Treatment Options for type 2 Diabetes in Adolescents and Youth Observational Follow-Up Study (TODAY2)

Data Release Documentation

April 15, 2022 Full Scale Data Release

Prepared by the TODAY2 Coordinating Center

The George Washington University Biostatistics Center
6110 Executive Blvd, Suite 750
Rockville, MD 20852

Telephone: (301) 881-9260
Fax: (301) 881-3767

Sponsored by
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health

Table of Contents

1 INTRODUCTION

1.1 General Information

TODAY2 was a non-interventional observational follow-up study of participants originally enrolled in the TODAY study. Briefly, TODAY was a multi-center randomized clinical trial designed to evaluate the safety and efficacy of three treatment regimens for type 2 diabetes (T2DM) in children and youth based on glycemic control. The three treatment regimens were: (1) metformin alone, (2) metformin plus rosiglitazone, and (3) metformin plus an intensive lifestyle intervention. The clinical trial recruited and followed participants for a minimum of two years. Participants were randomized within two years of the diagnosis of T2DM. The primary outcome of the clinical trial was time to treatment failure as defined in one of two ways:

1. **HbA1c ≥ 8% over a 6-month period.** All regularly scheduled HbA1c values must be ≥ 8% over a 6-month period. If any single value is < 8%, after which HbA1c re-elevates to ≥ 8%, the clock will restart at the time of the re-elevation. At least two consecutive measurements must be ≥ 8% over 6 months.
2. **Inability to wean from temporary insulin therapy due to metabolic decompensation.** Participants who experience metabolic decompensation requiring temporary use of insulin, who cannot safely be weaned from insulin within three months, will be classified as treatment failures.

The TODAY study was conducted from 2004-2011. At the conclusion of TODAY, the cohort continued to be followed in TODAY2 using a two-phase approach. During the first phase (2011-2014), which occurred immediately post-intervention, participants who enrolled had the option to continue to have their diabetes care managed by a TODAY clinical center or by a community physician. Those managed by the TODAY study team followed a standard treatment algorithm based on the current standard of care which involved glycemic management with metformin and/or insulin and were evaluated quarterly. Those who sought care in the community were only evaluated annually. During the second phase of TODAY2 (2014-2020), the diabetes management for all participants was transitioned to the community with annual in person examinations.

This document describes the complete public release of TODAY2 data. Data collected during the TODAY randomized clinical trial (2004-2011) are archived and available (released June 2014) to the public through the NIDDK Repository: https://repository.niddk.nih.gov/studies/today .

1.2 Data Collection Schedule

The table lists the schedule of data collection, measurements, and assessments included in the TODAY2 data and specimen repository.

Data Collection		
Measurement/Assessment	**Phase 1 (2011-2014)**	**Phase 2 (2014-2020)**
Historical data (a)	Annually	Annually
HbA1c	Quarterly	Annually
Blood for storage	Annually	Annually
Urine for storage	Annually	Annually
Fasting glucose, insulin, C-peptide, and proinsulin	Annually	Participant study years 6 and 9
OGTT (b)	Annually	Participant study years 6 and 9
Serum creatinine (c)	Annually	Annually

Data Collection		
Measurement/Assessment	**Phase 1 (2011-2014)**	**Phase 2 (2014-2020)**
LFTs	Annually	Annually
BMI	Quarterly	Annually
Blood pressure	Quarterly	Annually
Lipids (d)	Annually	Annually
Other laboratory values (e)	Annually	First annual visit
Physical exam (f)	Quarterly	Annually
Diabetes management	Quarterly	Annually
Diabetes complications	Quarterly	Annually
Interim history	Quarterly	Semi-annually
Psychosocial and QoL (g)		Annually
Life stressor (g)	Annually	Annually
Healthcare usage	Annually	Annually
Cardiovascular risk factors (h)	Annually	
Peripheral neuropathy (i)	Annually	Annually
Microalbuminuria, macroalbuminuria	Annually	Annually
Hypertension	Quarterly	Annually
Dyslipidemia (LDL, Triglyceride)	Annually	Annually
Retinopathy screening		Once
Echocardiogram		Once
Arterial stiffness	Once	Once
Sleep (j)		Once
Other misc. questionnaires (k)		Final visit
Pregnancy (l)	As needed	As needed

(a) Historical data include family and medical history, and demographics (including socioeconomic status).

(b) 75g oral glucose tolerance test (OGTT) with measurements of glucose, insulin, and C-peptide concentrations at 0, 30, 60, 90 and 120 minutes during Phase 1 and at 0 and 30 minutes during Phase 2. Samples for stimulated insulin were not collected once a participant began treatment with insulin.

(c) Serum creatinine is used to calculate creatinine clearance (calculated via the Cockcroft-Gault formula).

(d) Lipids included LDL, HDL, total cholesterol, and triglycerides. Through 5 years after enrollment, lipoproteins were also measured.

(e) Other laboratory assays were performed annually through the first year of Phase 2. For a full listing, see Section 5.5.1.

(f) A comprehensive physical exam including blood pressure, height, and weight was performed as indicated; Tanner staging was performed annually during Phase 1 only.

(g) The psychosocial and quality of life (QoL) battery includes the Beck Depression Inventory (BDI), Patient Health Questionnaire (PHQ), Eating Disorder Diagnostic Scale (EDDS), and Pediatric Quality of Life Inventory (PEDsQL). Life stressors were assessed annually via the Annual Life Events Inventory during both phases.

(h) Cardiovascular risk factors include fibrinogen, c-reactive protein, homocysteine, plasminogen activator inhibitor-1, and interleukin-6.

(i) Assessed using the Michigan Neuropathy Screening Instrument (MNSI) and monofilament. In the final year a graduated tuning fork and sharp sensation test were added.

(j) Survey instruments include the Epworth Sleepiness Scale, Berlin Sleep Questionnaire, Pittsburgh Sleep Quality Index, Horne-Ostburg Morningness-Eveningness Questionnaire, and polysomnograms.

(k) Other miscellaneous questionnaires include the Diabetes Distress Scale Questionnaire, the Material Needs Insecurities Scale Questionnaire, and the Offspring questionnaire.

(l) Pregnancy, post-delivery, and perinatal information extracted from obstetric and pediatric medical records was obtained when a pregnancy was reported.

1.3 Randomization

Randomization during the TODAY clinical trial was stratified by clinical center to ensure balance among the treatment groups with respect to anticipated differences in the participant populations. Randomization began in July 2004 and ended in February 2009. Details on randomization design and implementation are included in the TODAY data release. The TODAY2 study was designed as an observational follow-up of the original TODAY cohort.

2 DATA RELEASE INFORMATION

2.1 General Information

- No personal identifying information is included.
- Variables collecting the same information in TODAY and TODAY2 have consistent variable names other than the visit number which is changed in TODAY2 to denote the different phase (i.e., variable WEIGHT in the TODAY data release has the same variable name – WEIGHT - in this TODAY2 data release).
- The variable RELEASEID uniquely identifies each participant. It consists of a 2-digit study identifier (65), followed by a random 5-digit identifier which uniquely identifies the participant. RELEASEID is consistent for each participant throughout the TODAY and TODAY2 datasets.
- The variable DAYS represents the number of days between randomization and the time of measurement/assessment. No dates or specific time points are included in any of the datasets.
- The variable PVISIT is a four character value that identifies the visit the measurements/assessments were taken. PVISIT and RELEASEID are used to match a participant's information across multiple datasets. The variable starts with the letter 'P' followed by the visit month at which the measurements/assessments were taken (i.e., P075 represents a month 75 study visit).
- Data for all participants who were part of the TODAY2 cohort are included, regardless of whether they enrolled in Phase 1 only, Phase 2 only, or both phases.

  Exception: Data from 22 participants with monogenic diabetes mutations who were previously clinically diagnosed to have T2DM are excluded from ALL datasets, including those provided in the addendum.
- In accordance with HIPAA regulations and to protect the identification of TODAY participants, the data has been modified to ensure that no participant is identifiable. For example, data was sorted into small clearly-identifiable groups (sex*race) and collapsed if the sample size was small.
- Only research data are included in the released dataset, including data from all clinical visits, laboratory data, and special tests. Non-research data, including tracking forms, are not included. Adverse event and serious adverse event data were collected but are not included in the data release outside of the reporting of complications and comorbidities of diabetes and death.
- All available data from each form and central unit database are included.

2.2 Data Location

Data are released from the TODAY2 Coordinating Center at the George Washington University Biostatistics Center to the Data Repository at the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) of the National Institutes of Health.

2.2.1 Structure of the SAS Data Files

- Multiple datasets are available as transport files in the TODAY2 Release library. One transport file exists for each form or dataset.
- The files are included as SAS datasets within transport files with the same name as the embedded form or dataset name and the extension XPT. The SAS code to import each dataset is given below:

```
libname today2 'directory for the destination of SAS datasets on
your host';
filename tranfile 'name of the transport files on your host';
proc cimport data=today2.data infile=tranfile; run;
```

For example to import file TODAY2.BDI:

```
libname today2 'c:\mysasfiles';
filename tranfile 'c:\myxptfiles\bdi.xpt';
proc cimport data=today2.bdi infile=tranfile; run;
```

- The contents of variables in these datasets are provided via two means:
  - Form data have a companion form to the dataset that details the variable names and coding.
  - For non-form data, a listing of variable names, descriptions, and coding is included.

