Physical activity and risks of breast and colorectal cancer: a Mendelian randomisation analysis
Nikos Papadimitriou, Niki Dimou, Konstantinos Tsilidis, Barbara Banbury, Richard Martin, Sarah Lewis, Nabila Kazmi, Timothy Robinson, Demetrios Albanes, Krasimira Aleksandrova, et al.

To cite this version:
Nikos Papadimitriou, Niki Dimou, Konstantinos Tsilidis, Barbara Banbury, Richard Martin, et al.. Physical activity and risks of breast and colorectal cancer: a Mendelian randomisation analysis. Nature Communications, 2020, 11 (1), 10.1038/s41467-020-14389-8. hal-02872359

HAL Id: hal-02872359
https://u-paris.hal.science/hal-02872359v1
Submitted on 27 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Physical activity has been associated with lower risks of breast and colorectal cancer in epidemiological studies; however, it is unknown if these associations are causal or confounded. In two-sample Mendelian randomisation analyses, using summary genetic data from the UK Biobank and GWA consortia, we found that a one standard deviation increment in average acceleration was associated with lower risks of breast cancer (odds ratio [OR]: 0.51, 95% confidence interval [CI]: 0.27 to 0.98, P-value = 0.04) and colorectal cancer (OR: 0.66, 95% CI: 0.48 to 0.90, P-value = 0.01). We found similar magnitude inverse associations for estrogen positive (ER+ve) breast cancer and for colon cancer. Our results support a potentially causal relationship between higher physical activity levels and lower risks of breast cancer and colorectal cancer. Based on these data, the promotion of physical activity is probably an effective strategy in the primary prevention of these commonly diagnosed cancers.

#A full list of authors and their affiliations appears at the end of the paper.
Breast and colorectal cancer are two of the most common cancers globally with a combined estimated number of 4 million new cases and 1.5 million deaths in 2018. Physical activity is widely promoted along with good nutrition, maintaining a healthy weight, and refraining from smoking, as key components of a healthy lifestyle that contribute to lower risks of several non-communicable diseases such as cardiovascular disease, diabetes, and cancer.

Epidemiological studies have consistently observed inverse relationships between physical activity and risks of breast and colorectal cancer. However, previous epidemiological studies have generally relied on self-report measures of physical activity which are prone to recall and response biases and may attenuate 'true' associations with disease risk. More objective methods to measure physical activity, such as accelerometer, have seldom been used in large-scale epidemiological studies, with the UK Biobank being a recent exception in which -100,000 participants wore a wrist accelerometer for 7-days to measure total activity levels.

Epidemiological analyses of these data will provide important new evidence on the link between physical activity and cancer, but these analyses remain vulnerable to other biases of observational epidemiology such as residual confounding (e.g. low physical activity levels may be correlated with other unfavourable health behaviours) and reverse causality (e.g. preclinical cancer symptoms may have resulted in low physical activity levels).

Mendelian randomisation (MR) is an increasingly used tool that uses germline genetic variants as proxies (or instrumental variables) for exposures of interest to enable causal inferences to be made between a potentially modifiable exposure and an outcome. Unlike traditional observational epidemiology, MR analyses should be largely free of conventional confounding owing to the random independent assignment of alleles during meiosis. In addition, there should be no reverse causation, as germline genetic variants are fixed at conception and are consequently unaffected by the disease process.

We used a two-sample MR framework to examine potential causal associations between objective accelerometer-measured physical activity and risks of breast and colorectal cancer using genetic variants associated with accelerometer-measured physical activity identified from two recent genome-wide association studies (GWAS). We examined the associations of these genetic variants with risks of breast cancer and colorectal cancer.

