Target trial emulation of physical activity and cardiovascular disease risk: What is impact of the exposure assessment method?

Matthew Ahmadi[1,2,3], Borja del Pozo Cruz[4], Raaj Biswas[2,3], Armando Teixeira-Pinto[3], Dot Dumuid[5], Marla Beauchamp[6], Joanne McVeigh[7], Mark Hamer[8], Emmanuel Stamatakis[1,2,3]

1 Mackenzie Wearables Research Hub

2 Charles Perkins Centre, University of Sydney

3 Faculty of Medicine and Health, University of Sydney

4 Department of Sport Sciences, Faculty of Medicine, Health, and Sports, Universidad Europea de Madrid, Villaviciosa de Odón, Madrid, 28670, Spain

5 Alliance for Research in Exercise, Nutrition and Activity, Allied Health & Human Performance, University of South Australia, Adelaide, South Australia, Australia

6 School of Rehabilitation Science, McMaster University, Hamilton, ON, Canada

7 Curtin School of Allied Health, Curtin University, Perth, Western Australia

8 Institute of Sport Exercise and Health, Division of Surgery and Interventional Sciences, University College London, United Kingdom

Abstract

Background: Target trial emulation (TTE) designs provide a framework for strengthening causal inference in observational research, but it is unknown how vulnerable they are to substantial error when the emulated intervention (exposure) measurement is imprecise. In physical activity epidemiology specifically, correcting for confounding via TTE designs but not addressing the large measurement errors arising from self-reports (e.g. questionnaires, which typically capture partial behavioural accounts with low or very low precision) creates uncertainty about possible dominance of type 2 error biases arising from such novel designs. High resolution wearables-based methods capture most movement, providing an assessment of physical activity behaviour with substantially less measurement error. No study has examined how measurement method influences causal inference in physical activity research.

Objectives:

We applied TTE methodology to sub-samples of the UK Biobank cohort with repeat exposure measurements, to compare the effects of an emulated physical activity intervention on incident CVD risk, when the physical activity exposure was quantified using self-report vs. wearable devices.

Methods:

We emulated a Randomised Controlled Trial, by identifying physically inactive adults (<150 moderate-to-vigorous physical activity (MVPA) mins/week) who had repeat assessments for both wearable and self-reported physical activity. At re-examination, participants were categorized into intervention (adopted ≥150 MVPA mins/week) or control (remained physically inactive) groups. Participants in each group were propensity score-matched to balance lifestyle behaviours, demographic, and health factors. Cumulative risk for CVD incidence was assessed through survival curves, hazard ratios, risk ratios, using Fine-Gray subdistribution and Poisson regression models.

Results:

The wearables analytic sample included 490 participants (245 per arm; mean incident CVD follow-up 4.4 years), and the self-report sample included 11,302 participants (5,651 per arm; mean follow-up 6.3 years). In wearables assessments,

guideline-adherent participants had markedly lower cumulative CVD risk (8.0% vs. 17.0%; hazard ratio [95%CI] = 0.59 [0.36, 0.98]; relative risk = 0.45 [0.28, 0.72]). In contrast, self-report assessments showed near-identical risk trajectories for intervention and control groups (21.6% vs. 21.2%; hazard ratio = 0.98 [0.89, 1.08]; relative risk = 0.92 [0.84, 1.00]). A series of sensitivity analyses confirmed these findings.

Conclusion:

Reliance on self-reports of physical activity in TTE studies may obscure true emulated intervention effects due to non-differential misclassification, increasing considerably risk of Type II error. Exposure assessment using wearable devices may be essential for valid causal inference in TTE studies of physical activity and CVD risk. Future TTE studies should prioritise objective measurements to accurately inform public health policy and guidelines.

Introduction

Randomised controlled trials (RCTs) are widely regarded as the gold standard for causal inference in evaluating intervention effects. By randomly assigning participants to treatment or control groups, RCTs minimise differences between study arms aside from the intervention itself, thus allowing any observed differences in outcomes to be attributed causally to the treatment. However, the logistics, costs, and ethical concerns associated with extended follow-up, such as hospitalization and mortality, often limit the feasibility of traditional RCTs[1 2]. Currently, physical activity guidelines are developed using almost exclusively observational prospective studies, the large majority of which used self-reports to quantify the exposure[3-5].

