

Statistical Analysis Plan (SAP)

Physical activity and microvascular complications in adults with type 2 diabetes: A UK Biobank study

Version 2.0

Version date: 10 November 2022

SAP authors: Jakob Tarp, Frederik Pagh Kristensen

Statistical analyst: Jakob Tarp

Statistical advisors and co-authors:

Miguel A. Sanchez-Lastra, Knut Eirik Dalene, Borja Del Pozo Cruz, Mathias Ried-Larsen, Reimar W. Thomsen, Ding Ding, Ulf Ekelund

Affiliations:

JT: Department of Clinical Epidemiology, Department of Clinical Medicine, Aarhus University & University Hospital, Aarhus, Denmark

FPK: Department of Clinical Epidemiology, Department of Clinical Medicine, Aarhus University & University Hospital, Aarhus, Denmark

AS-L: Department of Special Didactics, Faculty of Education and Sports Sciences, University of Vigo, Pontevedra, Spain

KED: Department of Sports Medicine, Norwegian School of Sports Sciences, Oslo, Norway

BDPC: Biomedical Research and Innovation Institute of Cádiz (INiBICA) Research Unit, Puerta del Mar University Hospital, University of Cádiz, Spain

MR-L: The Centre of Inflammation and Metabolism & the Centre for Physical Activity Research, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

RWT: Department of Clinical Epidemiology, Department of Clinical Medicine, Aarhus University & University Hospital, Aarhus, Denmark

DD: Prevention Research Collaboration, Sydney School of Public Health, The University of Sydney, Camperdown, NSW Australia

UE: Department of Sports Medicine, Norwegian School of Sports Sciences, Oslo, Norway

Background

The goal of diabetes care is a patient-centered approach for optimizing quality of life and prevention of complications (1). Microvascular complications are common in type 2 diabetes, specifically diabetic retinopathy, diabetic neuropathy, and diabetic kidney disease. Microvascular complications can be prevented or delayed with appropriate glycemic, blood pressure and lipid management. Physical activity or exercise improves metabolic control so adults with higher levels of physical activity would be expected to have lower risk of microvascular complications. However, for the 2018 rendition of the Physical Activity Guidelines for Americans, the Physical Activity Guidelines Advisory Committee was unable to determine the relationship between physical activity and progression of retinopathy, neuropathy or nephropathy due to a lack of data.

AIMS

Determine dose-response association, including minimal effective dose and maximal risk reduction, between leisure-time physical activity and microvascular complications (diabetic retinopathy, diabetic neuropathy, and diabetic kidney disease) in adults with type 2 diabetes.

METHODS

Study design and setting

The study is nested within the UK Biobank population-based prospective cohort study. UK Biobank is designed to study the interrelations between environment, lifestyle, and genes, with the aims of improving the prevention, diagnosis, and treatment of chronic diseases. A total of 502,682 participants (approximately 5.5% of 9.2 million invited) aged 37 to 82 years were recruited via 22 assessment centers across England, Wales, and Scotland between 2006 and 2010. At the assessment centers participants completed a touch-screen questionnaire, an interview with a nurse, and a wide variety of physical measurements and biological sampling (2). A subsample has attended a repeat assessment of all data collected at the baseline examination. Data has been linked with several electronic registries for ongoing follow-up on health status. Ethical approval to establish the UK Biobank cohort was obtained by the North-West Research Ethics Committee and participants gave written informed consent before data collection.

Study population

We identified individuals with prevalent type 2 diabetes in the UK Biobank from the baseline assessment (2006-2010) and from the 1st repeat assessment (2012-2013), i.e., including individuals developing diabetes between baseline and repeat assessment.