**Results**

**MR estimates for breast cancer.** We estimated that a 1 standard deviation (SD) (8.14 milligravities) increment in the genetically predicted levels of accelerometer-measured physical activity was associated with a 49% lower risk of breast cancer for the instrument using the 5 genome-wide-significant SNP instrument (odds ratio [OR]: 0.51, 95% confidence interval [CI]: 0.27 to 0.98, P-value = 0.04, Q-value = 0.062) (Table 1), and a 41% lower risk for the extended 10 SNP instrument (OR: 0.59, 95% CI: 0.42 to 0.84, P-value = 0.003, Q-value = 0.012). An inverse association was only found for estrogen receptor positive breast cancer (ER+ve) (5 SNP instrument, OR: 0.45, 95% CI: 0.20 to 1.01, P-value = 0.054, Q-value = 0.077; extended 10 SNP instrument, OR: 0.53, 95% CI: 0.35 to 0.82, P-value = 0.004, Q-value = 0.004), and not estrogen receptor negative (ER-ve) breast cancer (Table 1); although this heterogeneity by subtype was not statistically different (I² = 16%; P-heterogeneity by subtype = 0.27). There was some evidence of heterogeneity based on Cochran’s Q (P-value < 0.05) for the breast cancer analyses; consequently, for these models random effects MR estimates were used (Table 1). MR estimates for each of the SNPs associated with accelerometer-measured physical activity in relation to breast cancer risk are presented in Fig. 1 and Supplementary Fig. 1. Scatter plots (with coloured lines representing the slopes of the different regression analyses) and funnel plots of the accelerometer-measured physical activity and breast cancer risk association for the extended 10 SNP instrument are presented in Supplementary Figs. 2 and 3.

**Mendelian randomisation estimates for colorectal cancer.** For colorectal cancer, a 1 SD increment in accelerometer-measured physical activity level was associated with a 34% lower risk (OR: 0.66, 95% CI: 0.59 to 1.01, P-value = 0.035, Q-value = 0.022) for the 5 SNP instrument, and a 40% lower risk for the extended 10 SNP instrument (OR: 0.60, 95% CI: 0.47 to 0.76, P-value = 2.4 × 10⁻⁵; Q-value = 0.0002) (Table 1). The inverse effect estimate was stronger for women (OR: 0.57, 95% CI: 0.36 to 0.90, P-value = 0.02, Q-value = 0.036), while there was weak evidence for an inverse association for men (OR: 0.79, 95% CI: 0.50 to 1.23, P-value = 0.29, Q-value = 0.31); this heterogeneity did not meet the threshold of significance (I² = 0%; P-heterogeneity by sex = 0.34). For colorectal subsite analyses, accelerometer-measured physical activity levels were inversely associated with risks of colon cancer (OR per 1 SD increment OR: 0.64, 95% CI: 0.44 to 0.94, P-value = 0.02, Q-value = 0.036); while there was weak evidence for an inverse association between accelerometer-measured physical activity levels and rectal cancer (OR: 0.70, 95% CI: 0.43 to 1.14, P-value = 0.15, Q-value = 0.18). Similar results by sex and subsite for colorectal cancer were found for the extended 10 SNP instrument (Table 1). MR estimates for each individual SNP associated with accelerometer-measured physical activity in relation to colorectal cancer risk are presented in Fig. 2 and Supplementary Figs. 4–6. Scatter plots (with coloured lines representing the slopes of the different regression analyses) and funnel plots of the accelerometer-measured physical activity and colorectal cancer risk association for the extended 10 SNP instrument are presented in Supplementary Figs. 7 and 8.

**Evaluation of assumptions and sensitivity analyses.** The strength of the genetic instruments denoted by the F-statistic was ≥10 for all the accelerometer-measured physical activity variants and ranged between 27 and 56 (Table 2). Little evidence of directional pleiotropy was found for all models that used the extended 10 SNP instrument (MR-Egger intercept P-values > 0.06) (Table 1). The estimates from the weighted-median approach for the extended 10 SNP instrument were consistent with those of inverse-variance weighted (IVW) models (Table 1). The MR pleiotropy residual sum and outlier test (MR-PRESSO) method identified the SNPs rs11012732 and rs55657917 contained within the extended 10 SNP instrument as pleiotropic for breast cancer, but similar magnitude associations were observed when these variants were excluded from the analyses (Supplementary Table 10). After examining Phenoscanner and GWAS catalogue, we found that several of the accelerometer-measured physical activity genetic variants were also associated with adiposity-related phenotypes (Supplementary Tables 11, 12). However, the results from the leave-one-SNP out analysis did not reveal any influential SNPs driving the associations (Supplementary Tables 13–18). Additionally, similar results were found when the 5 adiposity-related SNPs were excluded from the extended 10 SNP genetic instrument (Supplementary Table 19). Further, the results from the multivariable MR analyses adjusting for BMI using the extended 10 SNP instrument were largely consistent with these models.
unchanged from the main IVW results (Supplementary Table 20). Finally, a similar pattern of results was found when GWAS effect estimates adjusted for BMI were used for 5 SNP genetic instrument (Supplementary Table 21).

