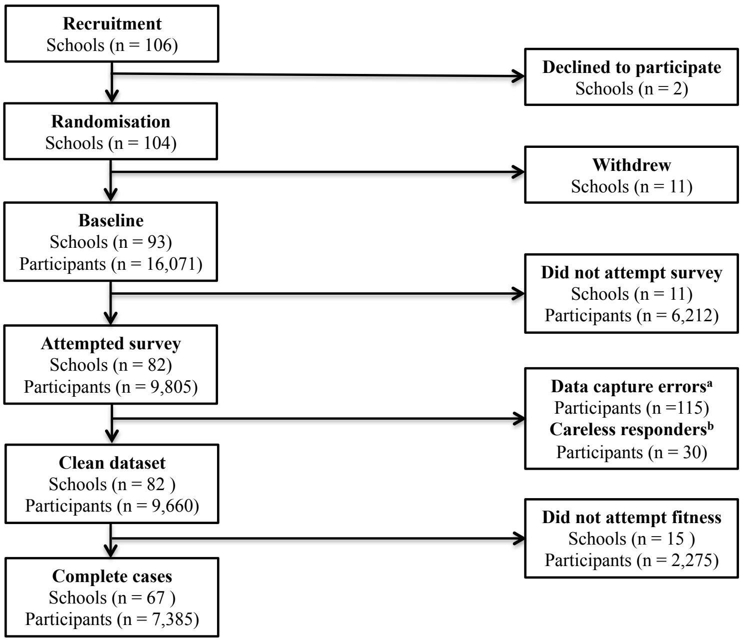
## Supplementary Material

### Supplementary Fig. 1: Flow chart showing Fit to Study recruitment, baseline sample and final subsample (n = 7,385)




[a]We removed 115 cases with missing data, caused by software failing to record participants' responses.
[b]Items reported in this study were from a larger battery arranged in three blocks of items measured on similar Likert scales. We took a conservative approach to identifying careless responders, searching for long strings of identical answers by position on the scale, accounting for reversed items searching for long strings of identical answers and removing participants who gave the same answer (by position on the scale, accounting for reversed items) throughout all blocks (25, 30 and 16 items per block). We calculated the mean string length per block (MSL) and removed 30 participants who gave the same answer throughout one block, and recorded a string of MSL+3SD in at least one more block.

### Secondary Analyses

We re-ran each of the four multilevel models exploring associations between physical activity, fitness and internalising and externalising symptoms using the full, imputed data set (n=16,017), controlling for nuisance variables as before. We explored predictors of the probability of missingness using multilevel logistic regression. Analyses revealed that data were not missing completely at random, lending support for the missing at random assumption. Data were imputed using multilevel multiple imputation available in the *jomo* package in R (1). *Jomo* uses a multivariate normal modelling approach fitted by Markov Chain Monte Carlo (MCMC). We used joint model substantive-model compatible imputation (2) within jomo to ensure compatibility between the imputation model and the substantive model of interest, which included multiple interaction terms. The imputation model included all variables that were part of the substantive model, as well as any variable (pre- or post-intervention) that had a correlation $\geq 0{\cdot}3$ with the model variables themselves, or, alternatively, with missingness in these variables. Due to the presence of post-intervention data in the imputation models, we split the imputations by treatment arm. We employed a common covariance matrix and, after a burn-in of at least 4,000 iterations, we imputed each data set 50 times, specifying 1,000 iterations between imputations. The MCMC chains were inspected to check for convergence. Operations were performed on transformed outcome variables (3); the reference category for sex was female rather than male as in the primary analyses. Model parameters and their standard errors were estimated for each imputed dataset and combined using Rubin's rules (4).

## Supplementary Table 1: Estimates of associations between activity, fitness and internalising symptoms in full, imputed data set (n = 16,017)

| Model 1 Estimate[a] | ß | 2·5% CI | 97·5% CI | Model 2 Estimate[a] | ß | 2·5% CI | 97·5% CI |
|---|---|---|---|---|---|---|---|
| PA | -0·14*** | -0·17 | -0·11 | Fit | -0·34*** | -0·37 | -0·30 |
| eFSM[b] | 0·08* | 0·01 | 0·16 | eFSM[b] | -0·02 | -0·10 | 0·05 |
| Sex[c] | -0·20*** | -0·25 | -0·16 | Sex[c] | -0·10*** | -0·15 | -0·06 |
| PA*eFSM | 0·004 | -0·06 | 0·07 | Fit*eFSM | -0·10 | -0·21 | 0·01 |
| PA*Sex | -0·10*** | -0·14 | -0·06 | Fit*Sex | 0·06** | 0·01 | 0·11 |
| PA*eFSM*Sex | 0·03 | -0·07 | 0·13 | Fit*eFSM *Sex | 0·17* | -0·03 | 0·32 |

