

BMJ Open Cohort profile: The MULTI sTudy Diabetes rEsearch (MULTITUDE) consortium

Elizabeth C Pino,^{1,2} Yi Zuo,^{1,2} Camila Maciel De Olivera,² Shruthi Mahalingaiah,^{3,4,5} Olivia Keiser,⁶ Lynn L Moore,² Feng Li,⁷ Ramachandran S Vasam,^{8,9,10} Barbara E Corkey,¹¹ Bindu Kalesan^{1,2,12}

To cite: Pino EC, Zuo Y, Maciel De Olivera C, *et al.* Cohort profile: The MULTI sTudy Diabetes rEsearch (MULTITUDE) consortium. *BMJ Open* 2018;**8**:e020640. doi:10.1136/bmjopen-2017-020640

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-020640>).

ECP and YZ contributed equally.

Received 17 November 2017

Revised 23 January 2018

Accepted 6 March 2018

ABSTRACT

Purpose Globally, the age-standardised prevalence of type 2 diabetes mellitus (T2DM) has nearly doubled from 1980 to 2014, rising from 4.7% to 8.5% with an estimated 422 million adults living with the chronic disease. The MULTI sTudy Diabetes rEsearch (MULTITUDE) consortium was recently established to harmonise data from 17 independent cohort studies and clinical trials and to facilitate a better understanding of the determinants, risk factors and outcomes associated with T2DM.

Participants Participants range in age from 3 to 88 years at baseline, including both individuals with and without T2DM. MULTITUDE is an individual-level pooled database of demographics, comorbidities, relevant medications, clinical laboratory values, cardiac health measures, and T2DM-associated events and outcomes across 45 US states and the District of Columbia.

Findings to date Among the 135 156 ongoing participants included in the consortium, almost 25% (33 421) were diagnosed with T2DM at baseline. The average age of the participants was 54.3, while the average age of participants with diabetes was 64.2. Men (55.3%) and women (44.6%) were almost equally represented across the consortium. Non-whites accounted for 31.6% of the total participants and 40% of those diagnosed with T2DM. Fewer individuals with diabetes reported being regular smokers than their non-diabetic counterparts (40.3% vs 47.4%). Over 85% of those with diabetes were reported as either overweight or obese at baseline, compared with 60.7% of those without T2DM. We observed differences in all-cause mortality, overall and by T2DM status, between cohorts.

Future plans Given the wide variation in demographics and all-cause mortality in the cohorts, MULTITUDE consortium will be a unique resource for conducting research to determine: differences in the incidence and progression of T2DM; sequence of events or biomarkers prior to T2DM diagnosis; disease progression from T2DM to disease-related outcomes, complications and premature mortality; and to assess race/ethnicity differences in the above associations.

INTRODUCTION

Type 2 diabetes mellitus (T2DM) is a chronic metabolic disease that can lead to complications in many body systems and increases the

Strengths and limitations of this study

- The primary strengths of the MULTI sTudy Diabetes rEsearch (MULTITUDE) consortium is the large sample size and generally long follow-up period that facilitates examination of type 2 diabetes mellitus (T2DM) risk and outcomes across the life course.
- Pooling consortium data allow us to provide insights into the evolution of T2DM risk factors and prediabetes in early life with greater statistical power than has been available previously.
- Furthermore, data from additional cohorts can be harmonised with the consortium to expand the MULTITUDE consortium to include more representative data or to improve the representation of minorities.
- Limitations include apparent heterogeneity of measures across cohorts, including variation in clinical methodology and technology, questionnaire data and diagnostic criteria.
- Long-term follow-up studies from our consortium that enrolled minorities only began tracking T2DM and cardiovascular disease events in the 1970s or later.

overall risk of chronic morbidity and premature death.¹ Globally, the age-standardised prevalence of T2DM has nearly doubled from 1980 to 2014, rising from 4.7% to 8.5% with an estimated 422 million adults living with the chronic disease.² It is currently the seventh leading cause of death in the USA with over 30 million Americans (9.4% of the US population) living with T2DM resulting in a total financial burden of US\$245 billion per year.¹ In adults, T2DM accounts for about 90%–95% of all diagnosed cases of diabetes and is commonly associated with obesity.¹

The risk of T2DM is associated with an interplay of genetic and metabolic factors including: ethnicity, family history of T2DM, previous gestational diabetes, polycystic ovary syndrome (PCOS), older age, overweight and obesity, unhealthy diet, physical inactivity and



For numbered affiliations see end of article.

Correspondence to

Dr Bindu Kalesan;
kalesan@bu.edu

smoking.^{3 4} The combination of increasing prevalence of T2DM and increasing lifespan of persons with diabetes may be altering the spectrum of morbidities that accompany T2DM. Complications include cardiovascular events (myocardial infarction (MI), heart failure, stroke), non-alcoholic fatty liver disease, kidney failure, and vision and neurological damage.⁵ The numerous and severe complications and increased years of life spent with T2DM indicate a need to better assess the trajectory of the disease and impact of various interventions and comorbidities, and the effect of attendant intermediate events on long-term outcomes.

An optimal approach to examining T2DM risk and disease progression involves the longitudinal examination of population-based cohorts. Further, integration and harmonisation of data from multiple population studies allows for sample sizes that are not obtained with individual studies. Furthermore, robust sample sizes and diverse cohorts improve the generalisability of results by increasing overall representativeness of the combined cohort.⁶⁻⁹ Another advantage to harmonising data across studies to create a single, large database is the facilitation of comparative effectiveness research.^{6 10}

The MULTI sTudy Diabetes rEsearch (MULTITUDE) Consortium was established in 2017 to harmonise data from 17 cohort studies and clinical trials and to facilitate a better understanding of the determinants, risk factors and outcomes associated with T2DM. The main research objectives of this project are to determine the relationship between the lifetime risk of T2DM and associated major risk factors, the transition-specific risk of adverse outcomes from T2DM diagnosis through intermediate morbidity to eventual mortality and to determine the temporal patterns of T2DM and related morbidity and mortality in the USA. Moreover, the MULTITUDE consortium enables evaluation of gender-specific outcomes in T2DM.

A comparable large-scale data harmonisation effort¹¹ has already been undertaken to better understand the long-term risks for cardiovascular disease (CVD) and to examine patterns of CVD development over the adult life course. However, similar projects focused on T2DM have thus far been limited in their sample size,^{12 13} participant demographic make-up,¹²⁻¹⁴ years of follow-up¹² or with their focus on improvements in care^{15 16} rather than on determining risk of diagnosis and adverse outcomes. Additionally, while multiple risk models¹⁷⁻¹⁹ aimed at early identification of patients at high risk of developing T2DM are already widely used in the clinical settings, these models typically only consider the patient's current state at the time of the assessment, ignoring the complex trajectory of events that leads up to the disease state. The MULTITUDE consortium aims to address these limitations.

