Physical activity and risk of lung cancer: a two-sample Mendelian randomization study

Sebastian E Baumeister[1,2], Michael F Leitzmann[3]; Christopher I Amos [4], Rayjean J Hung[5], Cancer in Lung of the International Lung Cancer Consortium (TRICL-ILCCO) [§], Lung Cancer Cohort Consortium (LC3) [§], Alexander Teumer[6*], Hansjörg Baurecht[3*]

[1] Chair of Epidemiology, LMU München, UNIKA-T Augsburg, Augsburg, Germany

[2] Independent Research Group Clinical Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Munich, Germany

[3] Department of Epidemiology and Preventive Medicine, University of Regensburg, Germany

[4] Biomedical Data Science, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA

[5] Lunenfeld-Tanenbaum Research Institute, Sinai Health System; Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

[6] Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany

* contributed equally

[§] Collaborators of the TRICL-ILCCO and LC3 consortium are listed in this article's supplementary note

Corresponding author: Sebastian E. Baumeister, PhD, Chair of Epidemiology, Ludwig-Maximilians-Universität München, UNIKA-T Augsburg, Neusässer Str. 47, 86156 Augsburg, Germany, Email: s.baumeister@unika-t.de

Word count: main text: 1683**;** abstract: 197

ABSTRACT

Background: Evidence from observational studies of physical activity has been inconclusive regarding its association with lung cancer risk. We investigated the causal relevance of physical activity for lung cancer using two-sample Mendelian randomization (MR) analysis.

Methods: Summary data of genome-wide association studies on physical activity and lung cancer were identified using PubMed and the GWAS catalog. Twenty six single nucleotide polymorphisms (SNP) known at $P < 5x10^{-8}$ to be associated with self-reported or accelerometer-assessed physical activity served as instrumental variables.

Results: Self-reported physical activity was associated with lower risk of overall lung cancer (inverse variance weighted odds ratio [OR] per standard deviation (SD) increase in : 0.48, 95% confidence interval [CI]: 0.32-0.73, $P=5.61x10^{-4}$, false discovery rate (FDR) = $1.12x10^{-3}$), squamous cell carcinoma, small cell carcinoma, and lung cancer in ever-smokers. Accelerometer-assessed physical activity was associated with lower risk of overall lung cancer (OR per 1-SD: 0.93, 95% CI: 0.90-0.97, $P=2.96x10^{-4}$, FDR = $1.68x10^{-3}$), adenocarcinoma, lung cancer in never-smokers, and lung cancer in ever-smokers.

Conclusions: The results support a potentially causal protective relationship between physical activity and risk of lung cancer and foster the hypothesis that enhancing physical activity may be an effective prevention strategy for lung cancer.

Introduction

Lung cancer causes more preventable death than any other cancer worldwide [1]. Smoking is the risk factor most strongly linked to all lung cancer subtypes [2]. Potential non-smoking related risk factors for lung cancer include environmental carcinogens, pulmonary fibrosis, genetic history, dietary factors, and insufficient physical activity [3, 4]. Several meta-analyes of observational studies suggested an inverse association between physical activity and lung cancer risk [5-7]. Yet, the evidence has been limited to current and former smokers in most studies [5-7]. Interpretation of this inverse association has been constrainted by potential confounding, as smoking causes lung cancer and renders physical activity more difficult [5, 8]. Reverse causation may also affect the association between physical activity and lung cancer risk, as the presence of lung cancer symptoms may lead to avoidance of physical activity. Accordingly, the World Cancer Research Fund/American Institute for Cancer Research [4] and a recent umbrella review [9] have categorized the overall evidence from observational studies as inconclusive. Mendelian randomization is a method that uses genetic variants as instrumental variables to uncover causal relationships in the presence of unobserved confounding and reverse causation [10]. Previous Mendelian randomization studies have established obesity, a variable closely related to physical activity, as a contributor to the development of lung cancer [11]. In the current study, we performed two-sample Mendelian randomization analyses to assess the causal association between physical activity and lung cancer.

Materials and Methods

The study design had three components: (1) identification of genetic variants to serve as instrumental variables for physical activity; (2) the acquisition of summary data for the genetic instruments from genomewide association studies on physical activity; (3) acquisition of instrumenting SNP-outcome summary data from genome wide association studies of lung cancer.