2.3 De-identified Data

The TODAY2 datasets were de-identified in the following manner. All personal identifiers were removed, including participant ID and other personal identifiers (date of birth, etc.), clinic ID, and all dates. In addition, variables that might identify a particular individual were collapsed into wider groupings. For example, multiple participant's total income categories were consolidated into the income categories of <$5,000, $5,000-$34,999, and $35,000.

BMI was also adjusted to protect the participant's identity. Where possible actual values were reported but those below or above a certain cut point had their actual data collapsed into a group that contained all individuals that also met that criterion. The upper and lower cut points for BMI are as follows:

Measure	Lower Cut Point	Upper Cut Point
BMI (kg/m^2)	≤28	>46

2.4 Structure of the Datasets

The structure of the datasets can vary. Some datasets include one record per participant (i.e., comorbidity and complication dataset, diabetes distress scale questionnaire, echocardiogram dataset). Other datasets include one record per participant per visit/assessment (i.e., visit dataset, laboratory dataset, BDI, PHQ, PWV). There are also special datasets that are structured according to the nature of the assessment. For example, the fundus/OCT datasets include multiple records per participant, one for each eye. Similarly, the pregnancy dataset includes multiple records per female participant, one per pregnancy and infant reported during the study. Sections 5-6 describe the data included in detail, including information on the structure of the dataset.

Most of the datasets included in this release contain data that were obtained during TODAY2 (2011-2020) only; however some datasets may also include information from the TODAY randomized clinical trial if those data were not included in the prior release. For example, data reporting on the development of complications and comorbidities, which were the main outcomes in TODAY2 (i.e., hypertension diagnosis, dyslipidemia diagnosis), were not part of the TODAY data release but are included now. In addition, data collected during TODAY but analyzed/assayed after the end of the randomized trial are included in this release, such as laboratory assays performed on blood or urine storage samples post-2011 (i.e., lipoproteins, sex hormones, adiponectin, cardiac biomarkers). This also includes TODAY data re-analyzed using new methodologies (i.e., DXA and OCT scans). Sections 5-6 describe the data included in detail, including specific information on whether the dataset includes TODAY2 data only or data from both TODAY and TODAY2.

The number of participants participating in each follow-up visit decreased over time. This is due to two reasons: 1) loss to follow-up and 2) staggered entry into the TODAY study (2004-2009), which means that the earliest randomized participants had visits >24 months during TODAY, while the latest randomized participants had these visits during TODAY2. The frequency of regularly-scheduled follow-up visits also differed over time (quarterly through 2014 and annually thereafter), which affected the number of participants participating at a given study visit.

RELEASEID is used to uniquely identify participants, and for select datasets where it applies, the PVISIT is used to identify the visit. The table below shows the number of participants at each regularly-scheduled follow-up visit in TODAY2. Note: earlier study visits can be found in the TODAY data release (variable: MVISIT).

Number of participants who completed in-clinic visits

Visit	N	Visit	N
P027	14	P081	232
P030	42	P084	458
P033	67	P087	183
P036	89	P090	174
P039	104	P093	155
P042	122	P096	448
P045	140	P099	109
P048	192	P102	92
P051	194	P105	66
P054	225	P108	451
P057	232	P111	20
P060	305	P114	3
P063	267	P120	449
P066	284	P132	443
P069	256	P144	368
P072	402	P156	271
P075	249	P168	175
P078	279	P180	69

2.5 Guidelines for Merging TODAY2 with TODAY Data

The variable RELEASEID, representing the unique participant ID, is common to all datasets (TODAY and TODAY2) and can be used to link records.

Datasets including multiple records per participant for each data collection visit can be merged using RELEASEID and PVISIT.

Datasets including variables with the same name, with the exception of the visit identifier (MVISIT or PVISIT), can be concatenated together using the SET statement in SAS. For example, the TODAY.CBL dataset included in the TODAY data release can be concatenated with the TODAY2.CBL dataset from this release. Similarly, the TODAY.BDI dataset included in the TODAY data release can be concatenated with the TODAY2.BDI dataset as they both contain the same variables. The MVISIT and PVISIT character variables can be converted to numeric values as needed for purposes of analysis.

Similar merging processes can be used for the lipoprotein and pregnancy datasets. The visit indicator in these datasets (LPMONTH in lipoprotein and CLOSVISIT in pregnancy) can be used to merge records with datasets containing a MVISIT or PVISIT indicator as needed, along with RELEASEID. Conversion of the visit indicator to a numeric variable may facilitate the merging process between records.

Datasets for special procedures (i.e., ECHO, PWV, OCT, FUNDUS) that were assessed twice during the study are not associated with a visit as they were collected during a set period of calendar time. For those datasets, the TIMEPOINT variable indicates whether the records correspond to the first assessment or the second assessment. The TIMEPOINT variable can be used to concatenate the TODAY and TODAY2 datasets. The variable DAYS represents the number of days between randomization and the time of the assessment.

3 COMPLICATIONS AND COMORBIDITIES DEFINITIONS

The primary aims of the TODAY2 study were to evaluate the long-term rates of development of complications and comorbidities in youth-onset T2DM and identify the associated risk factors. Primary and secondary outcomes were evaluated through rigorous study assessments (e.g., hypertension, LDL dyslipidemia, moderately increased albuminuria) or through a thorough review of medical records (e.g., myocardial infarction, pancreatitis, stroke). During TODAY2, we also continued to track glycemic control for participants who did not meet the primary outcome in TODAY.

The methodology for determining each of these events is detailed below and can also be found in the Supplemental Information of the TODAY2 primary outcome paper (NEJM, 2021).

3.1.1 Hypertension

Blood pressure was measured after a 5-minute rest with the participant in a sitting position using a CAS 740 monitor with standardized oscillometric cuff sizes at every visit. Three measurements were taken at 1-minute intervals. The average of the 2nd and 3rd systolic and diastolic measurements was calculated to obtain blood pressure at that visit.

Hypertension was defined as blood pressure ≥ 95th percentile for age, sex and height or systolic blood pressure (SBP) ≥ 130mmHg and/or diastolic blood pressure (DBP) ≥ 80mmHg on three occasions with less than 2 years between the first and second and second and third measurements, regardless of interim blood pressure. There were two definitions for hypertension for individuals on blood pressure-lowering medications: 1) 1-2 elevated blood pressures, as defined above, less than 2 years apart with the last elevated pressure followed immediately by the start of pharmacologic treatment; 2) elevated blood pressure and previously prescribed anti-hypertensive medication or ACE inhibitors prescribed for elevated albumin.

During the TODAY randomized trial and Phase 1 of TODAY2, blood pressure was measured at each quarterly visit. During Phase 2 of TODAY2, blood pressure was measured at each annual visit. Hypertension was defined as blood pressure ≥ 95th percentile for age, sex and height or systolic blood pressure ≥ 130mmHG and/or diastolic blood pressure ≥ 80mmHG on two consecutive annual visits or a single elevated blood pressure and previously prescribed anti-hypertensive medication.

3.1.2 Dyslipidemia

Measurements of cholesterol in low density lipoprotein (LDL-C) from frozen plasma samples were performed at the Northwest Lipid Metabolism and Diabetes Research Laboratories (Seattle, WA) by

enzymatic assay on the Hitachi 917 autoanalyzer using methods standardized to the Centers for Disease Control and Prevention Reference Methods. LDL-C was calculated by the Friedewald equation. If triglycerides were >400 mg/dl, a complete lipoprotein separation by ultracentrifugation was performed using the Lipid Research Clinics Beta Quantification procedure. LDL-C dyslipidemia and triglyceride dyslipidemia were defined according to "Cardiovascular risk reduction in high-risk pediatric participants: a scientific statement" and updated in the Summary Report of the Expert Panel on Integrated Guidelines for Cardiovascular Health and Risk Reduction in Children and Adolescents.

LDL-C Dyslipidemia:
LDL-C dyslipidemia was defined as two consecutive values of LDL-C ≥ 130 mg/dL without an intervening LDL-C value < 130 mg/dL, irrespective of the time between consecutive values. There were two definitions for LDL-C dyslipidemia for individuals on cholesterol-lowering medications: 1) a single LDL-C ≥ 130 mg/dL followed immediately by the start of pharmacologic treatment; 2) a single LDL-C ≥ 130 mg/dL and previously prescribed lipid-lowering medication.

Triglyceride Dyslipidemia:
Triglyceride dyslipidemia was defined as two consecutive values of triglycerides ≥ 150 mg/dL without an intervening triglyceride value < 150 mg/dL, irrespective of the time between consecutive values. There were two definitions for triglyceride dyslipidemia for individuals on cholesterol-lowering medications: 1) a single triglyceride value ≥ 150 mg/dL followed immediately by the start of pharmacologic treatment; 2) triglyceride value ≥ 150 mg/dL and previously prescribed lipid-lowering medication.

Any Dyslipidemia:
Any dyslipidemia was defined as a diagnosis of either LDL-C dyslipidemia or triglyceride dyslipidemia as defined above.

3.1.3 Renal Disease

The immunochemical measurement of albumin in urine was performed at the Northwest Lipid Metabolism and Diabetes Research Laboratories (Seattle, WA) using Dade Behring reagent on a Behring Nephelometer. The target value to the assay calibrator was assigned by the reference material CRM470 prepared by the International Federation of Clinical Chemistry (IFCC). Concentrations of creatinine in urine were determined using the Creatinine Plus enzymatic Roche reagent on a Modular P analyzer (Roche Diagnostics, Inc., Indianapolis, IN); the results are traceable to the IDMS reference method.