COHORT DESCRIPTION

Study inclusion and follow-up time

The 17 cohort studies and clinical trials in the MULTITUDE consortium were included based on their

availability as open source data from the National Heart, Lung, and Blood Institute (NHLBI) Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) and their relevance to T2DM risk and outcomes. All studies have been approved for the sharing of their data in this consortium per the NHLBI policy for data sharing from clinical trials and epidemiology studies (<http://www.nhlbi.nih.gov/funding/datasharing.htm>). These studies vary by study design, inclusion criteria, recruitment site and enrolment age, with recruited participants ranging in age from 3 to 88 years (table 1 and online supplementary table 1). Most studies included individuals with and without T2DM at baseline. All studies except the three cohorts of the Framingham Heart Study (FHS)²⁰ include non-Caucasian patients/participants, and all but two include both men and women. Participants from 45 US states and the District of Columbia are represented across the 17 cohorts/trials making up the consortium (online supplementary table 2).

The participants recruited into the prospective cohorts in the consortium tended to be free of prevalent comorbidities, while patients enrolled in the clinical trials all suffered from either CVD, T2DM, obesity or hypertension. A diagnosis of T2DM was part of the inclusion criteria for the Action to Control Cardiovascular Risk in Diabetes (ACCORD)²¹ and Bypass Angioplasty Revascularization Investigation in Type 2 Diabetes (BARI 2D)²² clinical trials, while the same diagnosis excluded patients from enrolling in the Systolic Blood Pressure Intervention Trial Primary Outcome Paper (SPRINT-POP)²³ trial. The number of enrolled individuals with diabetes within each individual cohort/trial is provided in table 1.

The study duration and follow-up examinations differ by study. Figure 1 demonstrates the calendar years of data collection and the time range in which examinations were performed. The earliest year of data collection was 1948 (inception of the FHS²⁰) and the most recent 2017, with six cohorts still continuing follow-up examinations. The shortest interval of total follow-up, 2 months, comes from the Functional Outcomes in Cardiovascular Patients Undergoing Surgical Hip Fracture Repair (FOCUS)²⁴ trial, while the longest follow-up of 69 years comes from the original cohort of the FHS.²⁰

Data collection methods

We pooled the data using three distinct steps of crosswalk, catalogue and harmonisation. Retrospective harmonisation is a complicated process since few established studies have used identical collection methods and procedures. We determined participating study attributes and type of information collected (eg, diagnoses, clinical laboratory values). As well, we documented information such as study designs, sampling protocols and data access policies in order to evaluate sources of study heterogeneity and feasibility of harmonisation.^{25 26} To enable harmonisation, we ensured that all the study-specific data items required to generate the target variables (tables 2-4; online

Table 1 Study design and enrolment details of cohorts included in the MULTI sTudy Diabetes rEsearch Consortium

Data sets	Study type	Sample size (n)	Follow-up time	Sex (%)	Race/ethnicity	Inclusion criteria	Enrolment age (years)	Enrolled with diabetes* (n)
ACCORD ²¹	Clinical trial	10251	4–8 years	M (61), F (39)	W, B, H, O	T2DM	40–79	10251
AFFIRM ³⁴	Clinical trial	4060	5 years	M (61), F (39)	W, O	Atrial fibrillation Atrial fibrillation+stroke/risk factor	≥65 <65	813
BARI 2D ²²	Clinical trial	2368	5 years	M (70), F (30)	W, B, H, A	T2DM and CAD	≥25	2368
ARIC ³⁵	Prospective cohort	15792	Five examinations to date	M (45), F (55)	W, B	Randomly selected from four ARIC field centres in MD, NC, MS, MN	45–64	1518
JHS ³⁶	Prospective cohort	5302	Three examinations to date	M (36), F (64)	B	Randomly selected from urban and rural areas of three counties in MS	35–84	1152
FHS cohort ²⁰	Prospective cohort	5209	32 examinations to date	M (45), F (55)	W	Framingham resident	28–74	109
FHS offspring ²⁰	Prospective cohort	5124	Nine examinations to date	M (48), F (52)	W	Offspring of FHS cohort or spouse of FHS offspring	5–70	77
FHS Gen3 ²⁰	Prospective cohort	4095	Two examinations to date	M (47), F (53)	W	At least one parent in FHS offspring or one grandparent in FHS cohort	19–72	101
FOCUS ²⁴	Clinical trial	2016	60 days	M (24), F (76)	W, B, A, O	Hip surgery patient with clinical evidence for CVD	≥50	508
MRFIT ³⁷	Clinical trial	12866	6–8 years	M (100)	W, B	At increased risk of death from CHD without clinical evidence of CHD	35–57	345
ALLHAT ³⁸	Clinical trial	33 357; 10355	4–8 years	M (53), F (47)	W, B, A, O	Hypertension and at least one other CHD risk factor	≥55	15701
BHS ³⁹	Prospective cohort	11796	Seven examinations to date	M (52), F (48)	W, B	Randomly selected from a biracial, semirural Louisiana community	3–37	33
CORAL ⁴⁰	Clinical trial	947	31–55 months	M (51), F (49)	W, B	Severe renal-artery stenosis with a history of hypertension	35–88	310
NGHS ⁴¹	Prospective cohort	2379	9 years	F (100)	W, B	Randomly selected from three clinic sites in California, Ohio and Maryland	9–10	0
OMNI heart ⁴²	Clinical trial	164	18 weeks	M (55), F (45)	W, B	Prehypertension or stage 1 hypertension, not medically managed	≥30	4
POWER-UP ⁴³	Clinical trial	390	24 months	M (20), F (80)	W, B, A, Bi	BMI of 30–50 kg/m ² and at least two of five components of the metabolic syndrome	≥21	84
SPRINT-POP ²³	Clinical trial	9361	3.26 years	M (65), F (35)	W, B, H, O	High blood pressure without diabetes or a history of CVD	≥50	47

*T2DM is defined in our cohort consortium as a random plasma glucose concentration ≥200 mg/dL, which may differ from definitions within individual cohorts.

A, Asian; ACCORD, Action to Control Cardiovascular Risk in Diabetes; AFFIRM, Atrial Fibrillation Follow-up Investigation of Rhythm Management; ALLHAT, Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial; ARIC, Atherosclerosis Risk in Communities; B, black; BARI 2D, Bypass Angioplasty Revascularization Investigation in Type 2 Diabetes; BHS, Bogalusa Heart Study; Bi, biracial; BMI, body mass index; CAD, coronary artery disease; CHD, coronary heart disease; CORAL, Cardiovascular Outcomes in Renal Atherosclerotic Lesions; CVD, cardiovascular disease; F, female; FHS, Framingham Heart Study; FOCUS, Functional Outcomes in Cardiovascular Patients Undergoing Surgical Hip Fracture Repair; H, Hispanic; JHS, Jackson Heart Study; M, male; MRFIT, Multiple Risk Factor Intervention Trial for the Prevention of Coronary Heart Disease; NGHS, NHLBI Growth and Health Study; O, other; OMNI heart, Optimal Macronutrient Intake Trial to Prevent Heart Disease; POWER-UP, Practice Based Opportunities for Weight Reduction Trial at the University of Pennsylvania; SPRINT-POP, Systolic Blood Pressure Intervention Trial Primary Outcome Paper; T2DM, type 2 diabetes mellitus; W, white.