Selection of instrumental variables for physical activity

A genome-wide study of 377,234 UK Biobank participants identified 18 SNPs associated with self-reported moderate-to-vigorous physical activity (metabolic equivalents per week) at genome-wide significance ($P < 5 \times 10^{-8}$) [12], with an estimated heritability of 5% (Supplementary Table 1). In addition to those 18 SNPs, we used 8 SNPs associated with accelerometer-based physical activity (mean acceleration in milli-gravities) at genome-wide significance in 91,084 individuals [12]. The estimated SNP-based heritability for accelerometer-based physical activity was 14%.

Lung cancer data

Summary data for the association of the self-reported SNPs and the accelerometer SNPs with lung cancer (overall, adenocarcinoma, squamous cell carcinoma, small cell carcinoma, never-smokers, ever-smokers) were obtained from a GWAS on 29,266 cases and 56,450 controls [13]. That GWAS for lung cancer did not include the UK Biobank.

Statistical power

The a priori statistical power was calculated using an online tool at http://cnsgenomics.com/shiny/mRnd/ [14]. We assumed that the 18 self-reported physical activity SNPs explained 0.1% and the 8 accelerometer-based physical activity SNPs explained 0.4% of the phenotypic variable [12, 15, 16]. Given a type 1 error of 5%, we had sufficient statistical power (>80%) when the expected odds ratios (OR) per 1 standard deviation (SD) for overall lung cancer was ≤0.80 and ≤0.91 in genetically instrumented self-reported physical activity and accelerometer-based physical activity, respectively (Supplementary Table 2).

Statistical analyses

The principal analysis was conducted using the inverse-variance weighted (IVW) method, under a fixed effects model, alongside other methods to adress the violations of specific instrumental variable analysis assumptions: weighted median, MR-Egger and MR-Pleiotropy

RESidual Sum and Outlier (MR-PRESSO) [17, 18]. The results were presented as ORs and 95% confidence intervals (CIs) per 1-SD increment in self-reported moderate-to-vigorous physical activity or accelerometer-based physical activity. We tested potential directional pleiotropy by testing the intercepts of MR-Egger models [17]. Finally, we looked up each instrument SNP and its proxies (r^2>0.8) in Phenoscanncer [19] and the GWAS catalog [20] to assess any previous associations ($P<1x10^{-8}$) with potential confounders. We performed leave-one-out analyses and exlusion of potentially pleiotropic SNPs to rule out possible pleiotropic effects. To correct for multiple testing, we additionally reported the false discovery rate (FDR) according to Benjamini-Hochberg. Analyses were performed using the TwoSampleMR (version 0.4.25) [18] and MRPRESSO (version 1.0) packages in R (version 3.6.1). Reporting follows the STROBE-MR statement [21].

Results

We estimated that a 1-SD increment in genetically predicted self-reported physical activity was associated with a 52% (OR: 0.48, 95% CI: 0.32-0.73 $P=5.61x10^{-4}$, FDR=$1.12x10^{-3}$) lower risk of overall lung cancer (Table 1). Inverse associations for genetically predicted self-reported physical activity were also found for squamous cell carcinoma, small cell and lung cancer in ever-smokers. No associations were found for genetically predicted self-reported physical activity and adenocarcinoma or lung cancer in never-smokers. By comparison, genetically predicted accelerometer-based physical activity was inversely related to overall lung cancer (OR per 1-SD: 0.93, 95% CI: 0.90-0.97, $P=2.96x10^{-4}$, FDR=$1.68x10^{-3}$), adenocarcinoma, lung cancer in never-smokers, lung cancer in ever-smokers but showed no associations with squamous cell carcinoma or small cell carcinoma (Table 2).

The F-statistics for the strength of the genetic instruments were all ≥10 and ranged from 33 to 51 (Supplementary Table 1). The intercept test from the MR-Egger regression was statistically significant in the analysis for self-reported physical activity with overall lung cancer, squamous cell carcinoma, and lung cancer in ever smokers (Supplementary Table 3). In the Phenoscanner and GWAS databases, we identified two of the nine SNPs for self-reported

physical activity associated with cognitive traits (rs1043595) and blood lipid, blood protein, inflammation and cognitive decline (rs429358). Removing both SNPs in leave-one out analyses did not change the pattern of the results (Supplementary Table 4). However, removal of rs2854277 in leave one-SNP-out analyses attenuated associations of self-reported physical activity with overall lung cancer, adenocardinoma, and lung cancer in never-smokers, which indicated potential pleiotropy. MR-PRESSO detected one outlier SNP (rs429358) for the association between self-reported physical activity with overall lung cancer, squamous cell carcinoma, small cell carcinoma and lung cancer among ever smokers but MR estimates remained unaltered after removal of this outlier (Supplementary Table 4).