In scenarios where RCTs cannot be practically conducted, target trial emulation offers a rigorous framework for causal inference based on observational data[6]. This approach involves explicitly designing an observational analysis to mirror the protocol of a hypothetical RCT that addresses the causal research question at hand. The goal is to preserve the internal validity of RCTs while leveraging real-world data, thereby enhancing transparency and mitigating key biases often associated with observational research[7].

Target trial emulation encompasses a two-step process[8]. First, investigators specify a detailed protocol (including eligibility criteria, treatment strategies, treatment assignment procedures, timing of follow-up, outcomes, causal estimands, and analysis plans) that defines the ideal randomized trial to answer the research question. Second, the components of this protocol are then applied to observational data: eligible individuals are identified, treatment strategies are assigned, follow-up and censoring periods are delineated, and analytic techniques (such as propensity score matching or inverse probability of treatment weighting) are employed to minimise baseline differences in potential confounders between groups except for the treatment.

While target trial emulation designs have gained momentum in recent years[9] and is advancing observational population study designs[10] by systematically addressing biases that contribute to Type I error (false statistically significant effects), it remains vulnerable to Type II error (false statistically non-significant effects) when the intervention target variable is measured imprecisely. For target trial emulation studies

to yield valid causal inferences, accurate and unbiased measurement of the intervention is crucial[11]. This concern is especially pronounced in physical activity observational research, where reliance on self-reported data (e.g. questionnaires) introduces substantial reporting bias and non-differential misclassification error. Self-reports of physical activity capture partial behavioural accounts (mostly leisure-time activities lasting for at least 10-15 minutes) with low precision[12-14]. When such misclassification occurs, target trial emulation frameworks may be prone to Type II error, which may artificially nullify findings despite true benefits of physical activity. It is unclear if the markedly low effects sizes reported by the early application of TTE designs in self-reported physical activity research[15 16] are signs of such Type II error. In contrast, wearable devices are not susceptible to self-report bias and when matched with high resolution machine learning based methods provide greater accuracy in quantifying physical activity continuously with high measurement granularity.

We used a target trial framework to mirror an RCT intervention among physically inactive adults that estimated the effects of meeting activity guidelines[3] versus remaining physically inactive on incident CVD risk. To evaluate the impact of measurement type on causal inference estimates, we compared causal effects on CVD risk from self-report and wearable-based physical activity data.

Methods

Comprehensive descriptions of the UK Biobank design have been published previously [17]. In summary, between 2006 and 2010, around half a million adults aged 40–69 years from England, Scotland, and Wales were enrolled in this large-scale prospective cohort (overall response rate 5.5%).

Physical activity measurement

Self-reports

Baseline self-reported physical activity was measured between 2006-2010 when participants underwent a touchscreen-based IPAQ questionnaire[18] assessment. days. For the IPAQ, participants reported their frequency and duration per week of walking, and moderate and vigorous intensity activity lasting for at least 10 continuous minutes [18]. Physical activity was expressed as metabolic equivalent of task (MET) minutes and based on the International Physical Activity Questionnaire

scoring procedure where 600 MET-min/week is equivalent to the guideline of 150 minutes of MVPA/week.

The exposure window in the TTE analyses was defined as the time from physical activity re-examination (time zero) to CVD incidence or censoring, whichever came first, i.e. for self-report physical activity it began between 2012-2019.

Wearables

A subgroup of 103,712 participants later took part in the main accelerometer substudy conducted between 2013 and 2015. Each was mailed an Axivity AX3 wrist-worn triaxial accelerometer (sampling at 100 Hz) and instructed to wear it continuously on the dominant wrist for seven consecutive

For wearables-based physical activity, the exposure window began between 2017-2019, when a sub-cohort of 5067 individuals from this accelerometry sample with complete data were invited to repeat the accelerometer assessments[19], using a stratified sampling by sex and 10-year age bands. Participants were scheduled to complete four monitoring sessions approximately three months apart, with flexibility allowing missed early sessions to be compensated later. Among those invited, 3208 individuals (63%) agreed to participate. All data collection procedures were identical to those used in the original accelerometry substudy.