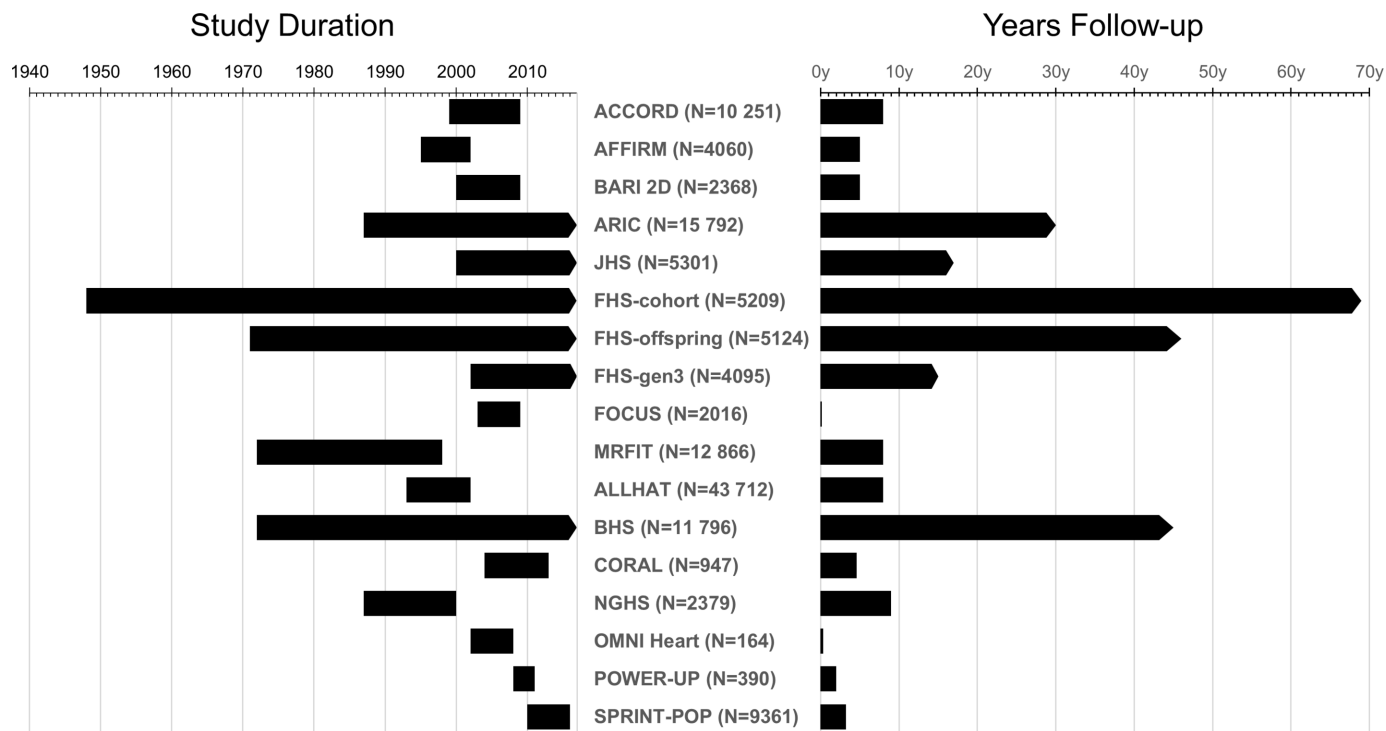


Figure 1 Follow-up time for each cohort/clinical trial included in the MULTI sTUDY Diabetes rEsearch Consortium Study duration: length of follow-up from initial year of enrolment to last year of follow-up data for each cohort. Years follow-up: cumulative years of follow-up for each cohort. ACCORD, Action to Control Cardiovascular Risk in Diabetes; AFFIRM, Atrial Fibrillation Follow-Up Investigation of Rhythm Management; ALLHAT, Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial; ARIC, Atherosclerosis Risk in Communities; BARI 2D, Bypass Angioplasty Revascularization Investigation in Type 2 Diabetes; BHS, Bogalusa Heart Study; CORAL, Cardiovascular Outcomes in Renal Atherosclerotic Lesions; FHS, Framingham Heart Study; FOCUS, Functional Outcomes in Cardiovascular Patients Undergoing Surgical Hip Fracture Repair; JHS, Jackson Heart Study; MRFIT, Multiple Risk Factor Intervention Trial for the Prevention of Coronary Heart Disease; NGHS, NHLBI Growth and Health Study; OMNI heart, Optimal Macronutrient Intake Trial to Prevent Heart Disease; POWER-UP, Practice Based Opportunities for Weight Reduction Trial at the University of Pennsylvania; SPRINT-POP, Systolic Blood Pressure Intervention Trial Primary Outcome Paper.

supplementary tables 3–5) were available and that the collected information was valid. The approach used to process data under a common format varies depending on the variables to be harmonised, the data collected by each study and the possibility to pool data.²⁶

Data crosswalk

The first step of harmonising data for the MULTITUDE consortium was the crosswalk of data measures (variables) across all studies. All available variables from individual studies within the consortium were identified and systematically entered in eight sections of (1) demographic data, (2) comorbidities, (3) laboratory values at diagnosis, (4) biomarkers, (5) medications, (6) ECG and echocardiogram (ECHO), (7) complications related to diabetes and (8) events. This crosswalk allows assessment of each variable and, in turn, allows us to determine the level of comparability between studies.

For example, in determining a baseline diagnosis of T2DM, individual studies may have dichotomous ‘yes’/‘no’ data on a history or diagnosis of T2DM. Alternatively, studies may only have data on fasting glucose levels, random plasma glucose levels or haemoglobin A1c levels. At this stage of the process, all relevant variables

were identified and collected from each study without alteration of the original data or creation of new MULTITUDE-specific target variables.

Data cataloguing

Following data element crosswalk, all variables were then catalogued based on their key characteristics and relevance in answering the research questions addressed. Recording of blood pressure may be different across studies: one study may have systolic and diastolic blood pressure measured by the technician and another could have reported values from medical records. Clinical outcomes can also be obtained from different sources such as medical records without independent adjudication, with independent adjudication or via self-report. All variables that are empirically similar or indicate the same measurement are grouped together and named under a common pooled variable. We evaluated which studies could provide data that enabled generation of each of the target variables and we qualitatively assessed the level of similarity between the study-specific and target variables.²⁶ All relevant information describing data elements and collection modes such as data dictionaries, questionnaires and standard operating

Table 2 Baseline comorbidity information included in the MULTI sTudy Diabetes rEsearch Consortium

Data sets	Metabolic disease				Cardiac disease				Vascular disease				Liver disease	Ageing diseases	Cancer	Depression
	T2DM	Obesity	Dyslipidaemia	Any	Angina	CAD	CHF	MI	Hypertension	PVD*	Stroke /TIA	Pulmonary disease				
ACCORD	+	+	+	+	+	+	+	+	+	+	+	+	+	-		+
AFFIRM	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
BARI 2D	+	+	+	+	+	+	+	+	+	+	+	+	+			
ARIC	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
JHS	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
FHS cohort	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	†‡
FHS offspring	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+\$
FHS Gen3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
FOCUS	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
MRFIT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ALLHAT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
BHS	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CORAL	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
NGHS	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+
OMNI heart	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
POWER-UP	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
SPRINT-POP	-	+	+	-	+	+	+	+	+	+	+	+	+	-	+	+

Dyslipidaemia refers to abnormally elevated cholesterol and/or triglycerides; Ageing diseases=dementia, Alzheimer's or Parkinson's; '-','-'=exclusion criteria for the clinical trial.