Discussion

The current study found that higher levels of genetically predicted self-reported and accelerometrer-based physical activity were associated with lower risks of overall lung cancer, several lung cancer subtypes, lung cancer in never- smokers, and lung cancer in ever-smokers. These findings are consistent with previous observational studies showing an inverse association between physical activity and risk of lung cancer. A large pooled analysis of 12 European and US cohort studies including 19,133 lung cancers reported a relative risk reduction of 24% (hazard ratio: 0.76; 95% CI: 0.71-0.77) comparing high and low levels of self-reported physical activity [22]. The most comprehensive meta-analysis comprising 20 cohort studies and 31,807 cases found a 17% relative reduction in lung cancer risk with highest versus lowest levels of physical activity (hazard ratio: 0.83; 95% CI: 0.77-0.90) [7]. Findings of another meta-analysis suggest no heterogeneity between lung cancer histologic subtypes [5]. Of note, the above-mentioned pooled analysis revealed an inverse associations in current- and former smokers and a null association in never-smokers [22]. Similarly, meta-analyses consistently found that physical activity was inversely associated with lung cancer among former and current smokers but unrelated to lung cancer among never smokers, although the number of study estimates among never smokers was limited [5-7].

The pooled analysis [22] and the meta-analyses [5-7] relied on observational study designs that are prone to uncontrolled confounding and reverse causation. By comparision, the current study used MR, which is less susceptible to some of the biases inherent in conventional observational studies. The use of two-sample MR enabled us to use the largest GWAs on lung cancer [13] to date. Our MR study also incorporated the largest GWAs on physical activity to increase the precision of SNP-physical activity estimates, to reduce the potential for weak instrument bias and to increase statistical power. However, our study also had certain limitations. First, the genetic instruments for self-reported physical activity and accelerometer-assessed physical activity explained a small fraction of the phenotypic variability, which resulted in some of the subgroup analyses being underpowered. Second, for the two-sample MR to provide unbiased estimates, the risk factor and outcome sample should come from the same underlying population. The discovery genome-wide association study of physical activity consisted of UK Biobank participants of European descent, aged 40 to 70 years [12]. The SNP-lung cancer associations were derived from cohort and case-control studies of men and women of European descent aged 18 years and older [13]. Given the limited age range of the UK Biobank and inclusion of European ancestry individuals only, our results may not be generalizable to other age groups or ancestral populations. Therefore, replication of our findings in other age groups and non-European populations is warranted. Third, we found significant effects on lung cancer risk for genetically predicted self-report physically activity. This is surprising given that self-report measures is prone to recall and response bias, while more objective methods to measure physical activity are less susceptible to these biases [23]. In contrast, previous MR studies on depression [15] and breast and endometrial cancers [24] discovered that the findings on the relationship with physical activity were specific to objectively measured, but not self-reported, physical activity. One possible explanation for this finding is that some of the genetic loci for self-reported physical activity are also related to cognitive performance, which might have introduced measurement bias [25]. Another explanation is that some of the SNPs for self-reported physical activity are also associated with blood lipids, inflammatory markers and waist-hip ratio [25], which might have introduced hori-

zontal pleiotropy. Nevertheless, it is reassuring that we also found association between accelerometer-based physical activity and lung cancers.

There was several potential biological explanations for the the inverse association between physical activity and lung cancer. First, physical activity increases pulmonary function [26]. A second explanation may be that regular physical activity lowers lowers biomarker levels associated with lung cancer (estradiol, estrone, C-reactive protein, oxidative stress, interleukin-6, tumor necrosis factor-α) [6, 7]. Lastly, physical activity may improve immune response and reverse DNA damage [27]. In conclusion, we found evidence that physical activity may be causally and inversely related to lung cancer risk. Further studies are warrented to replicate this association in other ethnic groups, to elucidate the mechanisms by which physical activity may prevent to lung cancer, and investigate the potential to intervene to reduce lung cancer risk.

Conflicts of Interest Disclosures: All authors disclose no conflict.

Funding/Support: Transdisciplinary Research for Cancer in Lung (TRICL) of the International Lung Cancer Consortium (ILCCO) was supported by grants U19-CA148127 and CA148127S1. ILCCO data harmonization is supported by the Cancer Care Ontario Research Chair of Population Studies to R.J.H. and the Lunenfeld-Tanenbaum Research Institute, Sinai Health System. The TRICL-ILCCO OncoArray was supported by in-kind genotyping by the Centre for Inherited Disease Research (26820120008i-0-26800068-1).