For the wearable devices, the acceleration signal was calibrated and non-wear periods were identified according to standard procedures[20]. Monitoring days were considered valid if wear time was greater than 16 hours, and to be included in analysis participants were required to have at least four valid monitoring days, with at least one of those days being a weekend day. MVPA was assessed using a 2-stage Random Forest physical activity intensity classifier schema (accuracy = 84% and F-score= 0.83). We have described the 2-stage schema in detail previously[21], and we have also appended it as **Supplemental Text 2**. In brief, Physical activity intensity was classified with an accelerometer-based activity machine learning scheme covering sedentary, light, moderate, and vigorous intensity physical activity. The 2-stage activity scheme uses features in the raw acceleration signal to identify and quantify time spent in different activity types and then intensities in 10s windows.

Target trial emulation sample

A target trial emulation approach enables the use of observational data to emulate a "target" trial (i.e., the RCT that would have been conducted, were it possible to do so), with exposures framed as hypothetical interventions over a predefined treatment/exposure period[8]. In the case of physical activity, the intervention group (e.g., the intervention group did not meet a specific threshold of weekly physical activity volume at baseline, but they met it at follow-up; the control group met the same weekly volume threshold in neither time points). We emulated a target trial in which the "intervention" was conceptualised as physically inactive participants at baseline being assigned to meet physical activity guidelines at follow up, compared with controls defined as not meeting guidelines at follow up (**Table 1**).

To emulate an intervention with a start and an end in adults who were not meeting physical activity guidelines and then enrolled into a physical activity intervention, we included participants from the UK Biobank cohort who had at least 2 assessments (baseline and follow-up) of wearables or self-report data. To emulate physically inactive adults entering into an intervention, participants were only included in the study if they did not meet physical activity guidelines at baseline assessment. Additional information for the target trial emulation design and timeline comparison of the "target trial" and "trial emulation" is provided in **Supplementary Text 1**.

Table 1: Comparison of target trial and emulated trial using UK Biobank data

	Target trial	**Target trial emulation with UK Biobank**
Population/ Eligibility criteria	Middle-age to older adults: -Who do not meet physical activity guidelines - With unimpaired physical function, -With no history of cardiovascular disease, -Who are prepared to participate in a physical activity intervention that will randomise them to an intervention arm to meet	Middle-age to older adults: -Who and do not meet physical activity guidelines - With unimpaired physical function, - with no history of cardiovascular disease

	physical activity guidelines or to a control arm	
Trial arms	Intervention: physically inactive adults who are allocated to a prescription to meet physical activity guidelines over a 5 to 12 year period	Intervention: physically inactive adults at baseline who meet physical activity guidelines at re-examination and followed for 5 to 12 year period (referred to hereafter as “treatment group”)
	Control: physically inactive adults who are asked not to change behaviour	Control: physically inactive adults at baseline who do not meet guidelines at re-examination (referred to hereafter as “control group”)
Assignment procedures	Random assignment to either intervention or control arm at recruitment	Propensity score matching of covariates between the intervention and control arm
Time zero / start of follow-up	Date of randomisation	Date of re-examination (when physical activity exposure was reassessed)
Follow-up period	5 to 12 years of intervention period	5 to 12 years of intervention period
Outcomes	Cardiovascular disease incidence (hospitalisation or mortality)	Cardiovascular disease incidence (hospitalisation or mortality)
Outcome ascertainment	Collected according to trial protocol	Linkage to national hospital and mortality records in England, Wales, and Scotland

Causal contrast	Intention-to-treat: comparison of groups as randomised, regardless of adherence	Per-protocol: comparison of groups defined by actual PA status at re-examination, with exposure fixed from that point onward
Effect measure	Cumulative risk ascertained through survival curves, risk ratios, and risk differences	Cumulative risk ascertained through survival curves, risk ratios, and risk differences

Emulated intervention: At the time of re-examination, these baseline inactive adults were classified into the intervention arm if they had adopted sufficient physical activity (≥150 MVPA-mins/week) or into the control arm if they continued to not meet guidelines. This approach allowed us to emulate the effect of enrolling into an intervention to meet guidelines versus remaining inactive and enables estimation of the impact of increasing physical activity on CVD incidence in a design analogous to a RCT. In doing so, it also strengthens the causal interpretation of the effect of increased physical activity.

Cardiovascular disease ascertainment

Participant data was linked with electronic health records for mortality through the National Health Service (NHS) Digital of England and Wales or the NHS Central Register and National Records of Scotland up through November 30th, 2022. Inpatient hospitalisation data were provided by either the Hospital Episode Statistics for England (October 31st, 2022), the Patient Episode Database for Wales (May 31st, 2022), or the Scottish Morbidity Record for Scotland (August 31st, 2022). CVD was defined as major CVD incidence including ischaemic heart disease, pulmonary heart disease, heart failure, stroke, and hypertensive heart disease.