*Peripheral vascular disease as measured by low Ankle-Brachial Index.

†Kidney disease measured by urine albumin, serum creatinine and potassium, or dialysis.

‡Only lung and thyroid.

\$Asked at examination 2.

ACCORD, Action to Control Cardiovascular Risk in Diabetes; AFFIRM, Atrial Fibrillation Follow-up Investigation of Rhythm Management; ALLHAT, Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial; ARIC, Atherosclerosis Risk in Communities; BARI 2D, Bypass Angioplasty Revascularization Investigation in Type 2 Diabetes; BHS, Bogalusa Heart Study; CAD, coronary artery disease; CHF, congestive heart failure; CORAL, Cardiovascular Outcomes in Renal Atherosclerotic Lesions; FHS, Framingham Heart Study; FOCUS, Functional Outcomes in Cardiovascular Patients Undergoing Surgical Hip Fracture Repair; JHS, Jackson Heart Study; MI, myocardial infarction; MRFIT, Multiple Risk Factor Intervention Trial for the Prevention of Coronary Heart Disease; NGHS, NHLBI Growth and Health Study; OMNI heart, Optimal Macronutrient Intake Trial to Prevent Heart Disease; POWER-UP, Practice Based Opportunities for Weight Reduction Trial at the University of Pennsylvania; PVD, peripheral vascular disease; SPRINT-POP, Systolic Blood Pressure Intervention Trial Primary Outcome Paper; TIA, transient ischaemic attack; T2DM, type 2 diabetes mellitus.

Table 3 Medications monitored in the MULTI sTUDY Diabetes rEsearch Consortium

Data sets	ACCORD	AFFIRM	BARI 2D	ARIC	JHS	FHS cohort	FHS offspring	FHS Gen3	FOCUS	MRFIT	HAT	BHS	CORAL	NGHS	OMNI heart	POWER-UP	SPRINT-POP
Any T2DM medication	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Insulin	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Oral medications	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Biguanides	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Sulfonylureas	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Meglitinides	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Alpha-glucosidase inhibitors	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Thiazolidinediones	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
DPP4 inhibitor	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Incretin mimetic	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Any dyslipidaemia medication	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Statins	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Fibrates	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Niacin and nicotinic acid	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Any CVD medication	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Aspirin	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Other inhibitors of platelet aggregation/anticoagulant	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Diuretic	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ACE inhibitor	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ARB	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Beta blocker	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Calcium channel blocker	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Peripheral alpha blocker	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Central alpha-adrenergic agonist	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Nitrates	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Antiarrhythmic	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Potassium supplements	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Any endocrine/metabolic medication	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Progesterin/oestrogens/birth control pills	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Glucocorticoids	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Androgens	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

ACCORD, Action to Control Cardiovascular Risk in Diabetes; AFFIRM, Atrial Fibrillation Follow-Up Investigation of Rhythm Management; ALLHAT, Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial; ARB, angiotensin receptor blockers; ARIC, Atherosclerosis Risk in Communities; BARI 2D, Bypass Angioplasty Revascularization Investigation in Type 2 Diabetes; BHS, Bogalusa Heart Study; CORAL, Cardiovascular Outcomes in Renal Atherosclerotic Lesions; CVD, cardiovascular disease; DPP4, dipeptidyl peptidase 4; FHS, Framingham Heart Study; FOCUS, Functional Outcomes in Cardiovascular Patients Undergoing Surgical Hip Fracture Repair; JHS, Jackson Heart Study; MRFIT, Multiple Risk Factor Intervention Trial for the Prevention of Coronary Heart Disease; NGHS, NHLBI Growth and Health Study; OMNI heart, Optimal Macronutrient Intake Trial to Prevent Heart Disease; POWER-UP, Practice Based Opportunities for Weight Reduction Trial at the University of Pennsylvania; SPRINT-POP, Systolic Blood Pressure Intervention Trial Primary Outcome Paper; T2DM, type 2 diabetes mellitus.

Table 4 Clinical outcomes measured in the MULTIsTUDy Diabetes REsearch Consortium

Data sets	T2DM-related events*										Cardiac events*					Vascular events*			
	All-cause mortality	Cause of death	T2DM	Renal failure†	Neuropathy‡	Retinopathy‡	CVD	Angina	CAD	CHF	MI	PCI/CABG	Hypertension	Stroke/TIA	Liver disease	Ageing diseases	Depression		
ACCORD	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
AFFIRM	+																		
BARI 2D	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
ARIC	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
JHS	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
FHS cohort	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
FHS offspring	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
FHS Gen3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
FOCUS	+																		
MRFIT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
ALLHAT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
BHS																			
CORAL	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
NGHS																			
OMNI heart																			
POWER-UP	+																		
SPRINT-POP	+																		

Ageing diseases=dementia, Alzheimer's or Parkinson's.

*All events include both fatal and non-fatal.

†Based on serum potassium and creatinine levels, and need for dialysis.

‡Based on amputations.

§Only 1973–1975.

ACCORD, Action to Control Cardiovascular Risk in Diabetes; AFFIRM, Atrial Fibrillation Follow-up Investigation of Rhythm Management; ALLHAT, Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial; ARIC, Atherosclerosis Risk in Communities; BARI 2D, Bypass Angioplasty Revascularization Investigation in Type 2 Diabetes; BHS, Bogalusa Heart Study; CABG, coronary artery bypass grafting; CAD, coronary artery disease; CHF, congestive heart failure; CORAL, Cardiovascular Outcomes in Renal Atherosclerotic Lesions; CVD, cardiovascular disease; FHS, Framingham Heart Study; FOCUS, Functional Outcomes in Cardiovascular Patients Undergoing Surgical Hip Fracture Repair; JHS, Jackson Heart Study; MI, myocardial infarction; MRFIT, Multiple Risk Factor Intervention Trial for the Prevention of Coronary Heart Disease; NHLBI Growth and Health Study; OMNI heart, Optimal Macronutrient Intake Trial to Prevent Heart Disease; PCI, percutaneous coronary intervention; POWER-UP, Practice Based Opportunities for Weight Reduction Trial at the University of Pennsylvania; SPRINT-POP, Systolic Blood Pressure Intervention Trial Primary Outcome Paper; TIA, transient ischaemic attack; T2DM, type 2 diabetes mellitus.

procedures was used for cataloguing and subsequently to assess the comparability of data collected by individual studies.²⁵

Continuing with the previous example, all variables from each study relevant to a baseline diagnosis of T2DM would then be categorised together. Some variables may be in more than one category: data on fasting glucose levels could be both categorised under a continuous fasting glucose target variable as well as a dichotomous T2DM diagnosis target variable. These inconsistencies or spread of target variables are thus defined and placed under appropriate pooled variable headers.