[a]Fully-adjusted model including age, sex, eFSM, nuisance covariates and school effects; physical activity/fitness and internalising scores are z-scored; internalising symptoms are square-root transformed
[b]Reference category: not eligible for FSM
[c]Reference category: female
*** p<0·001; **p<0·0125; * p <0·05

## Supplementary Table 2: Estimates of associations between activity, fitness and externalising symptoms in full, imputed dataset (n = 16,017)

| Model 3 Estimate[a] | ß | 2·5% CI | 97·5% CI | Model 4 Estimate[a] | ß | 2·5% CI | 97·5% CI |
|---|---|---|---|---|---|---|---|
| PA | -0·10*** | -0·13 | -0·07 | Fit | -0·12*** | -0·17 | -0·06 |
| eFSM[b] | 0·28*** | 0·22 | 0·35 | eFSM[b] | 0·19** | 0·07 | 0·32 |
| Sex[c] | 0·21*** | 0·16 | 0·25 | Sex[c] | 0·23*** | 0·17 | 0·30 |
| PA*eFSM | 0·01 | -0·05 | 0·07 | Fit*eFSM | -0·11 | -0·30 | 0·07 |
| PA*Sex | 0·03 | -0·00 | 0·07 | Fit*Sex | -0·06 | -0·005 | 0·14 |
| PA*eFSM*Sex | -0·01 | -0·12 | 0·09 | Fit*eFSM*Sex | 0·14 | -0·09 | 0·37 |

[a]Fully-adjusted model including age, sex, eFSM, nuisance covariates and school effects; physical activity/fitness and exterternalising scores are z-scored; externalising symptoms are rank-inverse transformed
[b]Reference category: not eligible for FSM
[c]Reference category: female
*** p<0·001; **p<0·0125; * p <0·05

## Supplementary Table 3: Active lifestyle and mental health variable loadings on respective canonical variates

| Mode 1 | | | | Mode 2 | | | |
|---|---|---|---|---|---|---|---|
| **Active Lifestyle** | **$R_s$** | **Mental Health** | **$R_s$** | **Active Lifestyle** | **$R_s$** | **Mental Health** | **$R_s$** |
| PA | 0·35 | Emotional | -0·28 | PA | -0·02 | Emotional | -0·20 |
| Habitual PA | 0·42 | Peer | -0·26 | Habitual PA | -0·02 | Peer | 0·08 |
| Fitness | 0·28 | Conduct | -0·15 | Fitness | 0·02 | Conduct | 0·14 |
| Attitude to PA | 0·46 | Hyperactivity | -0·12 | Attitude to PA | -0·04 | Hyperactivity | 0·08 |
| FSM | -0·06 | Pro-social | 0·23 | FSM | 0·05 | Pro-social | -0·23 |
| Sex | -0·05 | Global SE | 0·42 | Sex | -0·34 | Global SE | 0·00 |
| Lifestyle*FSM | 0·01 | Physical SE | 0·47 | Lifestyle*FSM | 0·01 | Physical SE | 0·06 |
| Lifestyle*Sex | 0·02 | | | Lifestyle*Sex | 0·02 | | |
| Lifestyle*FSM*Sex | 0·00 | | | Lifestyle*FSM*Sex | 0·01 | | |

## References


1. Quartagno M, Grund S, Carpenter J. jomo: A Flexible Package for Two-level Joint Modelling Multiple Imputation. R J. 2019;
2. Bartlett JW, Seaman SR, White IR, Carpenter JR, Initiative* ADN. Multiple imputation of covariates by fully conditional specification: accommodating the substantive model. Stat Methods Med Res. 2015;24(4):462–87.
3. Lee KJ, Carlin JB. Multiple imputation in the presence of non-normal data. Stat Med. 2017;36(4):606–17.
4. Rubin DB. Multiple imputation for nonresponse in surveys. Vol. 81. John Wiley & Sons; 2004.