Matching participants for lifestyle and health factors

After categorising participants according to whether they met physical activity guidelines (intervention) or not (control), we implemented propensity score matching to balance potential confounding lifestyle and health characteristics between groups. Propensity scores were estimated using a logistic regression model that included the

following covariates: age, sex, education, ethnicity, dietary pattern, alcohol consumption, smoking status, sleep duration, screen time, medication use (antihypertensive, antidiabetic, lipid-lowering agents), prevalent cardiovascular disease, prevalent cancer, and parental history of cardiovascular disease or cancer.

This approach simulates the exchangeability of randomised groups by ensuring that, at study entry, the distribution of lifestyle and health factors is similar between the intervention and control arms. Thus, differences in outcomes can more plausibly be attributed to the exposure of meeting physical activity guidelines, closely mirroring the comparison of a RCT. Matching on these baseline characteristics accounts for measured confounding and strengthens the causal inference from observational data in the target trial emulation framework. **Supplemental Figure 1** displays a Love plot showing the standardized mean differences for baseline covariates between the intervention and control groups, both before matching (full sample) and after matching (inclusion of only participants with similar lifestyle and health characteristics between the two groups). Based on standard practice, standardised mean differences with a value less than 0.1 is indicative of negligible imbalance[22].

Statistical analysis

We used Cox proportional hazard regression models with Fine-Gray subdistribution to address the competing risk for death, which allows a more accurate estimation of the cumulative incidence of the outcome in the presence of a competing event (non-CVD death) that precludes the occurrence of the event of interest (CVD hospitalisation or death). To mirror the analyses traditionally used in RCT's, in addition to hazard ratios, we generated cumulative risk curves from the hazard regression models. We also used modified Poisson regression models with robust standard errors to estimate risk ratios and risk differences. The control group was used as the reference, when applicable. If significant risk differences were observed, we conducted sensitivity analyses that included: 1) exclusion of participants with prevalent diseases, 2) exclusion of participants with an event in the first year of follow-up, 3) restricting analyses to older adults (>65 years) and censoring follow up time at 2 years (to mirror follow up time in an intervention trial), and 3) using the upper recommendation (300 minutes/week of MVPA) of the physical activity guidelines to delineate our control and treatment groups. We used R 4.2.1 (R Core

Team, 2017) to conduct all analyses. This study was reported in accordance with the TARGET guidelines (**Supplemental Table 1**).

Results

Our analytic sample for the wearables-based sample included 490 participants (245 in the control and 245 in the treatment arms) with a mean (SD) follow-up time of 4.4 (1.02) years and 63 CVD incidence events. Our self-report-based sample included 11,302 participants (5,651 in the control and 5,651 in the treatment arms) with a mean (SD) follow-up time of 6.3 (2.4) years and 1,642 CVD incidence events. **Table 2** displays the participant characteristics for both analytic samples.

Figure 1 presents the cumulative risk for CVD incidence for the treatment (meeting physical activity guidelines at re-examination) and control (not meeting physical activity guidelines at baseline and re-examination) participants. In the wearables assessed physical activity analyses (**Figure 1A**), participants in the control had a higher cumulative risk over 5 years of follow-up compared to the treatment participants. Specifically, at the end of follow-up participants in the control had a cumulative risk of 16.99% [95% CI = 12.83%, 20.95%] and participants in the treatment group had a risk of 8.06% [5.69%, 11.81%]. In contrast, self-reported physical activity showed minimal differentiation in cumulative risk between control and treatment participants (**Figure 1B**). The cumulative risk increased in parallel for both groups, with the control group reaching 21.56% [20.10%, 22.99%], and the treatment group reaching 21.16% [19.64%, 22.64%] over 10 years of follow-up.

Analyses using the control group as the reference showed that participants in the treatment group, as measured by wearables, had a hazard ratio of 0.59 [0.36, 0.98], while those based on self-report had a hazard ratio of 0.98 [0.89, 1.08]. The corresponding relative risk estimates were 0.45 [0.28, 0.72] for wearables and 0.92 [0.84, 1.00] for self-report. The %risk difference was –2.59% [–4.89%, -0.29%] for wearables compared to –1.24% [–3.23% to 0.75%] for self-reported assessment.