Discussion

In this MR analysis, higher levels of genetically predicted accelerometer-measured physical activity were associated with lower risks of breast cancer and colorectal cancer, with similar magnitude inverse associations found for ER<sup>+</sup>ve and for colon cancer. These findings indicate that population-level increases in physical activity may lower the incidence of these two commonly diagnosed cancers, and support the promotion of physical activity for cancer prevention.

A large body of observational studies has investigated how physical activity relates to risk of breast and colorectal cancer. In a participant-level pooled analysis of 12 prospective studies, when the 90th and 10th percentile of leisure-time physical activity were compared, lower risks of breast cancer (hazard ratio [HR]: 0.90, 95% CI: 0.87 to 0.93), colon cancer (HR: 0.84, 95% CI: 0.77 to 0.91), and rectal cancer (HR: 0.87, 95% CI: 0.80 to 0.95) were found. Similarly, inverse associations between total physical activity and risks of postmenopausal breast and colorectal cancer were recently reported in meta-analyses of all published prospective cohort data by the WCRF/AICR Continuous Update Project.

These observational studies relied on self-report physical activity assessment methods that are prone to measurement error, which may attenuate associations towards the null. In addition, causality cannot be ascertained from such observational analyses as they are vulnerable to residual confounding and reverse causality. Further, logistical and financial challenges prohibit

---

### Table 1 Mendelian Randomisation estimates between accelerometer-measured physical activity and cancer risk.

| Methods | No. Cases | Genom-wide significant SNPs (n = 5) from the GWAS by Doherty et al. | Extended number of SNPs (n = 10) from the GWAS by Klimentidis et al. |
|---------|-----------|---------------------------------------------------------------|---------------------------------------------------------------|
|         |           | Estimates (OR)<sup>a</sup> 95% CI P-value Q-value | Estimates (OR)<sup>b</sup> 95% CI P-value Q-value |
|         |           | for pleiotropy<sup>c</sup> or heterogeneity<sup>d</sup> | for pleiotropy<sup>c</sup> or heterogeneity<sup>d</sup> |
| Breast cancer | Inverse-variance weighted<sup>e</sup> | 122,977 0.51 0.27, 0.98 0.04 0.062 4.4 × 10<sup>-8</sup> | 0.59 0.42, 0.84 0.003 0.012 6.8 × 10<sup>-7</sup> |
| MR-Egger | 0.01 0.00, 2.01 0.09 0.16 | 0.55 0.09, 3.20 0.5 0.9 | 0.76 0.59, 0.98 0.03 |
| Weighted median | 0.61 0.42, 0.87 0.006 | 0.61 0.48, 0.90 0.008 | 0.66 0.48, 0.90 0.008 |
| ER<sup>v+w</sup> subset | Inverse-variance weighted<sup>e</sup> | 69,501 0.45 0.20, 1.01 0.054 0.077 8.5 × 10<sup>-9</sup> | 0.53 0.35, 0.82 0.004 0.004 3.1 × 10<sup>-7</sup> |
| MR-Egger | 0.03 0.00, 4.0 0.34 0.46 | 0.61 0.07, 5.26 0.65 0.9 | 0.66 0.48, 0.90 0.008 |
| Weighted median | 0.55 0.35, 0.85 0.008 | 0.61 0.48, 0.90 0.008 | 0.66 0.48, 0.90 0.008 |
| Inverse-variance weighted<sup>e</sup> | 21,468 0.95 0.44, 2.04 0.89 0.89 0.002 | 0.78 0.51, 1.22 0.27 0.3 0.01 | 0.78 0.51, 1.22 0.27 0.3 0.01 |
| MR-Egger | 0.01 0.00, 4.48 0.15 0.15 | 0.24 0.03, 1.81 0.17 0.24 | 0.7 0.47, 1.04 0.08 |
| Weighted median | 0.84 0.47, 1.47 0.53 | 0.6 0.47, 0.76 2.4 × 10<sup>-5</sup> 0.0002 0.5 | 0.6 0.47, 0.76 2.4 × 10<sup>-5</sup> 0.0002 0.5 |