Data harmonisation

We employed several established approaches²⁶ for harmonisation of MULTITUDE data while minimising bias that could be introduced by systematic differences in measurement techniques across cohorts:

1. A simple calibration model to transform one continuous measure into another continuous measure to operate at the same unit of measurement (eg, transferring weight in kilograms to weight in pounds).
2. An algorithmic transformation to harmonise continuous and categorical variables, or both, with different but combinable ranges or categories (eg, to identify a baseline diagnosis of T2DM: 'yes' on a 'history or diagnosis of T2DM', fasting glucose levels ≥ 126 mg/dL, random plasma glucose levels ≥ 200 mg/dL or haemoglobin A1c levels $\geq 6.5\%$).
3. A standardisation model that harmonises the same constructs measured using different scales (eg, blood cholesterol concentration). The distribution of the measure is compared across cohorts to assess for differences in accuracy and/or precision in the measure.¹¹

Data are only pooled if these methods are possible. We have also adjusted each analysis for cohort, which may help attenuate any confounding due to measurement differences or varying calendar decades across cohorts.

Baseline measures

The baseline demographic and patient/participant information are provided in online supplementary table 3 and table 1. Briefly, age, sex, race and smoking status were reported across all MULTITUDE cohorts. The majority of studies also include information on employment, education, hospitalisations, alcohol intake, dietary intake, physical activity, blood pressure and body mass index. A select number of cohorts provide information on a family history of T2DM or CVD.

The baseline comorbidity information included in the MULTITUDE consortium is shown in table 2. All studies provided baseline information on dyslipidaemia, hypertension and the T2DM status of patients/participants. The majority of cohorts/trials include information on obesity, CVD, pulmonary disease and kidney disease. Information on specific types of CVD is also provided in most studies.

Follow-up measures

Medications and clinical laboratory values monitored either at baseline or throughout the study follow-up period are detailed in table 3 and online supplementary table 4. All studies tracked whether patients/participants self-reported the use of any dyslipidaemia (including statins) or CVD medication. The majority of cohorts/trials reported the use of any T2DM (insulin or oral) medication, as well as specific CVD medications. Most studies also included information about the fasting glucose and insulin levels, and serum lipids, potassium and haemoglobin A1c.

A select number of studies collected information from ECGs and/or ECHOs over the course of the study, as shown in online supplementary table 5. Many of these studies also included information regarding cardiac dysfunctions such as dysrhythmia, QRS axis deviation, ventricular conduction defects, ST-segment abnormalities and left ventricular hypertrophy.

Collection of data on these intermediate end points and medical interventions across the lifespan enables us to better understand the evolution of T2DM and its interplay with CVD. This allows us to more confidently identify potential causal pathways to T2DM-related and CVD-related events.

Table 5 Baseline characteristics of participants of the MULTI sTUDY Diabetes rEsearch Consortium by diabetes status, n=135 156

	Total	T2DM diagnosis	No T2DM diagnosis	P values
N	135 156*	33 421	100 015	
Age, mean (SD)	54.3 (19.2)	64.2 (8.6)	51.4 (20.4)	<0.0001
Sex, %				<0.0001
Female	44.6	45.2	44.5	
Male	55.4	54.8	55.5	
Missing (n)	65			
Race, %				<0.0001
White	67.4	58.3	70.5	
Black	27.8	31.9	26.4	
Hispanic	2.8	5.1	2.0	
Other	2.0	4.7	1.1	
Missing (n)	4654			
Smoking, %				<0.0001
Yes	48.0	40.4	50.6	
No	52.0	59.6	49.4	
Missing (n)	4931			
BMI category, %				<0.0001
Underweight	6.6	0.2	8.7	
Normal	24.2	11.0	28.6	
Overweight	35.4	33.0	36.2	
Obese	33.9	55.7	26.6	
Missing (n)	4436			

Categorical variables were compared using χ^2 tests and continuous variables were compared using the Student's t-test.

*n=1720 for unknown diabetes status.

BMI, body mass index; T2DM, type 2 diabetes mellitus.

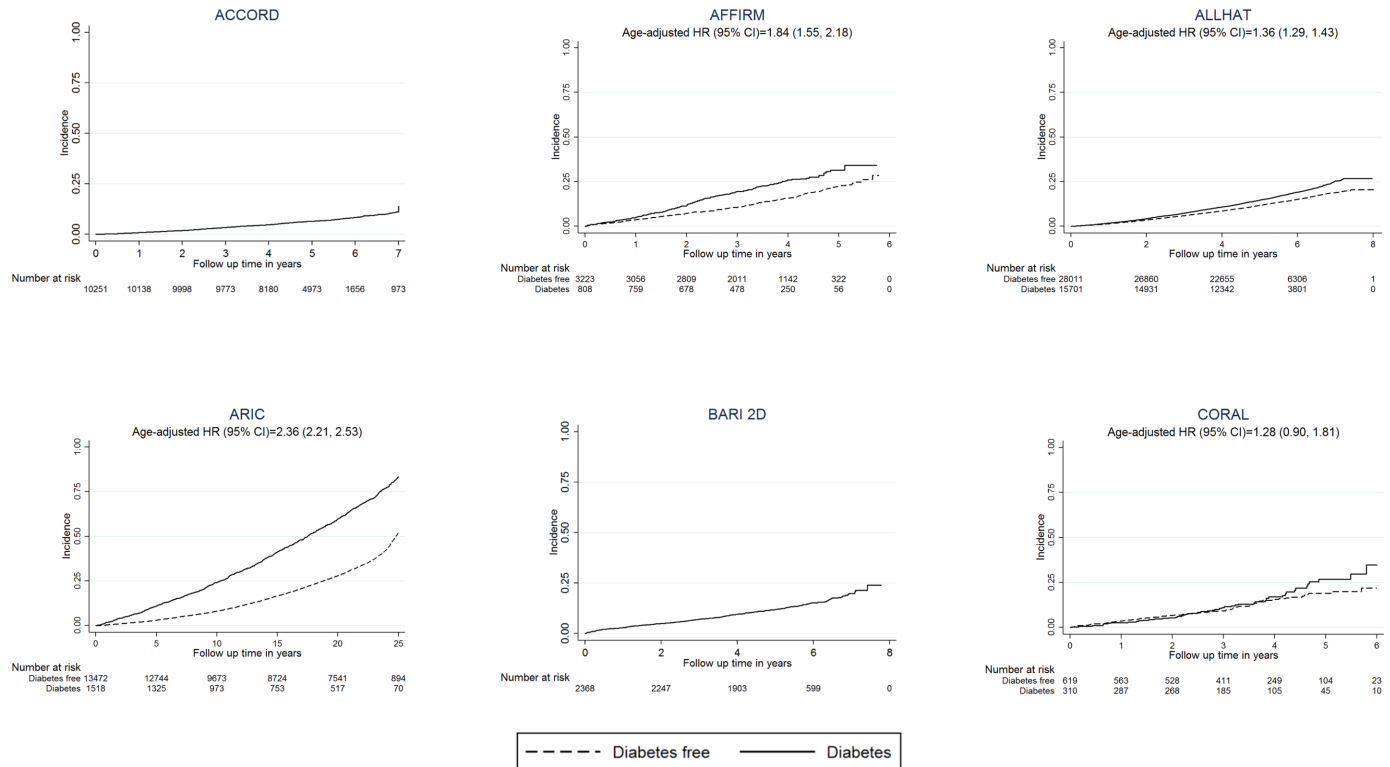


Figure 2 Cumulative incidence of all-cause mortality in the MULTI sTudy Diabetes rEsearch Consortium by baseline diabetes status Kaplan-Meier curves represent all-cause mortality. Cox proportional hazards regression models were used to estimate HRs and 95% CIs, adjusted for age. Several studies are not shown due to missing information: FOCUS, OMNI heart and POWER-UP had missing time to event data, BHS and NGHS mainly recruited children and did not track mortality. ACCORD, Action to Control Cardiovascular Risk in Diabetes; AFFIRM, Atrial Fibrillation Follow-Up Investigation of Rhythm Management; ALLHAT, Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial; ARIC, Atherosclerosis Risk in Communities; BARI 2D, Bypass Angioplasty Revascularization Investigation in Type 2 Diabetes; BHS, Bogalusa Heart Study; CORAL, Cardiovascular Outcomes in Renal Atherosclerotic Lesions; NGHS, NHLBI Growth and Health Study.