Sensitivity analysis in the wearables sample excluding participants with an event in the first year of follow-up, prevalent CVD, re-defining the PA intervention to meet the upper threshold recommendation (300 mins/week of MVPA) to delineate control and treatment groups, and restricting analyses to participants 65 years and older (with

exclusion of events after the first 2 years of follow-up) showed results that were consistent with the main analysis (**Supplemental Figures 2 and 3**).

Discussion

We found differential causal estimates for trials emulating the same physical activity intervention defined by two measurement types (self-reported vs wearables-captured physical activity. Our findings showed a robust association between physical activity guidelines and lower CVD incidence risk when the emulated treatment (e.g. physical activity) was measured using wearable devices, but no evident risk difference when using self-reported physical activity. Informing future target trial emulation studies, our findings indicate that participant group matching for demographic, behavioural, and clinical characteristics, to account for Type I error (e.g. falsely statistically significant effects due to confounding) significantly raises the risk of Type II error (e.g. falsely statistically non- significant effects due to poor exposure measurement). Well-balanced study groups on lifestyle and health confounders can strengthen causal inference, but the validity of risk contrasts, such as CVD incidence in our current study, seems to also be largely dependent on the accuracy of the emulated intervention measurement method. In our wearables-based analyses, the separation in cumulative risk trajectories by control and treatment groups was markedly clear, while self-report measures resulted in overlapping risk curves and clearly null findings, regardless of greater power from higher sample size and events. Our findings suggest future target trial emulation studies should include precise exposure assessment to minimise Type II error and avoid missing true effects of the exposure-outcome relationship due to non-differential misclassification.

Our findings demonstrate that the measurement method used to assess the emulated intervention exposure may be a critical determinant of error and bias in target trial emulation designs. When target trial emulation studies rely solely on self-reported physical activity, they maybe at risk of substantial misclassification from inherent reporting biases, which can lead to an increased likelihood of Type II error and null results even when a true association or causal effect exists. Objective measurement methods, such as accelerometers, provides more accurate exposure

measurements; this negates the risks of reporting biases, minimises non-differential misclassification, and thus strengthening causal inference.

As population-based cohorts increasingly incorporate wearables-based assessment of physical activity and other lifestyle behaviours, it maybe essential for future target trial emulation studies to prioritise high-precision measurement modalities for exposure assessments. Demonstrating accuracy in the assessment of the emulated trial's exposure should be a key criterion in the design and reporting of target trial emulation studies. Further, the comparisons in our study also show larger sample sizes and a higher number of events cannot fully compensate for inaccuracy of the exposure measurement. The reliability and generalisability of causal estimates and interpretations in future physical activity target trial emulation studies may depend on the method used to assess the emulated intervention, with wearables assessment potentially serving as a crucial foundation for robust public health evidence from such studies.

Strengths and limitations

Strengths of our study include its target trial emulation design, the use of both wearable devices and self-reported physical activity measures, the inclusion of participants with repeat examination data, and electronic health record linkage for CVD incidence. This allowed direct comparison of exposure assessment methods within the target trial emulation framework to elucidate the potential risks for Type II error. Limitations of our study include the low response rate (5.5%) in the UK Biobank cohort; however, prior research suggests that poor representativeness does not necessarily influence the observed relationships between physical activity and health outcomes[23]. Although the target trial emulation design strengthens causal inference, it cannot definitively establish causality, e.g. unmeasured confounding may still exist despite rigorous analytic methods addressing measured confounding.

Conclusion

In this target trial emulation, causal estimates of physical activity and CVD incidence varied markedly by measurement type. Wearables-based physical activity showed robust protective associations, whereas self-reported measures yielded null results, despite larger sample sizes. These findings highlight that exposure misclassification from self-report may substantially increase Type II error, obscuring true effects.

Objective, wearable-derived measures—now increasingly available in population cohorts—should be prioritised to enhance validity of causal inference in future target trial emulation studies, to strengthen the evidence base linking physical activity with cardiovascular health. Without addressing this issue, the physical activity evidence base risks distortion in the years to come as more target trial emulation analyses of existing self-report data are published, leading to de-prioritisation in public health and clinical care policy, further worsening economic and health impacts or unhealthy lifestyles in the long term.