Creatinine clearance was calculated by the Cockcroft formula using serum creatinine values. In addition, the Full Age Spectrum (FAS) combined serum creatinine and cystatin C equation, which has been validated in children and adults, was used to calculate estimated glomerular filtration rate (eGFR). eGFR was also estimated via the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation using serum creatinine values.

Moderately and Severely Increased Albuminuria:
Moderately increased albuminuria was defined as an urine albumin/creatinine ratio (ACR) ≥ 30 mg/g on at least 2 out to 3 visits (including a first-morning sample) within a 6-month period or 3 or more ACR values ≥ 30 mg/g with no drop in between, regardless of time duration, during TODAY and during the first 3 years of TODAY2 follow-up (Phase 1). Subsequently, moderately increased albuminuria was defined as an ACR ≥ 30 mg/g on two consecutive annual visits or an ACR ≥ 30 mg/g and previously prescribed ACE inhibitor. Severely increased albuminuria was defined similarly using an ACR ≥ 300 mg/g.

Kidney Disease:
Kidney disease was defined as either moderately or severely increased albuminuria as defined above.

Hyperfiltration:
Hyperfiltration was defined as an eGFR ≥ 135 ml/min/1.73m^2 at two consecutive visits.

Rapid eGFR Decline:
Rapid eGFR decline was defined as an eGFR decline > 3 ml/min/1.73m^2 per year and/or ≥ 3.3% at two consecutive visits.

3.1.4 Nerve Disease

Nerve disease was assessed by two methods: Michigan Neuropathy Screening Instrument (MNSI) exam and Semmes-Weinstein 5.07 10-gram monofilament. Neuropathy was defined by an abnormality on at least two consecutive exams.

Michigan Neuropathy Screening Instrument:
The MNSI examination is a validated screening tool for diabetic peripheral neuropathy consisting of 4-items administered by a health professional conducting a direct examination of each foot. Vibratory sensation is assessed using a 128-Hz tuning fork applied to the dorsal surface of the great toe. Scoring of the exam was as follows: 1) each foot with any physical abnormality received a score of 1; 2) each foot with an ulcer received a score of 1; 3) vibration sensation was considered normal (0 points) if the examiner felt vibration for less than 10 seconds after the participant stopped feeling the vibration; it was scored as impaired (0.5 points) if the examiner felt it for >10 seconds and absent (1 point) if the subject could not feel the vibration at all; and 4) reflexes at the ankle were considered normal (0 points); reduced (0.5 points) if they could only be elicited with the Jendrassic maneuver; and absent (1 point) if they could not be elicited at all. The total possible score was added across both feet (range 0 to 8 points). The MNSI exam was considered abnormal if the score (across both feet) was > 2 on at least two consecutive exams.

Monofilament Examination:
The monofilament examination consisted of applying a Semmes-Weinstein 5.07 10-gram monofilament to the dorsum of the great toe of each foot 10 times. Correct identification of at least 8 of the 10 applications on each foot was considered normal. The monofilament examination was considered abnormal if there were <8/10 correct responses on at least two consecutive exams.

Graduated Tuning Fork and Sharp Sensation Test:
This test consisted of a semi-quantitative evaluation of vibration and pinprick sensation using a Rydel-Seiffer 128 Hz Graduated Tuning Fork and Neuropen® device (US Neurologicals, LLC). Vibration assessments were performed twice on the great toe of each foot and scored 2 if the participant was unable to feel any vibration, 1 if vibration sensation was lost when the tuning fork indicator was ≤4, and 0 when the tuning fork indication was >4, for a total score of 0-8. A score of >5 was considered abnormal. Pinprick sensation was tested bilaterally using two Owen Mumford Neuropens. The blunt, “dull” Neurotip and exposed, “sharp” Neurotip were applied to the dorsum of the great toe midway between the nail fold and the DIP joint 8 times each and recorded. Incorrect identification on ≥5/8 of the trials indicated impaired pinprick sensation.

Nerve Disease:
Nerve disease (neuropathy) was defined as either an abnormal MNSI examination or abnormal monofilament results as defined above. Graduated tuning fork and sharp sensation test are not included in the nerve disease definition.

3.1.5 Eye Disease

Eye disease was assessed by 7-field stereoscopic fundus photographs taken by certified photographers during the final year of the TODAY randomized trial and again approximately seven years later during TODAY2. The fundus photographs were graded according to the Early Treatment Diabetic Retinopathy Study (ETDRS) protocol by graders at a centralized reading center at the University of Wisconsin masked to treatment group assignment using a 25-step scale representing the range of diabetic retinopathy in each eye. Based upon the grading in the worse eye, each participant was classified as having no retinopathy, very mild non-proliferative retinopathy (NPDR), mild NPDR, moderate NPDR, moderately severe NPDR, severe NPDR, early or stable, treated proliferative diabetic retinopathy (PDR), and high risk PDR. Additionally, Ocular Coherence Tomography (OCT) was utilized to determine the presence of clinically significant macular edema, and to assess retinal thickness across sectors and layers of the retina.

Eye Disease:
Eye disease was defined, for either eye, as either an ETDRS grade score ≥ 20 (very mild NPDR or above) or clinically significant macular edema.

3.1.6 Microvascular Disease

Microvascular disease was defined as any kidney, nerve, or eye disease event as defined in the sections above.

3.1.7 Extended Primary Outcome (Loss of Glycemic Control)

The primary outcome during the TODAY randomized trial was loss of glycemic control, which was strictly defined and adjudicated by an independent committee. During TODAY2, we continued to monitor glycemia in participants who had not lost control during the TODAY study. During TODAY, loss of glycemic control was defined as HbA1c ≥8% for 6 months or sustained metabolic decompensation requiring insulin. During TODAY2, loss of glycemic control was defined by two consecutive measures of HbA1c ≥8%.

3.1.8 Comorbidity Assessment

A rigorous adjudication process was established to track and document each reported comorbid event occurring outside of study visits and diagnosed by a participant's regular medical provider. During semi-annual visits, each participant underwent a structured interview by a medical provider to assess if any relevant events had occurred since the previous visit. Every affirmative answer initiated a review process that included obtaining and submitting relevant medical records to the Coordinating Center. A Comorbidity Assessment Committee (CAC), comprised of physician experts selected from the study group, was charged with reviewing all reported comorbidities. CAC was further divided into seven sub-committees: heart, vascular, cerebrovascular, renal, nerve, eye, and liver/pancreas/gallbladder. The assigned member of CAC assessed each submitted event utilizing criteria established a priori based on national guidelines. If questions arose, the case was discussed by the full CAC. In some cases, participants reported the same medical event at different visits. This resulted in some repeated records for the same event in the same participant; if this occurred, the case was re-evaluated by CAC. Discretion is advised to avoid over counting events that are non-transient (e.g., congestive heart failure).

Detailed descriptions and definitions for the adjudicated comorbid medical events are provided below and can also be found in the Supplemental Information of the TODAY2 primary outcome paper (NEJM, 2021).

Definitions

Heart Events

Arrhythmia: Evidence of electrocardiogram (EKG), Holter monitor, or event monitor having been performed with the results reviewed and interpreted by a cardiologist and available in the medical record, or the insertion of a pacemaker or implantable cardioverter defibrillator.

Coronary Artery Disease (CAD): ≥ 50% occlusion of any coronary artery by angiography or the report of a coronary artery bypass grafting or revascularization procedure.

Congestive Heart Failure (CHF): Hospital admission with a principal diagnosis of heart failure, pulmonary edema by chest X-ray, or symptoms, signs, or physical findings consistent with CHF.

Left Ventricular Systolic Dysfunction: Mild, moderate, or severe dysfunction based upon ejection fraction from angiogram or through noninvasive testing by echocardiogram, magnetic resonance, computed tomography, or nuclear test.

Myocardial Infarction (MI): Cardiologist interpretation of a definitive MI on EKG and/or significantly elevated enzymes. The enzymes considered were troponin T and troponin I elevated to a value that indicates myocardial necrosis in the laboratory performing the test and/or CPK and CPK-MB elevated to twice the upper limit of normal for the laboratory performing the test.

Vascular Events

Peripheral Artery Disease/Vascular Insufficiency (PAD): Ankle brachial index ≤ 0.9, an imaging study (angiogram, Doppler ultrasound (US), magnetic resonance angiogram (MRA), or computed tomography (CT) scan) demonstrating > 50% stenosis in any peripheral artery (subclavian, femoral, iliac); amputation of an extremity for severe arterial vascular insufficiency; or vascular surgery for reconstruction, bypass, or percutaneous revascularization in the arteries of the lower and upper extremities.

Renal Artery Disease: Imaging study (angiogram, Doppler US, MRA, or CT scan) demonstrating > 50% stenosis in either renal artery or vascular surgery for reconstruction, bypass, or percutaneous revascularization in the renal arteries.

Deep Vein Thrombosis (DVT): Doppler US demonstrating a non-compressible vessel or an imaging study (angiogram, Doppler US, MRA, or CT scan) demonstrating the presence of a clot.

Cerebrovascular Events

Stroke: Acute episode of neurological dysfunction caused by focal or global brain, spinal cord, or retinal vascular injury as a result of hemorrhage or infarction confirmed by angiogram, computed tomography, magnetic resonance imaging, neurosurgery, or autopsy.