Prevalent type 2 diabetes is determined by the algorithm developed by Eastwood et al. (3) or from measured Hba1c ≥ 48 mmol/mol. The algorithm is based on combining information on self-reported diabetes, insulin use, age of diabetes onset, and ethnicity obtained from a questionnaire in addition to self-reported diabetes, use of medications (Table 1), and age at diabetes diagnosis obtained from an interview with a trained nurse. Type 1 diabetes is removed from the sample by combining information on insulin use, time from diagnosis to initiation of insulin use, and age of diagnosis (3). Gestational diabetes is removed by combining information on age of diagnosis, current use of medication, and no concurrent report of type 2 diabetes. These criteria identify 29,236 individuals with type 2 diabetes.

Table 1. List of medications used to determine diabetes status in UK Biobank;

	Brand or product as listed in UK Biobank showcase	UK Biobank code
Insulin medication		
	insulin product	1140883066
Metformin medication		
	metformin	1140884600
	glucophage 500mg tablet	1140874686
	rosiglitazone 1mg / metformin 500mg tablet	1141189090
Non-metformin diabetes medications		
	troglitazone	1141153254
	pioglitazone	1141171646
	rosiglitazone	1141177600
	acetohexamide	1140857584
	chlorpropamide	1140874706
	tolazamide	1140874664
	tolbutamide	1140874674
	glibornuride	1140857494
	gliclazide	1140874744
	glipizide	1140874646
	glipizide product	1141157284
	gliquidone	1140874658
	glimepiride	1141152590
	repaglinide	1141168660
	nateglinide	1141173882
	amaryl 1mg tablet	1141156984
	daonil 5mg tablet	1140874724
	semi-daonil 2.5mg tablet	1140874726
	diamicron 80mg tablet	1140874746
	glibenese 5mg tablet	1140874650
	minodiab 2.5mg tablet	1140874652
	repaglinide	1141168660
	nateglinide	1141173882
	starlix 60mg tablet	1141173786
	acarbose	1140868902
	glucobay 50mg tablet	1140868908
	avandia 4mg tablet	1141177606
	actos 15mg tablet	1141171652

Outcomes

The primary outcomes are incidence of diabetic retinopathy, diabetic neuropathy, and diabetic kidney disease.

Microvascular complications are coded according to the *International Classification of Diseases, Injuries, and Causes of Death, Tenth Revision* obtained by linkage with hospital episode statistics as provided by the UK Biobank. We will use the longest available follow-up, seeking updates as they become available. Current censoring dates are 30 September 2021 (England), 31 March 2016 (Wales), and 31 July 2021 (Scotland).

The following ICD-10 codes are used;

Diabetic retinopathy	E103, E113, E123, E133, E143 H330, H332, H333, H334, H335 H350, H353, H356, H357, H358, H359 H360 H430, H431, H438 H540, H541 I708
Diabetic neuropathy	E104, E114, E124, E134, E144 G590 G629 G632 G990
Diabetic nephropathy	E102, E112, E122, E132, E142 I120 I131, I132 N180, N181, N182, N183, N184, N185, N188, N189 N19 Z992

Exposure data sources

Physical activity exposures are obtained from a touch-screen questionnaire (<https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/TouchscreenQuestionsMainFinal.pdf>).

Questionnaires were completed at study centres along with clinical measurements and a structured interview conducted by a nurse. For participants satisfying all inclusion criteria (detailed later), data obtained at the first visit will be used.

Assessment of physical activity

Our exposure variable is leisure-time physical activity (LTPA). LTPA is derived by summarizing information on frequency and duration of walking for pleasure, light DIY, heavy DIY, strenuous sport, and 'other exercises'. We will assign the following MET-values; walking: 3.3 METs, light DIY: 2.25 METs, heavy DIY: 4.5 METs, strenuous sports: 8.0 METs, and 'other exercises': 4.5 METs (4).