Data Resource: This research was done using the UK Biobank, a major biomedical database, under application 25813. The authors thank all of the participants and professionals contributing to the UK Biobank. Data is available upon application to the UK Biobank

Funding: National Heart Foundation (Dr. Ahmadi; APP107158) and by a National Health and Medical Research Council Investigator Grant (Dr. Stamatakis; APP1194510

References

1. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *American journal of epidemiology* 2016;183(8):758-64.
2. Cribb L, Moreno-Betancur M, Wu Z, et al. Moving beyond the prevalent exposure design for causal inference in dementia research. *The Lancet Healthy Longevity* 2025;6(2)
3. Bull FC, Al-Ansari SS, Biddle S, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. *British Journal of Sports Medicine* 2020;54(24):1451-62. doi: 10.1136/bjsports-2020-102955
4. Piercy KL, Troiano RP, Ballard RM, et al. The Physical Activity Guidelines for Americans. *JAMA* 2018;320(19):2020. doi: 10.1001/jama.2018.14854
5. Ross R, Chaput JP, Giangregorio LM, et al. Canadian 24-Hour Movement Guidelines for Adults aged 18-64 years and Adults aged 65 years or older: an integration of physical activity, sedentary behaviour, and sleep. *Appl Physiol Nutr Metab* 2020;45(10 (Suppl. 2)):S57-s102. doi: 10.1139/apnm-2020-0467
6. Hernán MA, Dahabreh IJ, Dickerman BA, et al. The target trial framework for causal inference from observational data: why and when is it helpful? *Annals of Internal Medicine* 2025;178(3):402-07.
7. Ding D, Tarp J, Giovannucci E, et al. Causal inference in physical activity epidemiology research: an opportunity for better evidence: BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine, 2025:963-65.
8. Hernán MA, Wang W, Leaf DE. Target trial emulation: a framework for causal inference from observational data. *Jama* 2022;328(24):2446-47.
9. Anderson AH. Invited commentary: target trial emulation-a call for more widespread use. *Am J Epidemiol* 2025;194(3):659-61. doi: 10.1093/aje/kwae222
10. Nguyen B, Clare P, Mielke GI, et al. Physical activity across midlife and health-related quality of life in Australian women: A target trial emulation using a longitudinal cohort. *Plos Medicine* 2024;21(5):e1004384.
11. Scola G, Chis Ster A, Bean D, et al. Implementation of the trial emulation approach in medical research: a scoping review. *BMC Medical Research Methodology* 2023;23(1):186.
12. Gill JM, Chico TJ, Doherty A, et al. Potential impact of wearables on physical activity guidelines and interventions: opportunities and challenges. *Br J Sports Med* 2023;57(19):1223-25. doi: 10.1136/bjsports-2023-106822 [published Online First: 20230807]
13. Sattler MC, Ainsworth BE, Andersen LB, et al. Physical activity self-reports: past or future? *Br J Sports Med* 2021;55(16):889-90. doi: 10.1136/bjsports-2020-103595 [published Online First: 20210203]
14. Lee IM, Keadle SK, Matthews CE. Fitness Trackers to Guide Advice on Activity Prescription. *JAMA* 2023 doi: 10.1001/jama.2023.19332 [published Online First: 20231009]
15. Guo F, McGee EE, Chiu YH, et al. Evaluating recommendation-based dietary and physical activity strategies for prostate cancer prevention: a target trial emulation in the Health Professionals Follow-up Study. *Am J Epidemiol* 2025;194(2):449-59. doi: 10.1093/aje/kwae184
16. Nguyen B, Clare P, Mielke GI, et al. Physical activity across midlife and health-related quality of life in Australian women: A target trial emulation using a longitudinal cohort. *PLoS Med* 2024;21(5):e1004384. doi: 10.1371/journal.pmed.1004384 [published Online First: 20240502]
17. Collins R. What makes UK Biobank special? *The Lancet* 2012;379(9822):1173-74.
18. Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. *Medicine & science in sports & exercise* 2003;35(8):1381-95.