Cerebrovascular Disease in Absence of Stroke: Carotid ultrasound or angiogram demonstrating ≥ 50% narrowing of one or more carotid arteries or any major extracranial or intracranial vessels to the brain; cerebral (e.g., carotid) or cervical artery revascularization surgery; or percutaneous intervention.

Transient Ischemic Attack (TIA): Clinical diagnosis in clinic, emergency department, or hospital records or symptoms in the absence of infarct on imagining studies (angiogram, US, MRI, or CT).

Renal Events

Chronic Kidney Disease (CKD): GFR < 60 mL/min/1.73 m^2 for at least 3 months.

End Stage Renal Disease (ESRD): Kidney dialysis or kidney transplant.

Nerve Events

Diabetic Peripheral Neuropathy (DPN): Nerve conduction velocity test result consistent with DPN or diagnoses by a neurologist. If medical records indicated a diagnosis of DPN by a physician other than a

neurologist the full medical record were reviewed by the assigned CAC member and diagnosis confirmed on recorded history and physical exam findings.
Autonomic Neuropathy: Diagnosis of autonomic neuropathy from cardiology records (if diagnosis was by a physician other than a cardiologist, affirmation by the assigned CAC member based on full record review was required); gastroparesis diagnosed by a gastroenterologist; or gastric emptying study or test procedure (sphincter manometry, orthostatic or postural hypotension, urodynamics test, or post-void residual test).
Mononeuropathy: Mononeuropathy diagnosed by a neurologist. If medical records indicated a diagnosis of mononeuropathy by a physician other than a neurologist the full medical record were reviewed by the assigned CAC member and diagnosis confirmed on recorded history and physical exam findings

Eye Events
Non-Proliferative Diabetic Retinopathy (NPDR): Diagnosis of NPDR by an ophthalmologist or optometrist using funduscopic examination or fundus photography.
Proliferative Diabetic Retinopathy (PDR): Diagnosis of PDR by an ophthalmologist or optometrist including using funduscopic examination or fundus photography, or documentation of laser therapy, injection therapy, vitrectomy, and/or intravitreal injection for treatment of proliferative retinopathy.
Macular Edema (ME): Diagnosis of ME made by an ophthalmologist or optometrist using funduscopic examination or fundus photography or laser therapy, injection therapy, and/or intravitreal injection for treatment of macular edema.
Vitreous Hemorrhage: Diagnosis of vitreous hemorrhage by an ophthalmologist or optometrist using funduscopic examination or fundus photography.
Blindness Due to Diabetes: Diagnosis of blindness due to diabetes by an ophthalmologist or optometrist using a visual acuity test. Due to diabetes means directly due to or as a complication of treatment of retinopathy, macular edema, or vitreous hemorrhage.
Cataracts: Diagnosis of cataracts by an ophthalmologist or optometrist or documented cataract surgery.
Glaucoma: Diagnosis of glaucoma by an ophthalmologist or optometrist or documented laser surgery.

Liver, Pancreas, and Gallbladder Events
Cirrhosis: Nodular liver surface contour or other markers of cirrhosis from imaging; evidence of portal hypertension (e.g., varices) from imaging or endoscopy; evidence of cirrhosis from liver biopsy; or documentation of liver transplant.
Pancreatitis: Two of the following: (1) symptoms of abdominal pain, severe epigastric pain, or other clinical presentation; (2) characteristic findings of pancreatitis on MRI, CT scan, or transabdominal US; (3) an amylase or lipase value > 3 times the upper limit of normal.
Gallbladder Disease: Gallstones detected during gallbladder imaging or endoscopic retrograde cholangio-pancreatography or gallbladder surgery.

Other Events
In addition to the comorbid event data assessed by CAC, TODAY2 collected information on additional targeted adverse events. For each identified event below, medical records were obtained by each clinical site and the information necessary to confirm the event was extracted, reviewed, and recorded.

Bariatric Surgery: Bariatric surgery including type of procedure, complications, length of hospital stay, and need for additional surgery.
Diabetic Ketoacidosis (DKA): Hospitalization or emergency department diagnosis of DKA (HCO_3- < 15 mEq/L or pH < 7.3 mM and moderate or large urine ketone or serum ketone ≥ 4.0 mM).
*Hyperosmolar Hyperglycemic Syndrome (HHS): G*lucose concentration > 600 mg/dL, serum osmolality > 330 mOsm/kg, and serum bicarbonate concentration > 15 mEq/L, and urine ketone concentration < 15 mg/dL.

Fractures: Fractures confirmed by x-ray or imagining test report with the date, location, and cause of the fracture noted.
Sleep Apnea: Sleep study or polysomnogram performed overnight in a sleep lab with documented apnea-hypopnea index ≥ 5.0.
Lower Extremity Ulcers: Diagnosis of ulcer (not including necrobiosis alone), along with location (foot/lower leg, right/left).
Cancer: Diagnosis of cancer, including type and location.

4 STATISTICAL CONSIDERATIONS

A general overview of statistical analysis considerations is provided below. Further details can be found in the published study manuscripts.

4.1 Time to Complication or Comorbidity

For select complication and comorbidity events (e.g., hypertension, dyslipidemia), we enumerated the amount of time between randomization and the event. This is represented as DAYSTOxxx in all release datasets (i.e., DAYSTOHTN, DAYSTOLDL). If the participant had the event, DAYSTOxxx represents the time in days between randomization and the event. If the participant did not experience the event, DAYSTOxxx represents the number of days between randomization and the last visit when the measure was assessed (i.e., censoring time). The time to event variables are defined such that when a participant first experiences the event, they are forever classified as having the event, irrespective of event resolution.

4.2 Primary Analysis

Kaplan-Meier estimates were used to estimate cumulative incidences for the time to the first occurrence of complications and comorbidities outcomes. The outcomes were analyzed individually or as aggregate outcomes (i.e., any microvascular events was evaluated as the first occurrence of any kidney, nerve, or eye disease event). Associations with risk factors were evaluated using univariate and multivariate Cox proportional hazards regression models. Participants with the event at baseline were excluded from time-to-event analyses. The proportional-odds cumulative logit model for ordinal data was used to evaluate the association between the number of complications and the risk factors.

4.3 Repeated Measures

To account for repeated measurements over time, generalized linear mixed models were used to assess covariate effects on the mean of each quantitative risk factor over repeated time points, and generalized estimating equation models were used to assess effects on the prevalence of each binomial risk factor over repeated time points. The TODAY/TODAY2 visit number was included as a class effect. Covariates measured repeatedly over time entered the models as fixed or time-dependent covariates, as appropriate.

Multivariable Cox proportional hazards regression models were used to estimate the effect of risk factors (entered as fixed or time-dependent covariates in the models) on the risk of time-to-event outcomes (i.e., hypertension). The models used time to the first occurrence of any time-to-event outcomes. Participants with the event at baseline were excluded from time-to-event analyses.

Multivariable linear regression or logistic regression models were utilized to analyze risk factors in relation to continuous and binary outcomes evaluated twice during the study (i.e., FUNDUS, ECHO, PWV). Slopes were estimated from repeated measures linear regression models to represent the change over time between the first and second assessment.

5 FILE DESCRIPTIONS

5.1 Data Forms

5.1.1 General

Each form is available as a PDF for use in approved data-release analyses only. Instructions for completing each form are included on each form. The TODAY2 form identifier can be found at the top left corner of the page along with the form name. The forms included in the data release have been modified to match the data provided in the datasets.

Data entry included responses in both data boxes and checkboxes on the data collection forms. In general, 'other' responses with write-in information were removed from the database and from the form in order to protect against the identification of the participant.

Over the course of TODAY and TODAY2 some forms remained fixed while others were modified. The forms included with this data release represent the final form versions. Any question or set of questions that was not collected in a prior form version is indicated on the form or in the form section below.

5.1.2 Variable Names on Data Forms

- Variable names for each released dataset are embedded in blue on the data form.
- All datasets are HIPAA compliant. Information that might identify a specific participant has been excluded from the datasets and questions that captured this information have been removed from the forms.
- Coding and formats for all variables are found on the original data form except where described below.
- The numerical value for checkbox style categorical variables is noted to the lower right of the checkboxes on the form.
- All free text information that was written-in on forms has been excluded from the datasets. In some cases, the write-in information has been recoded into broader categories.

5.2 Datasets for Non-Form Data

Data not collected on forms but for which datasets are included in this release are as follows:

Created Datasets: Data in these created datasets were not included as part of the June 2014 TODAY repository release. They include data from both TODAY and TODAY2.

- Adjudicated medical events (AME): One record per participant per adjudicated medical event diagnosed.
- Comorbidities and complications (COMORB): One record per participant.
- Targeted medical events (TME): One record per participant per targeted medical event.

Central Unit and Core Datasets: Data in these datasets include data from TODAY2 only, or from TODAY and TODAY2 as specified.

- Laboratory data (CBL): One record for each participant per data collection schedule. Includes data from TODAY2 only.

- Echocardiogram (ECHO): One record per participant with results from the TODAY2 echocardiogram assessment. Includes data from TODAY2 only.
- Eye Exam (FUNDUS): One record per participant and per eye. Includes data from TODAY2 only.
- Liposcience (LIPO): One record for each participant per data collection schedule. Includes data from both TODAY and TODAY2.
- Optical Coherence Tomography (OCT): One record per participant per eye and per data collection visit. Includes data from both TODAY (regraded scans) and TODAY2.
- Polysomnogram (PSG): One record per participant with data from the TODAY2 polysomnogram assessment.
- Arterial stiffness - Pulse Wave Velocity and Heart Rate Variability (PWV): One record per participant per arterial stiffness assessment. The arterial assessment was conducted twice, 5 years apart, during TODAY2.
- Speckle tracking and strain (SPECKLE): One record per participant per echocardiogram assessment. Includes data from both TODAY and TODAY2.