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

References

1. Bray F, Ferlay J, Soerjomataram I*, et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68(6):394-424.
2. Islami F, Goding Sauer A, Miller KD*, et al.* Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. CA: a cancer journal for clinicians 2018;68(1):31-54.

3. Rivera GA, Wakelee H. Lung Cancer in Never Smokers. Adv Exp Med Biol 2016;893:43-57.
4. World Cancer Research Fund International, American Insitute for Cancer Research. Diet, nutrition, physical activity and cancer: a global perspective. third expert report. In; 2018.
5. Schmid D, Ricci C, Behrens G*, et al.* Does smoking influence the physical activity and lung cancer relation? A systematic review and meta-analysis. Eur J Epidemiol 2016;31(12):1173-1190.
6. Brenner DR, Yannitsos DH, Farris MS*, et al.* Leisure-time physical activity and lung cancer risk: A systematic review and meta-analysis. Lung Cancer 2016;95:17-27.
7. Liu Y, Li Y, Bai YP*, et al.* Association Between Physical Activity and Lower Risk of Lung Cancer: A Meta-Analysis of Cohort Studies. Front Oncol 2019;9:5.
8. Samet JM. Lung Cancer, Smoking, and Obesity: It's Complicated. J Natl Cancer Inst 2018;110(8):795-796.
9. Rezende LFM, Sa TH, Markozannes G*, et al.* Physical activity and cancer: an umbrella review of the literature including 22 major anatomical sites and 770 000 cancer cases. Br J Sports Med 2018;52(13):826-833.
10. Burgess S, Foley CN, Zuber V. Inferring Causal Relationships Between Risk Factors and Outcomes from Genome-Wide Association Study Data. Annu Rev Genomics Hum Genet 2018;19:303-327.
11. Carreras-Torres R, Johansson M, Haycock PC*, et al.* Obesity, metabolic factors and risk of different histological types of lung cancer: A Mendelian randomization study. PLoS One 2017;12(6):e0177875.
12. Klimentidis YC, Raichlen DA, Bea J*, et al.* Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including CADM2 and APOE. Int J Obes (Lond) 2018;42(6):1161-1176.
13. McKay JD, Hung RJ, Han Y*, et al.* Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. Nat Genet 2017;49(7):1126-1132.
14. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. Int J Epidemiol 2013;42(5):1497-501.
15. Choi KW, Chen CY, Stein MB*, et al.* Assessment of Bidirectional Relationships Between Physical Activity and Depression Among Adults: A 2-Sample Mendelian Randomization Study. JAMA Psychiatry 2019.
16. Papadimitriou N, Dimou N, Tsilidis KK*, et al.* Physical activity and risks of breast and colorectal cancer: A Mendelian randomization analysis. bioRxiv 2019:762484.
17. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. Hum Mol Genet 2018;27(R2):R195-r208.
18. Hemani G, Zheng J, Elsworth B*, et al.* The MR-Base platform supports systematic causal inference across the human phenome. Elife 2018;7.
19. Kamat MA, Blackshaw JA, Young R*, et al.* PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. Bioinformatics 2019; 10.1093/bioinformatics/btz469.
20. Buniello A, MacArthur JAL, Cerezo M*, et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 2019;47(D1):D1005-d1012.
21. Smith GD, Davies NM, Dimou N*, et al.* STROBE-MR: Guidelines for strengthening the reporting of Mendelian randomization studies. In: PeerJ Preprints; 2019.
22. Moore SC, Lee IM, Weiderpass E*, et al.* Association of Leisure-Time Physical Activity With Risk of 26 Types of Cancer in 1.44 Million Adults. JAMA Intern Med 2016;176(6):816-25.
23. Dowd KP, Szeklicki R, Minetto MA*, et al.* A systematic literature review of reviews on techniques for physical activity measurement in adults: a DEDIPAC study. Int J Behav Nutr Phys Act 2018;15(1):15.
24. Baurecht H, Leitzmann M, O'Mara T*, et al.* Physical activity and risk of breast and endometrial cancers: a Mendelian randomization study. medRxiv 2019:19005892.