19. Zisou C, Taylor H, Lacey B, et al. Reproducibility and associated regression dilution bias of accelerometer-derived physical activity and sleep in the UK Biobank. *medRxiv* 2025:2025.01. 16.25320679.
20. Ahmadi MN, Nathan N, Sutherland R, et al. Non-wear or sleep? Evaluation of five non-wear detection algorithms for raw accelerometer data. *Journal of Sports Sciences* 2020;38(4):399-404. doi: 10.1080/02640414.2019.1703301
21. Stamatakis E, Biswas R, Koemel N, et al. Dose-response of incidental physical activity against cardiovascular events and mortality. *Circulation (In Press)* 2025
22. Zhang Z, Kim HJ, Lonjon G, et al. Balance diagnostics after propensity score matching. *Ann Transl Med* 2019;7(1):16. doi: 10.21037/atm.2018.12.10
23. Stamatakis E, Owen KB, Shepherd L, et al. Is Cohort Representativeness Passe? Poststratified Associations of Lifestyle Risk Factors with Mortality in the UK Biobank. *Epidemiology* 2021;32(2):179-88. doi: 10.1097/EDE.0000000000001316

Table 2: Participant characteristics for wearables and self-reported study samples

	Wearables		**Self-report**	
	Control	**Intervention**	**Control**	**Intervention**
N	245	245	5,651	5,651
Age, years	64.4 (9.7)	64.6 (9.0)	60.8 (7.5)	60.8 (7.3)
Male (%)	108 (44.1)	117 (47.8)	2,861 (50.6)	2,843 (50.3)
Education (%)				
College	115 (46.9)	114 (46.5)	2,962 (52.4)	2,969 (52.5)
A/AS level	37 (15.1)	33 (13.5)	735 (13.0)	728 (12.9)
O level	42 (17.1)	45 (18.4)	963 (17.0)	952 (16.8)
CSE	8 (3.3)	12 (4.9)	152 (2.7)	155 (2.7)
NVQ/HND/HNC	11 (4.5)	12 (4.9)	321 (5.7)	323 (5.7)
Other	32 (13.1)	29 (11.8)	518 (9.2)	524 (9.3)
Ethnicity; white (%)	225 (91.8)	224 (91.4)	5,380 (95.2)	5,374 (95.1)
Fruits and vegetable servings/day	7.4 (4.3)	7.4 (3.9)	7.4 (4.2)	7.9 (4.3)
Alcohol consumption	11.9 (13.4)	11.9 (13.8)	14.2 (16.6)	14.2 (15.8)
Smoking history (%)				
Never	142 (58.0)	149 (60.8)	3,507 (62.1)	3,489 (61.7)
Previous	86 (35.1)	80 (32.7)	1,938 (34.3)	1,959 (34.7)
Current	17 (6.9)	16 (6.5)	206 (3.6)	203 (3.6)
Sleep duration, hrs/day	8.2 (1.1)	8.2 (1.0)	7.2 (1.0)	7.2 (1.0)
Medication use (%)	51 (20.8)	56 (22.9)	1,390 (24.6)	1,387 (24.5)
Parental history of CVD (%)	119 (48.6)	125 (51.0)	3,293 (58.3)	3,287 (58.2)
Parental history of cancer (%)	75 (30.6)	72 (29.4)	1,854 (32.8)	1,832 (32.4)

Screen time, hrs/day	3.8 (1.9)	3.8 (2.1)	3.9 (2.2)	3.9 (2.1)
Prevalent cancer (%)	5 (2.0)	4 (1.6)	196 (3.5)	202 (3.6)
Prevalent CVD (%)	13 (5.3)	11 (4.5)	462 (8.2)	485 (8.6)