Addendum Datasets: Data in these created datasets were not included as part of the June 2014 TODAY data release, but solely include data from TODAY.

- Additional laboratory measures (ADDCBL): One record per participant per data collection schedule. Obtained from storage blood and urine samples. Includes data from TODAY only.
- Re-analyzed Dual X-ray absorptiometry (DXA): One record per participant per data collection schedule. Re-analysis of Hologic and GE scans collected during TODAY. Includes data from TODAY only.

More details are provided in Sections 5.4, 5.5, and 6.

5.3 Data Collection Forms

This section pertains to forms either completed by the participant or by trained study staff. Unlike some of the datasets for non-form data described in the previous section (Section 5.2), all the form data in this section were collected during TODAY2 only, with the exception of the pregnancy data, which includes data from both TODAY and TODAY2.

5.3.1 TODAY2.BDI: Beck Depression Inventory (BDI)

This form was self-administered and used to collect information about symptoms of depression. The form was completed by the participant according to the instructions in the BDI-II manual. This form was collected at each annual visit after 2014 and at the final study visit. The value of PVISIT will be any annually regularly scheduled assessment visit number. In the rare case that a visit was conducted remotely, this form was not completed.

In order to expedite scoring, Questions 17 (sleeping pattern) and 19 (change in appetite) on the original form were recoded into a smaller number of categories, with possible response options 0, 1, 2, or 3. All other questions were unchanged and are provided as recorded. The same procedure was used for the archival of the TODAY data.

5.3.2 TODAY2.BERLIN: Berlin Sleep Questionnaire

This form was self-administered and used to identify those at risk for sleep apnea. Questions on height, weight, age, and gender were removed from the standard questionnaire as this information was collected elsewhere. The form was administered once at an annual visit between 2015 and 2017. The responses on this form were unchanged and are provided as recorded.

5.3.3 TODAY2.DDS: Diabetes Distress Scale Questionnaire

This form was self-administered and used to evaluate diabetes distress. The form was administered once at the final study visit. The responses on this form were unchanged and are provided as recorded.

5.3.4 TODAY2.EATING: Eating Behaviors Questionnaire

This form was self-administered and used to collect information about symptoms of anorexia nervosa, bulimia nervosa, and binge eating. The form was completed by the participant according to the instructions in the Eating Disorder Diagnostic Scale (EDDS) manual. This form was collected at each annual visit after 2014 and at the final study visit. The value of PVISIT will be any annually regularly scheduled assessment visit number.

The EDDS questionnaire is a validated questionnaire based on the Diagnostic and Statistical Manual, Fifth Edition (DSM-5) (American Psychiatric Association, 2013) criteria. Questions on weight, sex, age were removed from the standard questionnaire as this information was collected elsewhere as part of the standard visit. The responses on this form were unchanged and are provided as recorded.

5.3.5 TODAY2.EPWRTH: Epworth Sleepiness Questionnaire

This form was self-administered and used to evaluate the participant's general level of daytime sleepiness or average sleep propensity in daily life. The form was administered once at an annual visit between 2015 and 2017. The responses on this form were unchanged and are provided as recorded.

5.3.6 TODAY2.HEALTH: Healthcare Usage Questionnaire Version 1

This form was interview-administered and used to evaluate healthcare coverage and utilization. The wording and structure of the questions included on the form were adapted from questions in the Household Component of the Medical Expenditure Panel Survey (MEPS-HC). The form was administered every six months between 2015 and 2017. The form reports on practices in the last six months, including routine medical care, emergency medical care, healthcare access, barriers to care, and insurance coverage. Annual data regarding routine non-diabetes care and healthcare coverage from 2011 to 2015, while originally collected on a separate form, are also included.

5.3.7 TODAY2.HEALTH2: Healthcare Usage Questionnaire Version 2

This form was interview-administered and used to evaluate healthcare coverage and utilization. HEALTH2 is based on the original HEALTH form, but consists of additional instructive scripts, revised language, and new questions. The form was administered at annual data collection visits starting in 2017. The form reports on practices in the last six or twelve months, including sources of healthcare, healthcare usage, insurance coverage, and healthcare-associated costs.

5.3.8 TODAY2.LIFE: Life Stressor Questionnaire

This form was self-administered and used to evaluate sources of stress experienced by the participant over the past year. The form was administered at all annual visits. The responses on this form were unchanged and are provided as recorded.

5.3.9 TODAY2.MEQ: Morningness-Eveningness Questionnaire

This form was self-administered and used to evaluate circadian rhythm type. The form was administered once at an annual visit between 2015 and 2017. The responses on this form were unchanged and are provided as recorded.

5.3.10 TODAY2.MNI: Material Needs Insecurities Scale Questionnaire

This form was self-administered and used to assess material needs insecurities. The form was completed once at the final study visit. It is comprised of four categories, relating to medication, food, housing, and energy insecurity. The responses on this form were unchanged and are provided as recorded.

5.3.11 TODAY2.NEURO: Neuropathy Screening

This form was used to record the presence of neuropathy symptoms. It is comprised of three parts: The Michigan Neuropathy Screening Instrument (MNSI), the Semmes-Weinstein 5.70 10-gram monofilament test (SW-MF), and a semi-quantitative evaluation of vibration and pinprick sensation. Additional information on nerve disease assessment is provided in Section 3.1.4.

The first part of the form is the MNSI. Part Ia is the MNSI questionnaire (history), which was self-administered by the participant. Part Ib is the physical examination portion of the MNSI. Together they make up the full MNSI, which is a validated screening tool for detection of peripheral neuropathy. The second part is an additional exam, Semmes-Weinstein 5.70 10-gram monofilament, completed by study staff. The third and final part is the test to more quantitatively measure vibratory and pinprick sensation evaluated of vibration and pinprick sensation using a Rydel-Seiffer 128 Hz Graduated Tuning Fork and Neuropen® device (US Neurologicals, LLC). The exams in Part Ib, II, and III were completed by trained study staff (e.g., study physician, nurse practitioner, physician assistant).

Part Ia, PartIb/II, and Part III were originally separate forms, but to aid in the analysis of the data, were combined the parts for the current data release. Parts I and II of this form were administered at all annual visits and at the final study visit. Part III was administered only at the final study visit. The value of PVISIT will be any annually scheduled assessment visit number.

A question on the MNSI survey (Part Ia) asking about the occurrence of an amputation was removed from the original form due to the small sample size and potentially identifiable information. All other questions were unchanged and are provided as recorded.

5.3.12 TODAY2.OFFSP: Offspring Questionnaire

This form was used once at the final study visit to obtain information about the offspring of all participants. The form was interview-administered by trained study staff. There is one record per participant per child. The form captures birth outcomes as well as the current status of the child. If the participant had a PREG form, questions 5-10 were not completed on this form (data are represented in the TODAY2,PREG dataset). All data collected on this form are from participant report. The responses on this form were unchanged and are provided as recorded.

5.3.13 TODAY2.PEDSQA: Pediatric Quality of Life Inventory Adult Report

The adult questionnaire (PedsQL-A; ≥26 years) evaluates the participant's health-related quality of life. It was self-administered by the participant after introductory instructions from the administrator. If the administrator determined that the participant was unable to self-administer the PedsQL-A (e.g., due to illness, fatigue, reading difficulties), the PedsQL-A was read aloud to the participant. The administration followed the guidelines provided on the PedsQL website (www.pedsql.org). This form was collected at each annual visit after 2014 and at the final study visit. The responses on this form were unchanged and are provided as recorded.

5.3.14 TODAY2.PEDSQLT: Pediatric Quality of Life Inventory Teen Report

The teen questionnaire (PedsQL-T; 13-18 years) evaluates the participant's health-related quality of life. It was self-administered by the participant after introductory instructions from the administrator. If the administrator determined that the participant was unable to self-administer the PedsQL-T (e.g., due to illness, fatigue, reading difficulties), the PedsQL-T was read aloud to the participant. The administration followed the guidelines provided on the PedsQL website (www.pedsql.org). This form was collected at each annual visit after 2014 and at the final study visit. The responses on this form were unchanged and are provided as recorded.

5.3.15 TODAY2.PEDSQLYA: Pediatric Quality of Life Inventory Young Adult Report

The young adult questionnaire (PedsQL-YA; 19-25 years) evaluates the participant's health-related quality of life. It was self-administered by the participant after introductory instructions from the administrator. If the administrator determined that the participant was unable to self-administer the PedsQL-YA (e.g., due to illness, fatigue, reading difficulties), the PedsQL-YA was read aloud to the participant. The administration followed the guidelines provided on the PedsQL website (www.pedsql.org). This form was collected at each annual visit after 2014 and at the final study visit. The responses on this form were unchanged and are provided as recorded.