**Notes:**

- CI = confidence intervals, MR = Mendelian randomisation, OR = odds ratio, SNPs = Single nucleotide polymorphisms
- The estimates correspond to a standard deviation increase in physical activity
- P-value for heterogeneity based on Q statistic
- P-value for pleiotropy based on MR-Egger intercept
- P-value for heterogeneity based on Q statistic
- The estimates were derived from a random effects model due to the presence of heterogeneity based on Cochran's Q statistic

---

In a participant-level pooled analysis of 12 prospective studies, when the 90th and 10th percentile of leisure-time physical activity were compared, lower risks of breast cancer (hazard ratio [HR]: 0.90, 95% CI: 0.87 to 0.93), colon cancer (HR: 0.84, 95% CI: 0.77 to 0.91), and rectal cancer (HR: 0.87, 95% CI: 0.80 to 0.95) were found. Similarly, inverse associations between total physical activity and risks of postmenopausal breast and colorectal cancer were recently reported in meta-analyses of all published prospective cohort data by the WCRF/AICR Continuous Update Project.

These observational studies relied on self-report physical activity assessment methods that are prone to measurement error, which may attenuate associations towards the null. In addition, causality cannot be ascertained from such observational analyses as they are vulnerable to residual confounding and reverse causality. Further, logistical and financial challenges prohibit
randomised controlled trials of physical activity and cancer development. For example, it has been estimated that in order to detect a 20% breast cancer risk reduction, between 26,000 to 36,000 healthy middle-aged women would need to be randomised to a 5 year exercise intervention. Several trials on cancer survivors are registered and underway, and these may provide evidence of potential causal associations between physical activity and disease free survival and cancer recurrence; however, these interventions will not inform causal inference of the relationship between physical activity and cancer development. We therefore conducted MR analyses to allow causal inference between accelerometer-measured physical activity and risks of developing breast and colorectal cancer. The inverse associations we found were stronger for ER+ breast cancer and colon cancer, and are highly concordant with prior observational epidemiological evidence.

There is currently no standard method in translating accelerometer data into energy expenditure values, such as metabolic equivalent of tasks (METs). However, using an accepted threshold for moderate activity (e.g. fast walking) of 100 milli-gravity, 1-SD higher mean acceleration (~8 milli-gravity) equates to approximately 50 min extra moderate activity per week. Similarly, using an accepted threshold of 425 milli-gravity for vigorous activity (e.g. running), a 1-SD higher mean acceleration equates to approximately 8 min of extra vigorous activity per week. In our study, we found that such an increase in weekly activity translates to a 49 and 34% lower risks of developing breast and colorectal cancer, respectively.

Being physically active is associated with less weight gain and body fatness, and lower adiposity is associated with lower risks of breast and colorectal cancer. Since body size/adiposity is likely on the causal pathway linking physical activity and breast and colorectal cancer, it is challenging to disentangle independent effects of physical activity on cancer development. The close inter-relation between adiposity and physical activity is evident from 5 of the 10 SNPs in the extended genetic instrument for accelerometer-measured physical activity being previously associated with adiposity/body size traits. However, it is noteworthy that our results were unchanged when we excluded adiposity-related SNPs from this genetic instrument, and when we conducted multivariable MR analyses adjusting for body mass index (BMI). These results would therefore suggest that physical activity is also associated with breast and colorectal cancer independently of adiposity.

Multiple biological mechanisms are hypothesised to mediate the potential beneficial role of physical activity on cancer development. Greater physical activity has been associated with lower circulating levels of insulin and insulin-like growth factors, which promote cellular proliferation in breast and colorectal tissue and have also been linked to development of cancers at these sites. Higher levels of physical activity have also been associated with lower circulating concentrations of estradiol.