Events

Outcomes of interest are shown in table 4. Events were ascertained using each cohort's specific protocol and procedures. All but three studies within the consortium include data on all-cause mortality and the majority of studies provide information on cause of death. Both fatal and non-fatal events related to T2DM and CVD are tracked by most of the cohorts/trials in the MULTITUDE consortium. Specific types of CVD (angina, coronary artery disease, congestive heart failure, hypertension) as well as CVD-related events (MI, stroke/transient ischaemic event) and interventions requiring hospitalisation (percutaneous coronary intervention/coronary artery bypass grafting) are provided in many cohorts/trials. Initial diagnosis of T2DM as well as advanced stage outcomes of T2DM (renal failure, neuropathy, retinopathy) are also included in several studies.

Findings to date

The baseline characteristics of participants of the MULTITUDE consortium are presented in table 5. Among the 135156 participants included in the consortium, almost 25% (33 421) were diagnosed with T2DM at baseline. The average age of the participants was 54.3 years, while the average age of participants with diabetes was 64.2. Men (55.3%) and women (44.6%) were almost equally

represented across the consortium. Non-whites accounted for 31.6% of the total participants but 40% of those diagnosed with T2DM (<0.0001). Fewer individuals with diabetes reported being regular smokers than their non-diabetic counterparts (40.3% vs 47.4%, <0.0001). Over 85% of those with diabetes were reported as either overweight or obese at baseline, compared with 60.7% of those without T2DM (<0.0001).

Figures 2 and 3 show the age-adjusted incidence of all-cause mortality by baseline T2DM status for each study included in the MULTITUDE consortium. Due to the generally longer follow-up periods of prospective cohorts, we observed higher rates of overall mortality among these participants compared with clinical trials patients. We also observed more than two times the risk for mortality among individuals with T2DM in several prospective cohorts (HRs (95% CI): ARIC=2.36 (2.21 to 2.53), FHS cohort=1.28 (1.06 to 1.56), FHS offspring=2.65 (2.06 to 3.41), FHS Gen3=2.83 (1.20 to 6.70), Jackson Heart Study (JHS)=1.78 (1.52 to 2.10)) compared with lower risk among clinical trials (HR (95% CI): AFFIRM=1.84 (1.55–2.18), ALLHAT=1.36 (1.29 to 1.43), CORAL=1.28 (0.90 to 1.81), MRFIT=1.12 (0.79 to 1.58), SPRINT-POP=1.47 (0.37 to 5.92)).

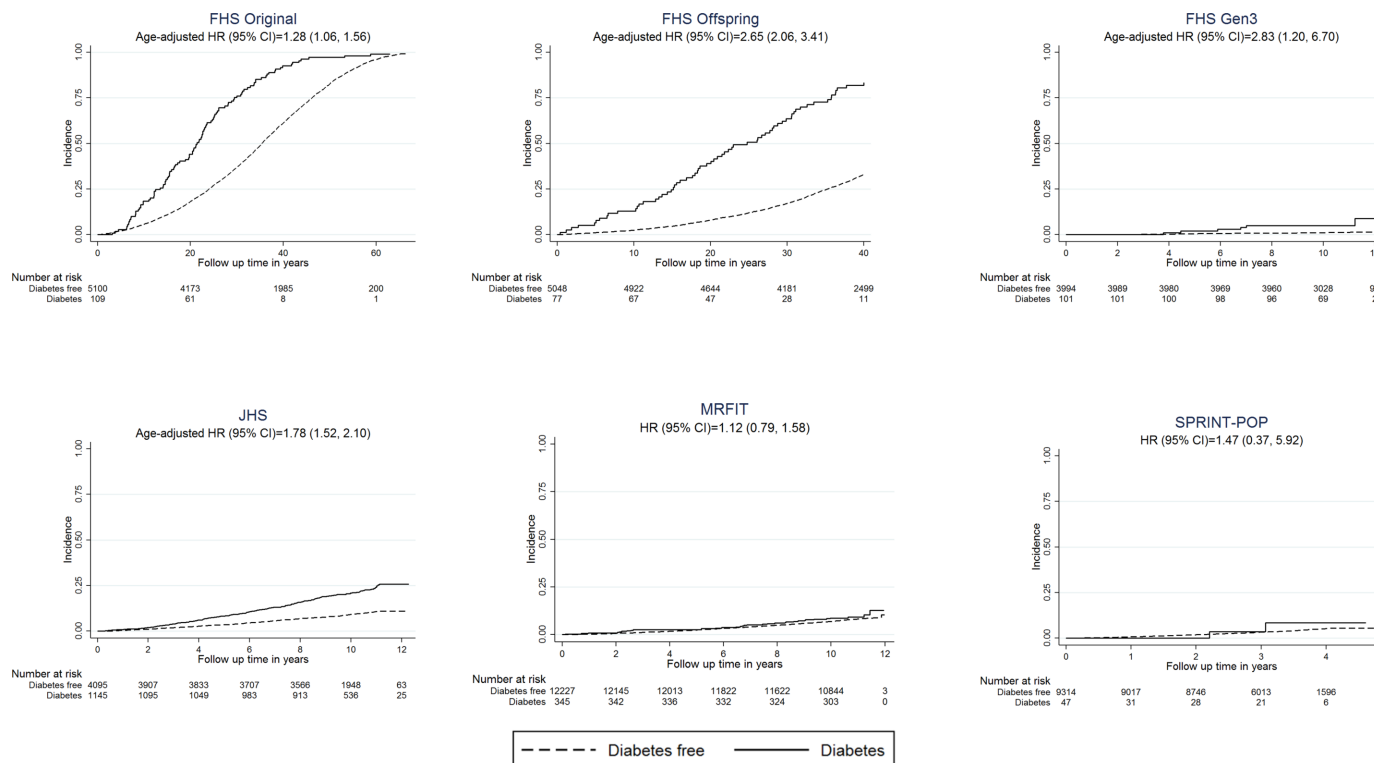


Figure 3 Cumulative incidence of all-cause mortality in the MULTI sTudy Diabetes REsearch Consortium by baseline diabetes status Kaplan-Meier curves represent all-cause mortality. Cox proportional hazards regression models were used to estimate HRs and 95% CIs, adjusted for age. Several studies are not shown due to missing information: FOCUS, OMNI heart and POWER-UP had missing time to event data, BHS and NGHS mainly recruited children and did not track mortality. FHS, Framingham Heart Study; JHS, Jackson Heart Study; MRFIT, Multiple Risk Factor Intervention Trial for the Prevention of Coronary Heart Disease; NGHS, NHLBI Growth and Health Study; SPRINT-POP, Systolic Blood Pressure Intervention Trial Primary Outcome Paper.