5.3.16 TODAY2.PEMD: Physical Measurements and Male Erectile Dysfunction

This form was used to obtain major physical examination observations including: HEENT (head, eyes, ears, nose, and throat), thyroid, lungs, heart, abdomen, extremities, skin, and neurologic (except MNSI). The form also includes a question related to male erectile dysfunction. The form was administered at each scheduled visit (quarterly or annual) during the first three years of TODAY2 (2011-2014) and at annual visits thereafter (2014-2020). The value of PVISIT will be any regularly scheduled assessment visit number. The form was completed by a study trained physician or nurse practitioner.

Items 2-5 were assessed during TODAY. Data for these items can be found in the TODAY.PE dataset included in the June 2014 release. All other items on this TODAY2.PEMD form were only assessed during TODAY2.

5.3.17 TODAY2.PHQ: Patient Health Questionnaire

This form was self-administered and used to evaluate symptoms of anxiety and depression. This form was collected at each annual visit after 2014 and at the final study visit. Only sections regarding depression (PHQ-8), panic disorders (PHQ-PD), anxiety (GAD-7), somatic symptoms (PHQ-15), and alcohol use were administered. The responses on this form were unchanged and are provided as recorded.

5.3.18 TODAY2.PREG: Pregnancy

This form was completed by trained study staff with data extracted from obstetric and pediatric medical records for each pregnancy, when permission to obtain the records was granted. In rare instances, some data were obtained from the participant directly in lieu of medical records. There is one record per participant per child. In the case of multiple gestations, the information on the pregnancy outcomes is repeated for each child, although there was only one pregnancy. The form contains pre-gestational, prenatal, pregnancy outcome, perinatal complication, and delivery information as well as information about live born neonates from birth and hospital stay. The data appears as collected other than for birthweight, birthweight z-score, and birthweight percentile which has the lower end of the range truncated to avoid identifying information.

There is no visit associated with the PREG form as it was collected only as needed and not at specified visit intervals. The nearest visit prior to pregnancy is noted on the form (variable CLOSVISIT) to aid in linking each pregnancy to the rest of the data.

5.3.19 TODAY2.PSQI: Pittsburgh Sleep Quality Index Questionnaire

This form was self-administered and used to evaluate sleep quality, sleep duration, and sleep timing. The standard form was modified slightly to collect information on work or school status at the time of the assessment and, if working, if the participant was working non-typical times. The form was administered once at an annual visit between 2015 and 2017. The responses on this form were unchanged and are provided as recorded.

5.3.20 TODAY2.VISIT: Clinical Visit Inventory

This form collects information on anthropometrics, blood pressure, concomitant medications, interval history, and menstrual and fertility status. Some measurements were only recorded at specified intervals, including healthcare access and family health history. The form was administered at each scheduled visit during TODAY2 – quarterly during Phase 1 (2011-2014) and annually in Phase 2 (2014-2020).

Values for height and weight were removed and replaced by BMI, with low and high values collapsed to protect the identity of the participant. Blood pressure was collected as a measurement taken 5 minutes after sitting, a second and third measurement after sitting for an additional minute each. The average of the second and third readings is reported. Information about specific medication usage was collapsed into broad categories of medications. Other prescription medications include antiarrhythmic, anticonvulsant, anticoagulant, antidepressant, anxiolytic, psychotropic, prescription eye drop, stimulant, and thyroid medications. Family health history was collected annually in the first three years, with a final inventory on parental health history taken at the last study visit. All other questions from the form have been removed since they were administrative in nature and/or related to safety.

5.4 Created Datasets

5.4.1 TODAY2.AME: Adjudicated Medical Events

This dataset includes one record per participant per adjudicated medical event diagnosed. Multiple records for the same participant indicate the event was diagnosed more than once (recurrent events). Recurrent events may include distinct events of the same type (i.e., new pancreatitis event reported at different visits) or unresolved events (i.e., same coronary artery disease event reported on repeated visits). For analysis purposes, first event occurrence may be considered for events that are clinically not

repeatable. All participants without any medical event diagnosed during the study are not included in this dataset. Information about the individual variables, their descriptions, and any associated coding are provided in the table below.

The majority of adjudicated comorbidity medical events occurred during the 2014-2020 study period. A limited number of events were diagnosed during the 2004-2014 study period. All events, regardless of when they occurred, are included in this dataset.

Variable Name	Description	Units / Coding
RELEASEID	Participant ID for NIDDK Database Repository	
AMENUM	Event occurrence number for diagnosed event (1st occurrence, 2nd occurrence, etc.).	1-5
DAYSTOAME	Days from randomization to diagnosed event	Days
AMETYPE	Type of event (broad disease categories)	1=Heart; 2=Vascular; 3=Cerebrovascular; 4=Liver; 5=Nerve; 6=Renal; 7=Eye; 8=Death
AMENAME	Specific event name	1=Arrhythmia 2=Coronary artery disease 3=Coronary heart failure 4=Left ventricular systolic dysfunction 5=Myocardial infarction 6=PAD/vascular insufficiency 7=Renal artery disease 8=Deep vein thrombosis 9=Stroke 10=Cerebrovascular disease in the absence of stroke 11=TIA 12=Pancreatitis 13=Gallbladder disease 14=Peripheral diabetic neuropathy 15=Autonomic neuropathy 16=Diabetic mononeuropathy 17=Chronic kidney disease 18=End stage kidney disease 19=Non-proliferative diabetic retinopathy 20=Proliferative diabetic retinopathy 21=Macular edema 22=Vitreous hemorrhage 23=Blindness due to diabetes 24=Cataracts 25=Glaucoma 26=Death

5.4.2 TODAY2.COMORB: Comorbidities

This dataset includes one record per participant. Definitions for classification of complications and comorbidities were established across the study phases (TODAY and TODAY2) incorporating differences in data collection frequency (see Section 3). Complications and comorbidities were not included in the TODAY data release therefore this dataset includes participant diagnoses encompassing the entire study spectrum, from randomization to study end (2004-2020). Detailed definitions of the complications and comorbidities can be found in Section 3 of this document.

Binary variables indicate whether the participant experienced the event at any point during the study (including baseline). Variables labeled as 'DAYSTOxxx" indicate the number of days between randomization and the event (for those participants who experienced the event) or between randomization and the last visit when the measure was assessed (i.e., censoring time). For composite outcomes (any dyslipidemia defined as presence of either LDL or triglyceride dyslipidemia), the 'DAYSTOxxx' variable represents the days from randomization to either event in the composite outcome, whichever occurred first.

Information about the individual variables, their descriptions, and any associated coding are provided in the table below.

Variable Name	Description	Units / Coding
RELEASEID	Participant ID for NIDDK Database Repository	
Hypertension		
HTN0	Baseline hypertension	1=Yes 0=No
HTN	Hypertension	1=Yes 0=No
DAYSTOHTN	Days from randomization to hypertension (or censoring date if no event)	Days
Dyslipidemia		
LDLDLP0	Baseline LDL dyslipidemia	1=Yes 0=No
LDLDLP	LDL dyslipidemia	1=Yes 0=No
DAYSTOLDL	Days from randomization to LDL dyslipidemia (or censoring date if no event)	Days
TGDLP0	Baseline triglycerides dyslipidemia	1=Yes 0=No
TGDLP	Triglycerides dyslipidemia	1=Yes 0=No
DAYSTOTG	Days from randomization to triglycerides dyslipidemia (or censoring date if no event)	Days
ANYDLP0	Baseline any dyslipidemia	1=Yes 0=No
ANYDLP	Any dyslipidemia	1=Yes 0=No
DAYSTOANYDLP	Days from randomization to any dyslipidemia (or censoring date if no event)	Days
Renal Disease		

MIC0	Baseline microalbuminuria	1=Yes 0=No
MIC	Microalbuminuria	1=Yes 0=No
DAYSTOMIC	Days from randomization to microalbuminuria (or censoring date if no event)	Days
MAC0	Baseline macroalbuminuria	1=Yes 0=No
MAC	Macroalbuminuria	1=Yes 0=No
DAYSTOMAC	Days from randomization to macroalbuminuria (or censoring date if no event)	Days
NEPHRO0	Baseline kidney disease	1=Yes 0=No
NEPHRO	Kidney disease	1=Yes 0=No
DAYSTONEPHRO	Days from randomization to kidney disease (or censoring date if no event)	Days
HYP0	Baseline hyperfiltration	1=Yes 0=No
HYP	Hyperfiltration	1=Yes 0=No
DAYSTOHYP	Days from randomization to hyperfiltration (or censoring date if no event)	Days
RAPID0	Baseline rapid eGFR decline	1=Yes 0=No
RAPID	Rapid eGFR decline	1=Yes 0=No
DAYSTORAPID	Days from randomization to rapid eGFR decline (or censoring date if no event)	Days
Nerve Disease		
DNE0	Baseline abnormal MNSI exam (score > 2)	1=Yes 0=No
DNE	Abnormal MNSI exam (score > 2)	1=Yes 0=No
DAYSTODNE	Days from randomization to abnormal MNSI exam (score >2) (or censoring date if no event)	Days
FILAM0	Baseline abnormal monofilament examination	1=Yes 0=No
FILAM	Abnormal monofilament examination	1=Yes 0=No
DAYSTOFILAM	Days from randomization to abnormal monofilament examination (or censoring date if no event)	Days
NEURO0	Baseline nerve disease	1=Yes 0=No
NEURO	Any nerve disease	1=Yes 0=No

DAYSTONEURO	Days from randomization to nerve disease (or censoring date if no event)	Days
Eye Disease		
RETINO	Eye disease (any diabetic retinopathy)	1=Yes 0=No
DAYSTORETINO	Days from randomization to eye disease (or censoring date if no event)	Days
Other		
MVD0	Baseline any microvascular disease	1=Yes 0=No
MVD	Any microvascular disease	1=Yes 0=No
DAYSTOMVD	Days from randomization to any microvascular disease (or censoring date if no event)	Days
NUMMVD	Number of any microvascular disease events	0 - 3
GLYC	Loss of glycemic control (extended primary outcome)	1=Yes 0=No
DAYSTOGLYC	Days from randomization to loss of glycemic control (or censoring date if no event)	Days

5.4.3 TODAY2.TME: Targeted Medical Events

This dataset includes one record per participant per targeted medical event. Multiple records for the same participant indicate the event occurred more than once (recurrent events). Recurrent events may include distinct events of the same type (i.e., new ulcers reported at different visits) or unresolved events (i.e., same ulcer reported on repeated visits). Study participants without any confirmed targeted medical event during the study are not included in this dataset.