![Fig. 1 Mendelian randomisation analysis for individual SNPs associated with accelerometer-measured physical activity in relation to breast cancer risk using the genetic instrument from the GWAS by Doherty et al.](https://doi.org/10.1038/s41467-020-14389-8)
estrone, and higher levels of sex hormone binding globulin\textsuperscript{28–30} which are themselves risk factors for breast cancer development\textsuperscript{31,32}. Physical activity has also been associated with improvements in the immune response with increased surveillance and elimination of cancerous cells\textsuperscript{33,34}. Higher levels of physical activity may also reduce systemic inflammation by lowering the levels of pro-inflammatory factors, such as C-reactive protein (CRP), interleukin-6 (IL-6) and tumour necrosis factor-alpha (TNF-\alpha)\textsuperscript{33,35,36}. Finally, emerging evidence suggests that the gut microbiome may play an important role in the physical activity and cancer relationship. Dysbiosis of the gut microbiome has been associated with increased risks of several malignancies, including breast and colorectal cancer\textsuperscript{37}. Changes in gut microbiome composition and derived metabolic products have been found following endurance exercise training with short-chain fatty acid concentrations increased in lean, but not obese, subjects\textsuperscript{38,39}.

A fundamental assumption of MR is that the genetic variants do not influence the outcome via a different biological pathway from the exposure of interest (horizontal pleiotropy). We conducted multiple sensitivity analyses using an extended 10 SNP genetic instrument for accelerometer-measured physical activity to test for the influence of pleiotropy on our causal estimates, and our results were robust according to these various tests. A potential limitation of our analysis is that the genetic variants explained a small fraction of the variability of accelerometer-measured physical activity, which may have resulted in some of the breast cancer subtype and colorectal subsite analyses being underpowered. In addition, our use of summary-level data precluded subgroup analyses by other cancer risk factors (e.g. BMI, exogenous hormone use). We were also unable to stratify breast cancer analyses by menopausal status; however, the majority of women in the source GWAS had postmenopausal breast cancer\textsuperscript{13}. Finally, 7-day accelerometer-measured physical activity levels of UK Biobank participants may not have been representative of usual behavioural patterns.

In conclusion, we found that genetically elevated levels of accelerometer-measured physical activity were associated with lower risks of breast and colorectal cancer. These findings strongly support the promotion of physical activity as an effective strategy in the primary prevention of these commonly diagnosed cancers.

**Methods**

**Data on physical activity.** Summary-level data were obtained from two recently published GWAS on accelerometer-measured physical activity conducted in ~91,000 participants from the UK Biobank\textsuperscript{11,12}. In the GWAS by Doherty et al.,\textsuperscript{11} BOLT-LMM was used to perform linear mixed models analyses that were adjusted for assessment centre, genotyping array, age, age\textsuperscript{2}, and season. This GWAS identified 5 genome-wide-significant SNPs (\(P\)-value < 5 × 10\textsuperscript{−8}) associated with accelerometer-measured physical activity. The estimated SNP-based heritability for accelerometer-measured physical activity in the UK Biobank is 14\%\textsuperscript{12}, suggesting that additional SNPs contributed to its variation. Consequently, we also used an accelerometer-measured physical activity instrument with an expanded number
Table 2 Summary information on accelerometer-measured physical activity SNPs used as genetic instruments for the Mendelian randomisation analyses.

| SNP          | Effect allele | Baseline allele | Chr | Positiona | Gene                        | EAF  | beta PAb | se PA | Nc | R2 | F-statistic |
|--------------|---------------|-----------------|-----|-----------|-----------------------------|------|---------|-------|----|----|-------------|
| rs6775319    | A             | T               | 3   | 18717009  | SATB1-AS1                   | 0.27 | 0.03    | 0.005 | 91,105 | 0.0003 | 27          |
| rs6895232    | T             | A               | 5   | 152659861 | LINC01470                   | 0.66 | 0.03    | 0.005 | 91,105 | 0.0003 | 30          |
| rs564819152  | A             | G               | 10  | 21531721  | SKID1A                      | 0.68 | 0.03    | 0.005 | 91,105 | 0.0003 | 31          |
| rs2696625    | G             | A               | 17  | 46249498  | KANS1L-AS1                  | 0.23 | 0.04    | 0.005 | 91,105 | 0.0004 | 44          |
| rs59499656   | T             | A               | 18  | 43188344  | RT2/YT4                    | 0.35 | 0.03    | 0.005 | 91,105 | 0.0004 | 32          |
| rs12045968   | G             | T               | 1   | 33225097  | ZNF362                      | 0.22 | 0.24    | 0.044 | 91,084 | 0.0003 | 30          |
| rs34174329    | C             | A               | 1   | 77984833  | DNAJB4                      | 0.91 | 0.31    | 0.056 | 91,084 | 0.0003 | 30          |
| rs6775319    | A             | T               | 3   | 18717009  | LOCI05376796                | 0.3  | 0.23    | 0.041 | 91,084 | 0.0003 | 30          |
| rs1252261    | G             | A               | 5   | 15267526  | LINC01470                   | 0.67 | 0.21    | 0.038 | 91,084 | 0.0003 | 31          |
| rs9293503    | T             | C               | 5   | 88653144  | LINC00461                   | 0.88 | 0.33    | 0.059 | 91,084 | 0.0003 | 31          |
| rs11012732   | A             | G               | 10  | 21541175  | MLLT10                      | 0.65 | 0.23    | 0.039 | 91,084 | 0.0004 | 33          |
| rs148193266  | C             | A               | 11  | 104657953 | RP11-681H10.1               | 0.02 | 0.51    | 0.092 | 91,105 | 0.0003 | 31          |
| rs1550435    | A             | C               | 11  | 74039044  | PML                         | 0.53 | 0.2     | 0.037 | 91,084 | 0.0003 | 32          |
| rs55657917   | G             | T               | 17  | 45767194  | CRHR1                       | 0.22 | 0.3     | 0.04  | 91,084 | 0.0006 | 56          |
| rs59499656   | T             | A               | 18  | 43188344  | RT2/YT4                    | 0.34 | 0.23    | 0.038 | 91,084 | 0.0004 | 36          |