25. Folley S, Zhou A, Hypponen E. Information bias in measures of self-reported physical activity. Int J Obes (Lond) 2018;42(12):2062-2063.
26. Emaus A, Thune I. Physical activity and lung cancer prevention. Recent Results Cancer Res 2011;186:101-33.
27. Hojman P. Exercise protects from cancer through regulation of immune function and inflammation. Biochem Soc Trans 2017;45(4):905-11.

Table 1 Mendelian randomization estimates between self-reported physical activity and cancer risk

	Method	OR[a]	(95% CI)[a]	P value	FDR
Overall lung cancer					
	Inverse-variance weighted (fixed effects)	0.48	0.32; 0.73	5.61×10^{-4}	1.12×10^{-3}
	Weighted median	0.87	0.46; 1.60	0.648	0.876
	MR PRESSO	0.75	0.46; 1.20	0.244	0.353
Adenocarcinoma	Inverse-variance weighted (fixed effects)	0.62	0.36; 1.09	0.096	0.121
	Weighted median	0.61	0.27; 1.39	0.236	0.876
	MR PRESSO	0.82	0.42; 1.59	0.563	0.543
Squamous cell carcinoma	Inverse-variance weighted (fixed effects)	0.28	0.15; 0.55	1.57×10^{-4}	0.93×10^{-4}
	Weighted median	0.55	0.20; 1.54	0.257	0.859
	MR PRESSO	0.52	0.23; 1.19	0.142	
Small cell carcinoma	Inverse-variance weighted (fixed effects)	0.21	0.07; 0.62	4.51×10^{-3}	6.77×10^{-2}
	Weighted median	0.40	0.07; 2.20	0.294	0.859
	MR PRESSO	0.21	0.05; 0.92	0.045	0.053
Never smoker	Inverse-variance weighted (fixed effects)	0.98	0.27; 3.22	0.906	0.906
	Weighted median	1.14	0.18; 7.21	0.891	0.890
	MR PRESSO	0.92	0.21; 4.20	0.923	0.924
Ever smoker	Inverse-variance weighted (fixed effects)	0.49	0.29; 0.82	6.73×10^{-3}	8.08×10^{-3}
	Weighted median	0.88	0.42; 1.84	0.728	0.876
	MR PRESSO	0.77	0.47; 1.24	0.294	0.354

MR PRESSO, MR Pleiotropy RESidual Sum and Outlier. OR (odds ratio) per one standard increase in metabolic-equivalent (MET)-minutes/week. CI, confidence interval. SNP, single-nucleotide polymorphism. FDR, false discovery rate

.

Table 2 Mendelian randomization estimates between accelerometer-based physical activity and lung cancer

	Method	OR [a]	95% CI	P value	FDR
Overall lung cancer					
	Inverse-variance weighted (fixed effects)	0.93	0.90; 0.97	2.96×10^{-4}	1.68×10^{-3}
	Weighted median	0.99	0.93; 1.05	0.682	0.957
	MR PRESSO	0.75	0.46; 1.20	0.244	0.411
Adenocarcinoma	Inverse-variance weighted (fixed effects)	0.89	0.84; 0.94	1.48×10^{-5}	8.93×10^{-4}
	Weighted median	0.92	0.84; 0.99	3.91×10^{-2}	0.112
	MR PRESSO	0.92	0.85, 1.00	0.102	0.308
Squamous cell carcinoma	Inverse-variance weighted (fixed effects)	0.96	0.90; 1.02	0.163	0.195
	Weighted median	1.00	0.91; 1.09	0.958	0.957
	MR PRESSO	0.99	0.93; 1.06	0.795	0.795
Small cell carcinoma	Inverse-variance weighted (fixed effects)	0.98	0.88; 1.08	0.647	0.647
	Weighted median	0.98	0.86; 1.13	0.828	0.957
	MR PRESSO	0.98	0.85; 1.13	0.754	0.795
Never smoker	Inverse-variance weighted (fixed effects)	0.86	0.76; 0.96	8.99×10^{-3}	1.35×10^{-2}
	Weighted median	0.85	0.72; 0.99	0.037	0.112
	MR PRESSO	0.86	0.77; 0.94	0.018	0.105
Ever smoker	Inverse-variance weighted (fixed effects)	0.93	0.89; 0.98	6.61×10^{-3}	1.32×10^{-2}
	Weighted median	0.95	0.89; 1.02	0.144	0.284
	MR PRESSO	0.97	0.93; 1.02	0.273	0.411

OR (odds ratio) per increase in mean acceleration (in milli-gravities). CI, confidence interval. SNP, single-nucleotide polymorphism. FDR, false discovery rate