Targeted medical events definitions are provided in Section 3.1.8. Because only one cancer was confirmed during the study, cancer is not reported. Information about the individual variables, their descriptions, and any associated coding are provided in the table below. The majority of targeted medical events occurred during the 2014-2020 study period. A limited number of events were confirmed during the 2004-2014 study period. All events, regardless of when they occurred, are included in the dataset.

Variable Name	Description	Units / Coding
RELEASEID	Participant ID for NIDDK Database Repository	
TMENUM	Event occurrence number for targeted event (1st occurrence, 2nd occurrence, etc.).	1-12
DAYSTOTME	Days from randomization to targeted event	Days
TMETYPE	Targeted event category	1=Bariatric surgery; 2=Problems with blood sugar/DKA; 3=Fracture; 4=Sleep (AHI index≥5); 5=Ulcer

5.5 Central Unit and Cores Datasets

5.5.1 TODAY2.CBL: Laboratory Data

This dataset includes laboratory results collected at in-person TODAY2 visits (quarterly during 2011-2014 and annually during 2014-2020). Blood and spot urine samples were obtained and processed immediately according to standardized procedures and shipped on dry ice for analysis at the TODAY central biochemical laboratory (Northwest Lipid Metabolism and Diabetes Research Laboratories, University of Washington, Seattle WA). Fasting blood draws (after a 10-14 hour overnight fast) were done at annual visits only. Blood and urine samples were not collected during pregnancy, lactation or immediately postpartum. The frequency of assessment of the laboratory measures varied over time during TODAY2:

- Hemoglobin A1c (HbA1c) was obtained at every visit.
- Lipids, alanine aminotransferase, aspartate aminotransferase, estimated creatinine clearance (calculated via the Cockcroft-Gault formula), serum creatinine values, urine albumin, urine creatinine, and very low density lipoprotein were obtained annually until the end of study.
- Fasting glucose and insulin, inflammatory markers, proinsulin, free fatty acid, fibrinogen, Apo-B, and LDL particles values were obtained annually through 2014.
- Cardiac biomarkers and vascular endothelial growth factor values were obtained annually through 2015. Those measures were obtained from storage samples collected only among participants who had granted permission for the use of their samples for future research.
- Potassium measurements were obtained at annual visits during 2011-2014 only.
- Serum Cystatin C and serum uric acid were obtained annually until the end of study. Those measures were obtained from storage samples collected only among participants who had granted permission for the use of their samples for future research.
- OGTTs were obtained from all participants after a 10-14 hour overnight fast at each annual visit during 2011-2014 and at participant study years (i.e., years since randomization) 6 and 9 during 2014-2020. 2-hour OGTTs were performed during 2011-2014 visits and 30-minute OGTTs during 2014-2020. Stimulated insulin was not collected from participants receiving exogenous insulin therapy.

Common OGTT-based variables (i.e., insulinogenic index, oral disposition index) that can be calculated based on the individual OGTT time points and have been reported in prior TODAY/TODAY2 publications are included. Creatinine clearance was estimated via the Cockcroft-Gault formula using serum creatinine values. In addition, estimates of glomerular filtration rates were calculated via the Full Age Spectrum (FAS) combined serum creatinine and cystatin C equation, which has been validated in children and adults, and via the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation using serum creatinine values. The laboratory results are outlined in the table below and are reported as provided from the laboratory.

Variable Name	Description	Units / Coding
RELEASEID	Participant ID for NIDDK Database Repository	
PVISIT	Release visit number	P27-P180 = Month 27 – Month 180
DAYS	Days from randomization to study visit	Days
Lab analyte		
HBA1C	Hemoglobin A1c	%
LDL	Low density lipoprotein cholesterol - without the contribution of IDL and Lp(a)	mg/dL

CHOL	Total cholesterol	mg/dL
HSCRP	High-sensitivity C-reactive protein	mg/dL
ALT	Alanine aminotransferase	U/L
AST	Aspartate aminotransferase	U/L
FFA	Free fatty acid	mEq/L
FIB	Fibrinogen	mg/dL
GLUCOSE	Fasting glucose	mg/dL
HOM	Homocysteine	µmol/L
LDLB	Apo B-100 in LDL fractions	mg/dL
LDLC	LDL cholesterol measures in LD particles only without the contribution of IDL and Lp(a)	mg/dL
PIN	Proinsulin	pM
RF	LDL Relative flotation rate (mean density of LDL particles)	
TRIG	Triglyceride	mg/dL
APOB	Total plasma apo B-100	mg/dL
CPEP	C-peptide	ng/mL
ESTCREATCLEAR	Estimated creatinine Clearance (Cockcroft-Gault formula)	mL/min
HDL	High density lipoprotein cholesterol	mg/dL
INS	Fasting Insulin	µU/mL
LDLCB	LDLC/ Apo B-100 in LDL fractions (LDLB) ratio	
SERUMCREAT	Serum creatinine	mg/dL
UALB	Urine albumin	mg/dL
UALBCREAT	Urine albumin creatinine ratio (multiply by 1000 for result in mg/g units)	
UCREAT	Urine creatinine	mg/dL
VB12	Vitamin B12	pg/mL
VLDL	Very low density lipoprotein	mg/dL
IL6	Interleukin 6	pg/mL
PAI1	Human plasminogen activator inhibitor-1	ng/mL
BNP	Brain Natriuretic Peptide	pg/mL
COPEPTIN	Copeptin	ng/mL
EPO	Erythropoietin	pg/mL
ESELECTIN	E-Selectin	pg/mL
FGF23	Fibroblast Growth Factor 23	pg/mL
ICAM1	Intercellular Adhesion Molecule 1	pg/mL
IGFBP1	Insulin-like Growth Factor Binding Protein 1	pg/mL
IL1	Inflammatory Cytokine IL-1	pg/mL
MCP1	Monocyte Chemoattractant Protein 1	pg/mL
POTASSIUM	Potassium	mmol/L
SERUMCYSTC	Serum Cystatin-C	mg/L
TNFA	Tumor Necrosis Factor Alpha	pg/mL
TNFR1	Tumor Necrosis Factor Receptor 1	pg/mL

TNFR2	Tumor Necrosis Factor Receptor 2	pg/mL
TROPONIN	Cardiac Troponin I	pg/mL
UACID	Serum Uric Acid	mg/dL
VCAM1	Vascular Cell Adhesion Molecule 1	pg/mL
VEGF	Vascular Endothelial Growth Factor	pg/mL
EGFR_FAS	Estimated glomerular filtration rate, calculated using the full age spectrum combined serum creatinine and cystatin C equation	mL/min/1.73 m^2
CKD_GFR	Estimated glomerular filtration rate, calculated using the CKD-EPI serum creatinine equation	mL/min/1.73 m^2
OGTT		
CPEP0MIN	C-peptide measurement right before consuming glucose solution	ng/mL
GLU0MIN	Glucose measurement right before consuming glucose solution	mg/dL
INS0MIN	Insulin measurement right before consuming glucose solution	µU/mL
CPEP30MIN	C-peptide measurement 30 minutes after consuming glucose solution	ng/mL
GLU30MIN	Glucose measurement 30 minutes after consuming glucose solution	mg/dL
INS30MIN	Insulin measurement 30 minutes after consuming glucose solution	µU/mL
CPEP60MIN	C-peptide measurement 1 hour after consuming glucose solution	ng/mL
GLU60MIN	Glucose measurement 1 hour after consuming glucose solution	mg/dL
INS60MIN	Insulin measurement 1 hour after consuming glucose solution	µU/mL
CPEP90MIN	C-peptide measurement 90 minutes after consuming glucose solution	ng/mL
GLU90MIN	Glucose measurement 90 minutes after consuming glucose solution	mg/dL
INS90MIN	Insulin measurement 90 minutes after consuming glucose solution	µU/mL
CPEP2HR	C-peptide measurement 2 hours after consuming glucose solution	ng/mL
GLU2HR	Glucose measurement 2 hours after consuming glucose solution	mg/dL
INS2HR	Insulin measurement 2 hours after consuming glucose solution	µU/mL
INSINV	Insulin inverse (1/fasting insulin)	mL/uU
INSINDEX	Insulinogenic index (calculated as $\Delta insulin_{0\text{-}30min}/\Delta glucose_{0\text{-}30min}$)	uU/mL per mg/dL
CPEPINDEX	C-peptide index (calculated as $\Delta C\text{-}peptide_{0\text{-}30min}/\Delta glucose_{0\text{-}30min}$)	ng/mL per mg/dL
ODI	Insulin-based oral disposition index (1/fasting insulin x $\Delta insulin_{0\text{-}30min}/\Delta glucose_{0\text{-}30min}$)	mL/uU x uU/mL per mg/dL
CODI	C-peptide-based oral disposition index (1/fasting insulin x $\Delta C\text{-}peptide_{0\text{-}30min}/\Delta glucose_{0\text{-}30min}$)	mL/uU x ng/mL per mg/dL

5.5.2 TODAY2.ECHO: Echocardiography

This dataset includes one record per participant. Transthoracic echocardiography was performed once during the last year of TODAY (2010-2011) and again five years later during TODAY2 (2015-2016). The data collected during TODAY was previously sent to the repository at the time of the June 2014 release. Herein, we include echocardiography data collected at the follow-up assessment during TODAY2. The protocol for measurement of left ventricular structure and function was identical at each of the two examinations with the exception that left ventricular global strain and mitral valve septal diastolic function measures were also assessed during the second evaluation.