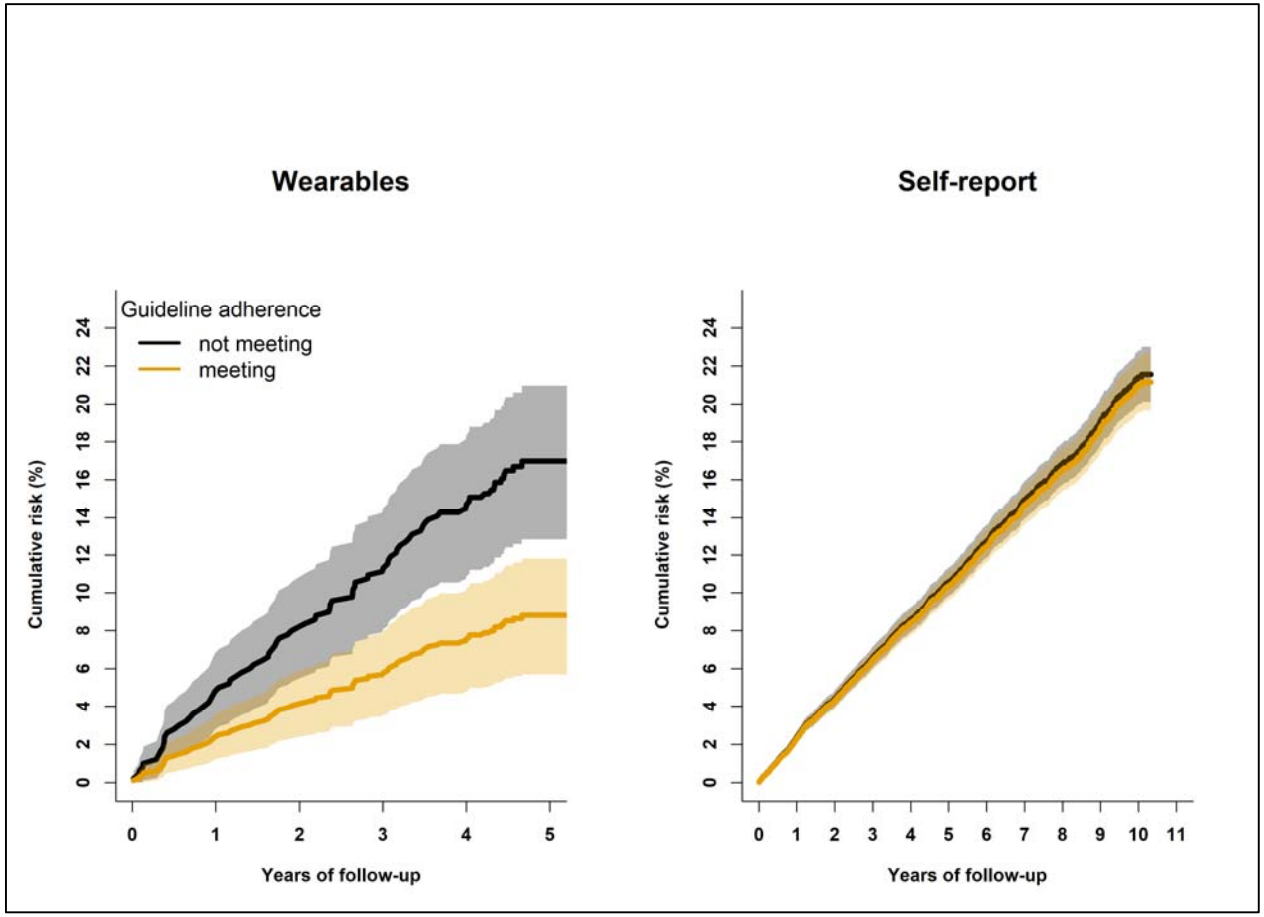
Values represent mean (SD) unless noted otherwise

Table 3: Cumulative risk, hazard ratio, relative risk, and %risk difference between control and intervention groups in the wearables and self-report samples

	Wearables		**Self-report**	
	Control	**Intervention**	**Control**	**Intervention**
Cumulative risk	16.99 [12.83, 20.95]	8.06 [5.69, 11.81]	21.56 [20.10, 22.99]	21.16 [19.64, 22.64]
Hazard Ratio	Ref	0.59 [0.36, 0.98]	Ref	0.98 [0.89, 1.08]
Relative Risk	Ref	0.45 [0.28, 0.72]	Ref	0.92 [0.84, 1.00]
%Risk Difference	Ref	-2.59 [-4.89, -0.29]	Ref	-1.24 [-3.23, 0.75]

Cumulative risk represents the risk at the maximum study follow-up time
Sample size: N=245 in both groups for the wearables sample and N= 5,651 in both groups for the self-report sample

Figure 1: Cumulative risk for cardiovascular disease incidence between control and intervention groups when using wearables or self-report to assess physical activity



Participants in the control (not meeting guidelines) and intervention (meeting guidelines) groups were matched for: age, sex, education, ethnicity, smoking history, diet, sleep duration, screen time, medication use, prevalent CVD, prevalent cancer, parental history of CVD, and parental history of cancer.
Sample size: N=245 in both groups for the wearables sample and N= 5,651 in both groups for the self-report sample