BM1 body mass index, Chr chromosome, EAF effect allele frequency, NA not available, PA physical activity, se standard error, SNP single nucleotide polymorphism

aPosition based on GRCh38.p12
bThe beta coefficients are expressed in milliglycories

N refers to the sample size of the initial GWAS from which the genetic variants were selected

The heterogeneity of causal effects by cancer subtype and sex was investigated by estimating the I2 statistic assuming a fixed-effects model.49

For causal estimates from MR studies to be valid, three main assumptions must be met: 1) the genetic instrument is strongly associated with the level of accelerometer-measured physical activity; 2) the genetic instrument is not associated with any potential confounder of the physical activity—cancer association; and 3) the genetic instrument does not affect cancer independently of physical activity (i.e. horizontal pleiotropy should not be present).49 The strength of each instrumental effect was calculated using the F-statistic using the following formula: \( F = (N - 2)/\left(1 - R^2\right) \), where \( R^2 \) is the proportion of the variability of the physical activity explained by each instrument and N the sample size of the GWAS for the SNP-physical activity association.49 To calculate \( R^2 \) for the 5-gene-wide significant SNP instrument we used the following formula: \( 2 \times EAF \times (1 - EAF) \times \beta^2 \); whereas for the extended 10 SNP instrument we used \( 2 \times EAF \times (1 - EAF) \times \beta^2 / (2 \times EAF \times (1 - EAF) \times \beta^2 + 2 \times EAF \times (1 - EAF) \times N \times SE(\beta^2)) \), where EAF is the effect allele frequency, \( \beta \) is the estimated genetic effect on physical activity, N is the sample size of the GWAS for the SNP-physical activity association and SE(\( \beta^2 \)) is the standard error of the genetic effect.49, FDR correction (Q-value) was performed using the Benjamini–Hochberg method.67