Conclusions

Using data from 17 harmonised cohort studies and clinical trials, the MULTITUDE consortium is a unique compilation that was established to facilitate a better understanding of the determinants, risk factors and outcomes associated with T2DM. Given the wide variation in demographics and all-cause mortality in the cohorts, MULTITUDE consortium will be a unique resource for conducting research to determine: (1) age, time period and cohort differences in the incidence and progression of T2DM (2) the sequence of events or biomarkers prior to T2DM diagnosis (3) disease progression from T2DM to CVD outcomes, T2DM complications and premature mortality and (4) to assess race/ethnicity differences in the above associations. Using the same harmonisation principles, this data resource can be extended to include a larger number of studies to provide a more comprehensive data infrastructure as relevant data are added to the BioLINCC repository. Several promising large-scale retrospective data analyses focused on gaining a better understanding of T2DM risk and outcomes are currently under way.^{27 28}

In our preliminary findings, we observed differences in demographics and all-cause mortality by baseline diabetes status. As has been previously shown,^{29 30} individuals with T2DM were more likely to be older, non-white and more overweight. However, interestingly, people diagnosed

with T2DM in the MULTITUDE consortium were less likely to be cigarette smokers, a known risk factor for the disease. This can be explained by the finding that smoking cessation is associated with weight gain and a subsequent increase in risk of diabetes,³¹ as well as the possibility that health providers and patients increase their efforts at smoking interventions after T2DM diagnosis.^{32 33} The relatively low HRs seen in the clinical trials compared with prospective cohorts is likely reflective of the comorbidities present as part of the inclusion criteria of individual studies, that is, all patients enrolled in SPRINT-POP clinical trial were previously diagnosed with high blood pressure.

Additionally, we can make preliminary conclusions regarding race differences in all-cause mortality by T2DM status. The three cohorts of FHS consist entirely of Northern whites, while the JHS-recruited Southern blacks. The risk ratio of all-cause mortality among JHS participants with T2DM more closely resembles the original FHS cohort, recruited in 1948 when CVD risk factors were largely unknown and medical interventions more limited, than the risk ratio from their contemporaries in FHS Gen3. This suggests that either individuals with T2DM are more protected from mortality in the JHS cohort or, perhaps, that all individuals in this cohort are more at risk for mortality compared with FHS Gen3. It is

likely that there is a complex interplay between genetics, lifestyle, culture and access to healthcare that remains to be explored further.

Strengths and limitations

The primary strength of the MULTITUDE consortium is the large sample size and generally long follow-up period that facilitates examination of T2DM risk and outcomes across the life course. Pooling data allow us to provide insights into the evolution of T2DM risk factors and prediabetes in early life with greater statistical power than has been available previously. Using the consortium data, we will be able to understand the variation in risk among different subgroups, including rare populations with T2DM and to observe the relationship of comorbid CVD and risk of outcomes in T2DM. Furthermore, data from additional cohorts can be harmonised with the consortium to expand MULTITUDE to include more representative data and to improve the representation of minorities.

The consortium also acknowledges a number of limitations. These include apparent heterogeneity of measures across cohorts, including variation in clinical methodology and technology, questionnaire data, and diagnostic criteria. As well, there are inherent differences in study design and methodology between clinical trials and cohort studies, which are combined in this consortium. The MULTITUDE consortium exclusively contains data from North American cohorts which may limit the generalisability of any significant findings to other global populations. We also acknowledge limited statistical power for specific subgroup analysis. Additionally, there is substantial possibility for birth cohort effects due to the trends in risk factors and development of medical therapies for prevention of T2DM and CVD.

While the individuals who have been enrolled in MULTITUDE studies the longest (FHS original cohort) have had their health and lifestyle monitored for almost 70 years and are in their 90s or 100s, this cohort was made up exclusively of Caucasians. Long-term follow-up studies from our consortium that enrolled minorities only began tracking T2DM and CVD events in the 1970s or later. It is likely that full exploration of causal mediators leading to T2DM and its related outcomes among non-whites has only recently become possible now that many participants have reached an age when incident T2DM and CVD events are just beginning to occur. MULTITUDE investigators will continue to update and expand the dataset to increase the representation of minority groups in the consortium.

Author affiliations

¹Center for Clinical Translational Epidemiology and Comparative Effectiveness Research, Boston University School of Medicine, Boston, Massachusetts, USA

²Department of Medicine, Section of Preventive Medicine and Epidemiology, Boston University School of Medicine, Boston, Massachusetts, USA

³Department of Obstetrics and Gynecology, Boston University Medical Campus, Boston, Massachusetts, USA

⁴Department of Physiology and Biophysics, Boston University School of Medicine, Boston, Massachusetts, USA

⁵Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, USA

⁶Institute of Global Health, University of Geneva, Geneva, Switzerland

⁷School of Statistics and Mathematics, Central University of Finance and Economics, Beijing, China

⁸Framingham Heart Study, Boston University's and National Heart, Lung, and Blood Institute's Framingham Heart Study, Boston, Massachusetts, USA

⁹Department of Medicine, School of Medicine, Boston University, Boston, Massachusetts, USA

¹⁰Department of Biostatistics and Epidemiology, School of Public Health, Boston University, Boston, Massachusetts, USA

¹¹Obesity Research Center, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA

¹²Community Health Sciences, Boston University School of Public Health, Boston, MA, USA

Acknowledgements The authors thank Mrithyunjay A. Vyliparambil of the University of Massachusetts Lowell, Massachusetts, for his contributions to the data crosswalk.

Contributors YZ, ECP and BK obtained the data. YZ performed the analysis. ECP drafted the manuscript. ECP, YZ, CMDO, SM, OK, LLM, FL, VSR, BEC and BK approved the manuscript, made significant contributions to the study and have read and approved the final version of the manuscript.

Funding The Evans Research Foundation at Boston University School of Medicine provided funding for the study. BK, ECP and YZ are supported by Evans Research Foundation. SM is supported by the Reproductive Scientist Development Grant (K12 HD000849); BEC is supported by the Boston Nutrition Obesity Research Center NIH Grant (P30 DK46200). OK is funded by a professorship grant from the Swiss National Science Foundation (no: 163878). FL is supported by the National Natural Science Foundation of China (No. 11501587)

Competing interests None declared.

Patient consent Not required.

Ethics approval Boston University IRB.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Details regarding access to the MULTITUDE consortium data for research purposes are available on our website. We are unable to redistribute the data due to data use agreement restrictions. However, all the individual cohorts are available on BioLINCC. Enquiries regarding use of the MULTITUDE consortium for specific research studies are welcomed in the form of a project request. For further information, please contact Bindu Kalesan at the Center for Clinical Translational Epidemiology and Comparative Effectiveness Research, Boston University School of Medicine (kalesan@bu.edu).

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- Centers for Disease Control and Prevention. *National diabetes statistics report: estimates of diabetes and its burden in the United States, 2017*. Atlanta, GA: US Department of Health and Human Services, 2017.
- World Health Organization. *Global report on diabetes: World Health Organization, 2016*.
- Mendis S. *Global status report on noncommunicable diseases 2014: World Health Organization, 2015*.
- Forouzanfar MH, Alexander L, Anderson HR, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet* 2015;386:2287–323.
- World Health Organization. *Global report on diabetes: WHO: WHO Press, 2016*.