The following four categories will be created as primary LTPA exposure variables (assuming moderate intensity as 3.0 METs);

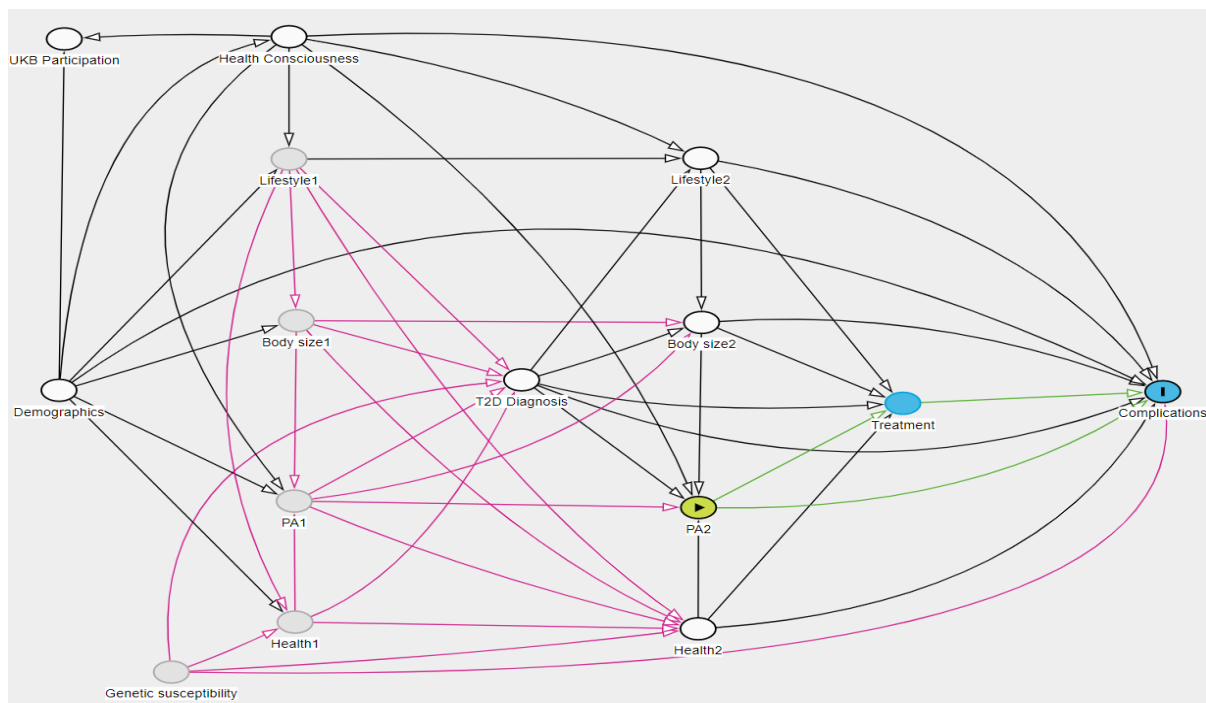
1. Zero LTPA - reference
2. >0-7.49 MET-hrs/week (below recommendation)
3. 7.5-14.9 MET-hrs/week (recommended levels)
4. ≥ 15 MET-hrs/week (above recommendation)

Quality control of physical activity data: Individuals reporting implausible high values will be excluded. Implausible high is defined as >24 hrs/day of physical activity (from IPAQ which asks about activity during all waking hours), TV-viewing, leisure time pc usage, and sleep. Sleep-time was assumed 8 hrs/day if missing.

Other variables

Variables used to mitigate confounding were selected based on a directed acyclic graph (DAG) (Figure).

Figure. Directed acyclic graph



Variables ending with 1: before type 2 diabetes

Variables ending with 2: after type 2 diabetes

The directed acyclic graph (DAG) was developed based on the template provided by Yang et al. (5). The DAG illustrates how lifestyle factors impacts the risk of developing type 2 diabetes as well as how lifestyle factors track over time (physical activity before type 2 diabetes predicts physical activity with type 2 diabetes). We have not included physical activity from other domains than leisure time in the DAG but will adjust for occupational and transportation physical activity.

Arrows from pre-diagnosis nodes to microvascular outcomes are omitted for simplicity.