Sensitivity analyses. Several sensitivity analyses were used to check and correct for the presence of pleiotropy in the causal estimates. Cochran’s Q was computed to quantify heterogeneity across the individual causal effects, with a P-value ≤0.05 indicating the presence of pleiotropy, and that consequently, a random effects IVW MR analysis should be used.58,59 We also assessed the potential presence of horizontal pleiotropy using MR-Egger regression based on its intercept term, where deviation from zero denotes the presence of directional pleiotropy. Additionally, the slope of the MR-Egger regression provides valid MR estimates in the presence of horizontal pleiotropy when the pleotropic effects of the genetic variants are independent from the genetic associations with the exposure.60,66 We also computed OR estimates using the complementary weighted-median method that can give valid MR estimates under the presence of horizontal pleiotropy when up to 50% of the included instruments are invalid.64 The presence of pleiotropy was also assessed using the MR-PRESSO. In this, outlying SNPs are excluded from the accelerometer-measured physical activity instrument and the effect estimates are reassessed.41 For all of the aforementioned sensitivity analyses to identify possible pleiotropy, we considered the estimates from the extended 10 SNP instrument as the primary results due to unstable estimates from the 5 SNP instrument. A leave-one-SNP out analysis was also conducted to assess the influence of individual variants on the observed associations. We also examined the selected genetic instruments and their proxies (r2 > 0.8) and their associations with secondary phenotypes (P-value < 5 × 10−8) in PhenoScanner (http://www.phenoscaner.msdl.scm.ac.uk) and GWAS catalog (date checked April 2019). For the extended 10 SNP instrument, we also conducted multivariable MR analyses to adjust for potential pleiotropy due to BMI because the initial GWAS on physical activity reported several strong associations (P-value < 10−5) between the identified SNPs and BMI.52 The new estimates correspond to the direct causal effect of physical activity with the BMI being fixed. The genetic data on BMI were
obtained from a GWAS study published by The Genetic Investigation of ANthropometric Traits (GIANT) consortium (Supplementary Table 9). Additionally, for the extended 10 SNP instrument, we also conducted analyses with adiposity-related SNPs (i.e., those previously associated with BMI, waist circumference, weight, or body/trunk fat percentage in GWAS studies at rsP-value < 10-8) excluded (n = 5; rs84517439, rs6775319, rs11012732, rs1550435, rs99499656). Finally, we conducted two-sample MR analyses using BMI adjusted GWAS estimates for the 5 SNP accelerometer-measured physical activity instrument. However, the MR results using the BMI adjusted GWAS estimates should be interpreted cautiously due to the potential for collider bias. All the analyses were conducted using the MendelianRandomisation and TwoSampleMR packages, and the R programming language.

### References

1. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: Cancer J. Clinicians 68, 394–424 (2018).
2. WHO. Global Status Report on Noncommunicable Diseases 2014 (WHO, 2014).
3. Moore, S. C. et al. Association of leisure-time physical activity with risk of 26 types of cancer in 1.44 million adults. JAMA Intern. Med. 176, 816–825 (2016).
4. World Cancer Research Fund/American Institute for Cancer Research, 2018.
5. Prince, S. A. et al. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int. J. Behav. Nutr. Phys. Act. 5, 56 (2008).
6. Doherty, A. et al. Large scale population assessment of physical activity using wrist worn accelerometers: The UK Biobank Study. PloS ONE 12, e0169649 (2017).
7. Davey Smith, G. & Ebrahim, S. ‘Mendelian randomisation’: can genetic epidemiology contribute to understanding environmental determinants of disease? Int. J. Epidemiol. 32, 1–22 (2003).
8. Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Davey Smith, G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Stat. Med. 27, 1133–1163 (2008).
9. Doherty, A. et al. GWAS identifies 14 loci for device-measured physical activity and sleep duration. Nat. Commun. 9, 5257 (2018).
10. Klimentidis, Y. C. et al. Genome-wide association study of habitual physical activity in over 277,000 UK Biobank participants identifies novel variants and genetic correlations with chromosome and obesity-related traits. bioRxiv https://doi.org/10.1101/179317 (2017).
11. Michailidou, K. et al. Association analysis identifies 65 new breast cancer risk loci. Nature 551, 92 https://www.nature.com/articles/nature24284#supplementary-information (2017).
12. Huzyge, J. R. et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nat. Genet. 51, 76–87 (2019).
13. WCRF-AICR. Physical Activity and the Risk of Cancer (World Cancer Research Fund/American Institute for Cancer Research, 2018).
14. Helmink, B. A., Khan, M. A. W., Hermann, A., Gopalakrishnan, V. & Wargo, J. A. The microbiome, cancer, and cancer therapy. Nat. Med. 25, 377–388 (2019).
15. Fernandez, D. M., Clemente, J. C. & Giannarelli, C. Physical activity, immune system, and the microbiome in cardiovascular disease. Front Physiol. 9, 763–763 (2018).
16. Allen, J. M. et al. Exercise alters gut microbiota composition and function in lean and obese humans. Med. Sci. Sports Exerc. 50, 747–757 (2018).
17. Amy, J. et al. Assessment of bidirectional relationships between physical activity and depression among adults: a 2-sample Mendelian randomization study. JAMA Psychiatry 76, 399–408 (2019).
18. Michailidou, K. et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat. Genet. 45, 353 (2013).
19. Brion, M.-J. et al., Shakibzadov, K. & Visscher, P. M. Calculating statistical power in Mendelian randomization studies. Int. J. Epidemiol. 42, 1497–1501 (2013).
20. Higgins, J. P. T., Thompson, S. G., Deeks, J. J. & Altman, D. G. Measuring inconsistency in meta-analyses. BMJ 327, 557–560 (2003).
21. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet. Epidemiol. 40, 304–314 (2016).
22. Burgess, S., Thompson, S. G. & Collaboration, C. C. G. Avoiding bias from weak instruments in Mendelian randomization studies. Int. J. Epidemiol. 40, 755–764 (2011).