6. Doiron D, Burton P, Marcon Y, *et al*. Data harmonization and federated analysis of population-based studies: the BioSHaRE project. *Emerg Themes Epidemiol* 2013;10:12.
7. Smith-Warner SA, Spiegelman D, Ritz J, *et al*. Methods for pooling results of epidemiologic studies: the Pooling Project of Prospective Studies of Diet and Cancer. *Am J Epidemiol* 2006;163:1053–64.
8. Khoury MJ. The case for a global human genome epidemiology initiative. *Nat Genet* 2004;36:1027–8.
9. Bath PA, Deeg D, Poppelaars JAN. The harmonisation of longitudinal data: a case study using data from cohort studies in The Netherlands and the United Kingdom. *Ageing Soc* 2010;30:1419–37.
10. Sachdev PS, Lipnicki DM, Kochan NA, *et al*. COSMIC (Cohort Studies of Memory in an International Consortium): an international consortium to identify risk and protective factors and biomarkers of cognitive ageing and dementia in diverse ethnic and sociocultural groups. *BMC Neurol* 2013;13:165–65.
11. Wilkins JT, Karmali KN, Huffman MD, *et al*. Data Resource Profile: the Cardiovascular Disease Lifetime Risk Pooling Project. *Int J Epidemiol* 2015;44:1557–64.
12. Szendroedi J, Saxena A, Weber KS, *et al*. Cohort profile: the German Diabetes Study (GDS). *Cardiovasc Diabetol* 2016;15:59.
13. Tanaka S, Tanaka S, Iimuro S, *et al*. Cohort profile: the Japan diabetes complications study: a long-term follow-up of a randomised lifestyle intervention study of type 2 diabetes. *Int J Epidemiol* 2014;43:1054–62.
14. Kearney PM, Harrington JM, Mc Carthy VJ, *et al*. Cohort profile: the cork and Kerry diabetes and heart disease study. *Int J Epidemiol* 2013;42:1253–62.
15. Arnold SV, Inzucchi SE, McGuire DK, *et al*. Evaluating the Quality of Comprehensive Cardiometabolic Care for Patients With Type 2 Diabetes in the U.S: the Diabetes Collaborative Registry. *Diabetes Care* 2016;39:e99–e101.
16. Morris AD. Considerations in assessing effectiveness and costs of diabetes care: lessons from DARTS. *Diabetes Metab Res Rev* 2002;18(Suppl 3):S32–S35.
17. Eddy DM, Schlessinger L. Archimedes: a trial-validated model of diabetes. *Diabetes Care* 2003;26:3093–101.
18. Clarke PM, Gray AM, Briggs A, *et al*. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). *Diabetologia* 2004;47:1747–59.
19. Wilson PW, Meigs JB, Sullivan L, *et al*. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Arch Intern Med* 2007;167:1068–74.
20. Tsao CW, Vasan RS. Cohort Profile: the Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. *Int J Epidemiol* 2015;44:1800–13.
21. Goff DC, Gerstein HC, Ginsberg HN, *et al*. Prevention of cardiovascular disease in persons with type 2 diabetes mellitus: current knowledge and rationale for the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial. *Am J Cardiol* 2007;99:S4–S20.
22. Frye RL, August P, Brooks MM, *et al*. A randomized trial of therapies for type 2 diabetes and coronary artery disease. *N Engl J Med* 2009;360:2503–15.
23. Wright JT, Williamson JD, Whelton PK, *et al*. A randomized trial of intensive versus standard blood-pressure control. *N Engl J Med* 2015;373:2103–16.
24. Carson JL, Terrin ML, Noveck H, *et al*. Liberal or restrictive transfusion in high-risk patients after hip surgery. *N Engl J Med* 2011;365:2453–62.
25. Doiron D, Raina P, Raina P, *et al*. Facilitating collaborative research: implementing a platform supporting data harmonization and pooling. *Nor Epidemiol* 2012;21.
26. Fortier I, Raina P, Van den Heuvel ER, *et al*. Maelstrom research guidelines for rigorous retrospective data harmonization. *Int J Epidemiol* 2017;46:103–5.
27. Indiana Biosciences Research Institute (IBRI). *2016 Annual Report*, 2016.
28. Colhoun H. A revolution in diabetes' prediction and prevention, through use of health data 2015. <https://www.axa-research.org/en/projects/helen-colhoun> (accessed 9 Jan 2018).
29. Shai I, Jiang R, Manson JE, *et al*. Ethnicity, obesity, and risk of type 2 diabetes in women: a 20-year follow-up study. *Diabetes Care* 2006;29:1585–90.
30. Narayan KM, Boyle JP, Thompson TJ, *et al*. Lifetime risk for diabetes mellitus in the United States. *JAMA* 2003;290:1884–90.
31. Wannamethee SG, Shaper AG, Perry IJ. Smoking as a modifiable risk factor for type 2 diabetes in middle-aged men. *Diabetes Care* 2001;24:1590–5.
32. Fan AZ, Rock V, Zhang X, *et al*. Trends in cigarette smoking rates and quit attempts among adults with and without diagnosed diabetes, United States, 2001–2010. *Prev Chronic Dis* 2013;10:E160.
33. Canga N, De Irala J, Vara E, *et al*. Intervention study for smoking cessation in diabetic patients: a randomized controlled trial in both clinical and primary care settings. *Diabetes Care* 2000;23:1455–60.
34. Wyse DG, Waldo AL, DiMarco JP, *et al*. A comparison of rate control and rhythm control in patients with atrial fibrillation. *N Engl J Med* 2002;347:1825–33.
35. Investigators A. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* 1989;129:687–702.
36. Sempos CT, Bild DE, Manolio TA. Overview of the Jackson Heart Study: a study of cardiovascular diseases in African American men and women. *Am J Med Sci* 1999;317:142–6.
37. The Multiple Risk Factor Intervention Trial Research Group. Multiple risk factor intervention trial: Risk factor changes and mortality results. *JAMA* 1982;248:1465–77.
38. ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *JAMA* 2002;288:2981–97.
39. Berenson GS. Bogalusa Heart Study: a long-term community study of a rural biracial (black/white) population. *Am J Med Sci* 2001;322:267–74.
40. Murphy TP, Cooper CJ, Dworkin LD, *et al*. The Cardiovascular Outcomes with Renal Atherosclerotic Lesions (CORAL) study: rationale and methods. *J Vasc Interv Radiol* 2005;16:1295–300.
41. Morrison J, Biro F, Campaigne B, *et al*. Obesity and cardiovascular disease risk factors in black and white girls: the NHLBI Growth and Health Study. *Am J Public Health* 1992;82:1613–20.
42. Carey VJ, Bishop L, Charleston J, *et al*. Rationale and design of the Optimal Macro-Nutrient Intake Heart Trial to Prevent Heart Disease (OMNI-Heart). *Clin Trials* 2005;2:529–37.
43. Wadden TA, Volger S, Sarwer DB, *et al*. A two-year randomized trial of obesity treatment in primary care practice. *N Engl J Med* 2011;365:1969–79.