As illustrated in the DAG, physical activity and health is connected through a reciprocal relationship where previous physical activity may affect current health, but current health may also impede current physical activity (PA1 -----> Health2 -----> PA2). Based on the data in UK Biobank, it is not possible to model this time-dependent relationship as we only observe participants when they already have diabetes. We thus consider Health2 as a strong confounding pathway.

Therefore, analysis will be based on the following logic;

- Remove from the analysis those individuals with a high risk of undiagnosed complications, potentially leading to reverse causation bias
- Multivariable-adjustment for duration of type 2 diabetes because complications tend to increase with time. We consider this an appropriate adjustment for poor health as physical activity does not affect duration of diabetes, but duration of diabetes may impact physical activity levels.
- No adjustment for use of glucose-, blood pressure-, or cholesterol lowering drugs (or their measured levels) in main model because physical activity may affect use of these drugs. They are therefore potential mediators of the effects of physical activity on microvascular diabetes complications.
- Under this DAG, there may be confounding from health consciousness. To address this, we have included an indicator for family history of major non-communicable diseases as a marker of health consciousness (may also be a marker of genetic susceptibility to those conditions).
- Remove from the analysis those individuals with comorbidities/physical limitations with a high risk of limitations to be physically active, i.e. these participants would not be eligible for a trial testing the effect of physical activity on microvascular complications (see below for a list of exclusion criteria).

List of confounding variables identified from the DAG and their operationalization

Acquired through a combination of self-report from an electronic questionnaire and reported during interview with a nurse

Age: used as time-scale

Sex: male/female

Ethnicity: European, South Asian, African Caribbean, other

Other PA:

Transportation: walking, cycling, passive

Occupation: Mainly sedentary, standing occupation, manual + heavy work

Lifestyle:

Smoking: never, previous, current

Diet: dietary quality index based on 1) ≤ 3 weekly servings of red meat and ≤ 1 serving/week of processed meat, 2) ≥ 2 servings/week of fish including one with oily fish, and 3) ≥ 400 grams of fruit and vegetables/day

Alcohol: never, previous, current < 3 times/week, current ≥ 3 times/week

Body size: BMI, continuous

SES:

Living with partner (yes/no)

Education: No qualifications, Other qualifications than college/university degree, University degree

Townsend index: continuous

Employment: unemployed, employed, retired

Health:

History of CVD (yes/no, any of myocardial infarction, stroke, ischemic stroke, intracerebral or subarachnoid haemorrhage, angina, or heart failure)

History of cancer (yes/no, not including non-melanoma skin cancers)

Depression (yes/no)

Loneliness (yes/no)

Years since T2D diagnosis (calculated as assessment data minus self-reported age of diabetes diagnosis. Individuals with T2 diabetes flagged from measured HbA1c are assigned 0 years since diagnosis).

T2D diagnosis:

T2D status ascertained from self-reported diabetes ("Has a doctor ever told you that you have diabetes?") or use of diabetes medication ('Medication for cholesterol, blood pressure or diabetes' [UK Biobank Fields 6177 or 6153]).

Health consciousness:

Family history of CVD, cancer or diabetes in biological or adoptive parents or siblings. This variable may also capture genetic risk of these conditions.

Statistical analysis

Descriptive data for continuous variables is described as means with standard deviations. Categorical and categorized variables are presented as proportions within strata of LTPA.

Exclusion criteria: The target population is adults with type 2 diabetes capable of doing physical activity.