**Data availability**

Data supporting the findings of this study are available within the paper and its supplementary information files.

**Received:** 23 August 2019; **Accepted:** 28 December 2019;
**Published online:** 30 January 2020

**NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-020-14389-8**

---

**References**

1. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: Cancer J. Clinicians 68, 394–424 (2018).
2. WHO. Global Status Report on Noncommunicable Diseases 2014 (WHO, 2014).
3. Moore, S. C. et al. Association of leisure-time physical activity with risk of 26 types of cancer in 1.44 million adults. JAMA Intern. Med. 176, 816–825 (2016).
4. World Cancer Research Fund/American Institute for Cancer Research, 2018.
5. Prince, S. A. et al. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int. J. Behav. Nutr. Phys. Act. 5, 56 (2008).
6. Doherty, A. et al. Large scale population assessment of physical activity using wrist worn accelerometers: The UK Biobank Study. PloS ONE 12, e0169649 (2017).
7. Davey Smith, G. & Ebrahim, S. ‘Mendelian randomisation’: can genetic epidemiology contribute to understanding environmental determinants of disease? Int. J. Epidemiol. 32, 1–22 (2003).
8. Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Davey Smith, G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Stat. Med. 27, 1133–1163 (2008).
9. Doherty, A. et al. GWAS identifies 14 loci for device-measured physical activity and sleep duration. Nat. Commun. 9, 5257 (2018).
10. Klimentidis, Y. C. et al. Genome-wide association study of habitual physical activity in over 277,000 UK Biobank participants identifies novel variants and genetic correlations with chromosome and obesity-related traits. bioRxiv https://doi.org/10.1101/179317 (2017).
11. Michailidou, K. et al. Association analysis identifies 65 new breast cancer risk loci. Nature 551, 92 https://www.nature.com/articles/nature24284#supplementary-information (2017).
12. Huzyge, J. R. et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nat. Genet. 51, 76–87 (2019).
13. WCRF-AICR. Physical Activity and the Risk of Cancer (World Cancer Research Fund/American Institute for Cancer Research, 2018).
14. Helmink, B. A., Khan, M. A. W., Hermann, A., Gopalakrishnan, V. & Wargo, J. A. The microbiome, cancer, and cancer therapy. Nat. Med. 25, 377–388 (2019).
15. Fernandez, D. M., Clemente, J. C. & Giannarelli, C. Physical activity, immune system, and the microbiome in cardiovascular disease. Front Physiol. 9, 763–763 (2018).
16. Allen, J. M. et al. Exercise alters gut microbiota composition and function in lean and obese humans. Med. Sci. Sports Exerc. 50, 747–757 (2018).
17. Amy, J. et al. Assessment of bidirectional relationships between physical activity and depression among adults: a 2-sample Mendelian randomization study. JAMA Psychiatry 76, 399–408 (2019).
18. Michailidou, K. et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat. Genet. 45, 353 (2013).
19. Brion, M.-J. et al., Shakibzadov, K. & Visscher, P. M. Calculating statistical power in Mendelian randomization studies. Int. J. Epidemiol. 42, 1497–1501 (2013).
20. Higgins, J. P. T., Thompson, S. G., Deeks, J. J. & Altman, D. G. Measuring inconsistency in meta-analyses. BMJ 327, 557–560 (2003).
21. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet. Epidemiol. 40, 304–314 (2016).
22. Burgess, S., Thompson, S. G. & Collaboration, C. C. G. Avoiding bias from weak instruments in Mendelian randomization studies. Int. J. Epidemiol. 40, 755–764 (2011).