- Diabetic retinopathy, diabetic neuropathy, diabetic kidney disease (defined according to outcome)
- Other subtypes of neuropathy (Hereditary, inflammatory, alcohol- or drug induced, infectious, connective disorders, nutritional or neoplastic).
- Chronic degenerative neurological problems (Parkinson's disease, dementia/Alzheimer's/cognitive impairment, motor neuron disease, myasthenia gravis, multiple sclerosis, other demyelinating disease)
- Metastatic cancer
- Chronic immunological/systemic diseases (Rheumatoid arthritis, vasculitis, giant cell/temporal arteritis, polymyalgia rheumatica, Wegners granulomatosis, microscopic polyarteritis, polyarteritis nodosa, systemic lupus erythematosus/sle, sjogren's syndrome/sicca syndrome, dermatopolymyositis, dermatomyositis, polymyositis, scleroderma/systemic sclerosis, chronic fatigue syndrome, antiphospholipid syndrome)
- Renal/kidney failure (renal failure requiring dialysis, renal failure not requiring dialysis, kidney nephropathy, IgA nephropathy, diabetic nephropathy, nephritis, glomerulonephritis)
- eGFR < 60
- Liver failure/cirrhosis/chronic pancreatitis
- Psychological/psychiatric problems (Schizophrenia, mania/bipolar disorder/manic depression, deliberate self-harm/suicide attempt, post-traumatic stress disorder, Anorexia/bulimia/other eating disorder)
- Substance abuse/dependency (Alcohol dependency, opioid dependency, other substance abuse/dependency)
- Anorexia/bulimia/other eating disorders
- Amyloidosis
- HIV/AIDS
- We are also excluding those who were pregnant, underweight (BMI >18.5), unable to walk, living in a care home or requiring attendance, disability or mobility allowance. These data are acquired using a combination of self-report (baseline questionnaire), clinical measurement, interview with a nurse, and electronic data linkage.
- Full information on putative confounding variables is needed for inclusion under the assumption that data is missing at random. We will include a standard flowchart detailing number of exclusions with reasons, including; missing exposure data, missing other data, excluded due to pre-existing conditions.

Primary outcomes:

Associations between LTPA and diabetic retinopathy, diabetic neuropathy, and diabetic kidney disease expressed as hazard ratios (HRs) from outcome-specific Cox proportional hazards models with age as the time scale and 95% confidence intervals (CI).

Analyses will be corrected for delayed entry. Participants are considered 'at-risk' from 1 year after attending a study centre. Participants are followed until development of first outcome-specific complication, emigration, loss to follow-up, withdrawal from the study or end of observation time, whichever occurs first. Statistical significance is set at $\alpha = 0.05$ (two-sided).

Four models with progressive levels of adjustment will be used;

- 1) A crude model (model 1) adjusted for sex and age.
- 2) Model 2, Model 1 + covariates listed in section 'other variables' except for BMI
- 3) Model 3 (main model), Model 2 including adjustment for BMI
- 4) Model 4, Model 3 + adjustment for glucose-lowering drugs, blood-pressure lowering drugs, statins, Hba1c, blood pressure, LDL-cholesterol, eGFR (when the outcome is kidney disease).

Some participants will have missing information on biochemical markers and will therefore not be included in model 4.

Continuous dose-response: All continuous dose-response analyses will be winsorized at the 95th percentile of the exposure distribution.

The shape of the dose-response association between MET-hrs/week of LTPA and outcomes will be explored using restricted cubic splines with 3 knots placed at the 10th, 50th and 90th percentiles of exposure distribution among individuals with non-zero LTPA. The reference category will be zero LTPA. Model fit with alternative knot placements and number of knots will be explored and evaluated using Akaike information criteria and likelihood-ratio tests.

We will estimate 10-year cumulative incidence functions, as measures of absolute risk, across categories of LTPA.

Effect modification:

Effect measure heterogeneity by demographic and clinical characteristics will be examined through stratification and evaluated statistically using the likelihood-ratio test (physical activity exposure-by-effect modifier interaction term + main effects compared to model 3 including the main effects only).

- sex (male/female)
- age (<60 vs ≥ 60 years of age)
- Diabetes duration (<5 vs ≥ 5 years)
- History of CVD
- Waist circumference, < 102 cm for men and 88 cm for women vs ≥ 102 cm for men and 88 cm for women)
- Baseline glycemic and cardiovascular risk factors (no. of treatment targets for Hba1c, systolic BP, and LDL-cholesterol achieved, categorization dependent on distribution in sample)
 - o Hba1c < 53 vs ≥ 53 mmol/mol
 - o Systolic BP < 140 vs ≥ 140 mmHg
 - o LDL-cholesterol < 2.6 vs ≥ 2.6 mmol/l

Model assumptions: Departure from proportional hazards assumption will be evaluated by tests and graphs of Shoenfeld residuals. If the model variables do not meet the proportional hazards assumption, a stratified Cox model will be employed for the relevant variables. If the violation of the proportional hazards assumption is related to statistical power, the variables in question may be adapted (simplified) to increase the number of cases within each stratum (for categorical variables).

Sensitivity analyses: The following sensitivity analyses are planned (based on model 3)

- Excluding ever smokers to examine robustness against residual confounding
- Exclude participants with a history of cancer
- Excluding participants classified as 'possible type 2 diabetes' if their Hba1c is < 48 mmol/mol
- Left-censoring first 3 years
- Only using primary diagnosis codes of "diabetes with neurological complications" since studies from Denmark have shown that secondary diagnosis codes often denote sequels after stroke (Christensen et al.).
- Using a Fine-Gray competing risk model with death as a competing event

Implementation of the SAP

The SAP will be used as a work chart for the statistical analysis and for drafting and completing the study report (scientific article). The SAP will be implemented using the following steps:

1. The SAP is circulated and approved by all co-authors. This is done prior to commencement of the statistical analysis. The SAP will be included as a supplementary file upon publication of the statistical analysis in a scientific journal.
2. Statistical analyses are performed (JT)
3. A preliminary report is drafted (FPK + JT) and circulated among co-authors
4. The report is revised and circulated among co-authors for further comments and final approval
5. When agreement about interpretation and conclusion is reached, the report is submitted to a scientific journal.

Anticipated outline of the study report (manuscript)

Table 1. Descriptive characteristics.

	No leisure time physical activity	Leisure time physical activity below recommendation	Leisure time physical activity at recommendation	Leisure time physical activity above recommendation
N (% Women)				
Age (years), mean (SD)				
BMI (kg/m ²), mean (SD)				
Waist circumference (cm), mean (SD)				
Body mass index categories (kg/m²), No. (%)				
18.5-25				
25-30				
30-35				
≥35				
LTPA (MET-hours/wk), mean (SD)				
Participation in sports, No. (%)				
Duration of diabetes (years), mean (SD)				
Townsend Index, mean (SD)				
Marital Status (living with partner), No. (%)				
Education, No. (%)				
No qualifications				
Other qualifications than college/university degree				
College/University degree				
Ethnicity, No. (%)				
European				
South Asian				
African Caribbean				
Other				
Occupational Physical activity, No. (%)				
Sedentary				
Some standing, No heavy				
Heavy manual work				

Not in employment				
Retired				
Transportation, No. (%)				
Passive				
Walking				
Cycling				
Working from home				
Not in employment				
Smoking, No. (%)				
Never				
Former				
Current				
Alcohol intake, No. (%)				
Never				
Former				
Current, <3 times/week				
Current, ≥3 times/week				
Diet quality index, No. (%)				
0 (lowest diet quality)				
1				
2-3 (highest diet quality)				
Family history of CVD, cancer or diabetes (yes), No. (%)				
History of CVD, No. (%)				
History of cancer, , No. (%)				
Depression (yes), No. (%)				
Loneliness (yes), No. (%)				
Doctor diagnosis or on treatment for type 2 diabetes (yes), No. (%)*				
Use of blood-glucose lowering drugs, No (%)**				
None				
Insulin only				
Non-insulin only				
Insulin and non-insulin				
Beta-blockers (yes), No. (%)				
Calcium-channel blockers (yes), No. (%)				

ACE-inhibitors (yes), No. (%)				
Thiazide diuretics (yes), No. (%)				
Loop diuretics (yes), No. (%)				
Potassium-sparing diuretics (yes), No. (%)				
Statins (yes), No. (%)				
Use of blood-pressure lowering drugs, No (%)				
0				
1				
2				
3 or more				
Hba1c (mmol/mol), mean (SD)***				
eGFR (mL/min/1.73m ²), mean (SD)***				
Systolic blood pressure (mmHg), mean (SD)***				
LDL-cholesterol (mmol/L), mean (SD)***				
Triglyceride (mmol/L), mean (SD)***				

*Individuals with type 2 diabetes identified from self-reported type 2 diabetes or use of glucose-lowering drugs. Remaining individuals identified from measured Hba1c.

**Reported at nurse interview, individuals identified with type 2 diabetes solely from measured Hba1c are not included in the denominator.

***Hba1c, n=, eGFR, n=, Systolic blood pressure, n=, LDL-cholesterol, n=, Triglyceride, n=

MET: metabolic equivalent, CVD: cardiovascular disease, ACE: angiotensin-converting enzyme

Table 2. Leisure Time Physical Activity and diabetic retinopathy, diabetic neuropathy, and diabetic kidney disease

	No leisure time physical activity	Leisure time physical activity below recommendation	Leisure time physical activity at recommendation	Leisure time physical activity above recommendation
Diabetic retinopathy				
N =, cases =	n / cases	n / cases	n / cases	n / cases
Crude incidence rate/1000 person-years				
Model 1 (sHR [95%CI])	1 [reference]			
Model 2 (sHR [95%CI])	1 [reference]			
Model 3 (sHR [95%CI])	1 [reference]			
Model 4 (sHR [95%CI])	1 [reference]			
Diabetic neuropathy				
N =, cases =	n / cases	n / cases	n / cases	n / cases
Crude incidence rate/1000 person-years				
Model 1 (sHR [95%CI])	1 [reference]			
Model 2 (sHR [95%CI])	1 [reference]			
Model 3 (sHR [95%CI])	1 [reference]			
Model 4 (sHR [95%CI])	1 [reference]			
Diabetic kidney disease				
N =, cases =	n / cases	n / cases	n / cases	n / cases
Crude incidence rate/1000 person-years				
Model 1 (sHR [95%CI])	1 [reference]			
Model 2 (sHR [95%CI])	1 [reference]			
Model 3 (sHR [95%CI])	1 [reference]			
Model 4 (sHR [95%CI])	1 [reference]			

Figure 1. Dose-response associations between leisure time physical activity and diabetic retinopathy, diabetic neuropathy, and diabetic kidney disease based on a restricted cubic spline

Figure 2. Leisure Time Physical Activity and diabetic retinopathy, diabetic neuropathy, and diabetic kidney disease stratified by participant characteristics

1. American Diabetes Association Professional Practice Committee. 4. Comprehensive Medical Evaluation and Assessment of Comorbidities: Standards of Medical Care in Diabetes-2022. *Diabetes care*. 2022;45(Suppl 1):S46-S59. doi:10.2337/dc22-S004
2. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779
3. Eastwood SV, Mathur R, Atkinson M, et al. Algorithms for the Capture and Adjudication of Prevalent and Incident Diabetes in UK Biobank. *PLoS One*. 2016;11(9):e0162388. doi:10.1371/journal.pone.0162388
4. Pearce M, Strain T, Kim Y, et al. Estimating physical activity from self-reported behaviours in large-scale population studies using network harmonisation: findings from UK Biobank and associations with disease outcomes. *Int J Behav Nutr Phys Act*. 2020;17(1):40. doi:10.1186/s12966-020-00937-4
5. Yang Y, Dixon-Suen SC, Dugue PA, Hodge AM, Lynch BM, English DR. Physical activity and sedentary behaviour over adulthood in relation to all-cause and cause-specific mortality: a systematic review of analytic strategies and study findings. *Int J Epidemiol*. 2021. doi:10.1093/ije/dyab181