A central Echocardiography Reading Center was used (Johns Hopkins University, A. I. duPont Hospital for Children) and studies were conducted by trained and certified technicians according to study protocol. Heart rate was measured as part of the echocardiogram. Studies were read by a single technician with random rereads for quality control using commercially available software (Digisonics, Houston, TX). Two-dimensional transthoracic echocardiograms were performed with the participant lying in a left lateral decubitus position to maximize image quality. Parasternal short axis, long axis, and apical views were obtained. This allowed measurement of left ventricular size and structure, tissue Doppler imaging of right and left ventricular inflow tracts, and for two dimensional images to allow later retrieval to obtain measurements of LV strain. Speckle tracking and strain data are provided in a separate dataset (see Section 5.5.8). Measurements were made according to the American Society of Echocardiography standards.

Since the echocardiography studies were all performed around the same time (2010-2011 and 2015-2016), they are not associated with a study visit. Not all participants with a follow-up echocardiography assessment collected during TODAY2 had an initial one performed during TODAY, and vice-versa. Records from participants were not included if the participant was pregnant or if their cardiovascular risk factor measurements were not collected within 3 months of the echocardiogram. The data included in the release are listed in the table below. The data indicate functional and anatomic changes in the heart, including left ventricular and lateral/septal E/Em, ejection fraction, and distensibility. Some measures were available at the follow-up echocardiography assessment but not at the initial one (e.g., 2D two chamber LV end diastolic volume, 2D two chamber LV ejection fraction), while other measures were only available at the initial assessment and not at the follow-up one (e.g., LA diameter adjusted for height, LV internal dimension diastole adjusted for height, LV mass/body surface area).

Variable Name	Description	Units / Coding
RELEASEID	Participant ID for NIDDK Database Repository	
DAYS	Days from randomization to echocardiogram	Days
ASCEND2D	2-D – aorta ascending	cm
DIASTO2D	2-D – aortic root end diastolic dimension	cm
SYSTOL2D	2-D – aortic root end systole systolic	cm
LAAREA2D	2-D – four chamber LA area	cm^2
DIAVOL2D	2-D – four chamber LV end diastolic volume	mm
LVEDV2D2CH	2-D – two chamber LV end diastolic volume	mm
LVEF2D2CH	2-D – two chamber LV ejection fraction	%
LVEF2D4CH	2-D – four chamber LV ejection fraction	%
LVESV2D2CH	2-D – two chamber LV end systolic volume	mm
LVESV2D4CH	2-D – four chamber LV end systolic volume	mm
LVE	Doppler diastology E/Em ratio – LV mitral valve lateral peak E	cm/second
LVEM	Doppler diastology E/Em ratio – LV lateral Em	cm/second

LVRATIO	Doppler diastology E/Em ratio – LV lateral E/Em	
LVSEPE	Doppler diastology LV mitral valve septal peak E	cm/second
LVSEPEM	Doppler diastology LV septal Em	cm/second
LVSEPRATIO	Doppler diastology LV septal E/Em	
LVTRIE	Doppler diastology LV mitral valve tricuspid peak E	cm/second
LVTRIEM	Doppler diastology LV tricuspid Em	cm/second
LVTRIRATIO	Doppler diastology LV tricuspid E/Em	
PEAKVELO	Doppler reg flow – tricuspid valve peak velocity	cm/second
RVSYSTOL	Doppler reg flow – tricuspid valve RV systolic pressure	mmHg
AORTROOT	MMODE aorta and LA – aortic root	cm
LADIMEN	MMODE aorta and LA – LA internal dimensions	cm
MMODEHR	MMODE LV and RV – heart rate	beats/minute
IVSDIAS	MMODE LV and RV – inter-ventricular septum diastole	cm
IVSSYST	MMODE LV and RV – inter-ventricular septum systole	cm
LVINDEX	MMODE LV and RV – LV cardiac index	liters/minute/m^2
LVOUTPUT	MMODE LV and RV – LV cardiac output	liters/minute
LVEJECT	MMODE LV and RV – LV ejection fraction	%
DIAVOLMM	MMODE LV and RV – LV end diastolic volume	mm
SYSVOLMM	MMODE LV and RV – LV end systolic volume	mm
PCTSHORT	MMODE LV and RV – LV % fractional shortening	%
LVDIAST	MMODE LV and RV – LV internal dimension diastole	cm
LVSYSTOL	MMODE LV and RV – LV internal dimension systole	cm
LVMASS	MMODE LV and RV – LV mass	g
WALLDIAS	MMODE LV and RV – LV posterior wall diastole	cm
WALLSYST	MMODE LV and RV – LV posterior wall systole	cm
LVSTROKE	MMODE LV and RV – LV stroke volume	ml
TAPSE	MMODE – tricuspid annular plane systolic excursion systolic dimension	cm
WALTHICK	Relative wall thickness = (walldias x 2) / lvdiast	
LVMASSHT	LV mass/ht2.7 where height in m	$g/m^{2.7}$
DOPPLERQC	Doppler quality score	0-<1=poor/not available, 1-<2=fair, 2-<3=good, 3+=excellent
MMODEQC	MMODE – quality score	0-<1=poor/not available, 1-<2=fair, 2-<3=good, 3+=excellent
OVERALLQC	Overall average quality score	0-<1=poor/not available, 1-<2=fair, 2-<3=good,

		3+=excellent
PLAXQC	Parasternal Long Axis quality score	0-<1=poor/not available, 1-<2=fair, 2-<3=good, 3+=excellent
SAXQC	Parasternal Short Axis quality score	0-<1=poor/not available, 1-<2=fair, 2-<3=good, 3+=excellent
APICALQC	Apical quality score	0-<1=poor/not available, 1-<2=fair, 2-<3=good, 3+=excellent

5.5.3 TODAY2.FUNDUS: Fundus Photography

This dataset includes one record per participant and per eye. Fundus photography was completed once during the last year of TODAY (2010-2011) and again seven years later during TODAY2 (2017-2018) using the same protocol as used during TODAY. The data collected during TODAY were previously sent to the repository at the time of the June 2014 release. Herein, we include Fundus photography data collected at the follow-up assessment during TODAY2.

Fundus photography was collected using the University of Wisconsin Fundus Photography Reading Center (FPRC; Madison, WI) Modified 7 Standard Digital Color Fundus Photography procedure (7M-D) and using FPRC-certified photographers. All photographs were graded centrally by graders masked to treatment, age, duration of diabetes, glycemic control, and other clinical characteristics, using the final Early Treatment Diabetic Retinopathy Study (ETDRS) grading scale. All images were captured using a digital system that was FPRC certified for color and capture capability (typically a minimum 3 megapixel resolution or higher with a preferred 5 or 6 megapixel resolution color image capture system). The modified 7-standard stereoscopic fields for color photography specified by the 7M-D procedure differs from the ETDRS 7-standard field protocol in the position of two fields: Field 1M and Field 3M are both modified to include the center of the macula, in Field 1M near the edge of the field and in Field 3M midway between the edge and center of the field. A tutorial on how to capture the images is located on the FPRC website (Fundus Photograph Reading Center).

Since the fundus photography studies were all performed around the same time (2017-2018), they are not associated with a study visit. Not all participants with a follow-up fundus photography assessment collected during TODAY2 had an initial one performed during TODAY, and vice-versa. The list of variables included in this data release was streamlined compared to the TODAY data release. Records from pregnant participants were not obtained beyond the 1st trimester and until at least 3 months postpartum. None of the TODAY participants had a diabetic retinopathy severity grading on the ETDRS scale at the eye level > 75. Variables at the participant level have the same value for each eye. Information about the individual variables, their descriptions, and any associated coding are provided in the table below.

VARIABLE	Description	Coding
RELEASEID	Participant ID for the NIDDK Database Repository	
DAYS	Number of days from randomization to Fundus photography completion	Days
Eye level variables		
EYE	Eye photographed	L=Left; R=Right
CSME	Clinically significant macular edema (ETDRS)	0=None 1=Questionable 2=Zone or RT ≥ 1 DA, part ≤ 1 DD from center (definite criteria 1) 3=RT or adjacent HE ≤ 500 μm from center (definite criteria 2) 8=Cannot grade 9=Not applicable
DRSEVERITY	Diabetic retinopathy severity grading on the ETDRS scale on eye level	10=MA and other characteristics absent 12=DR absent but other non-diabetic lesions present 14A=HE definite; MA absent; 14B=SE definite; MA absent 14C=IRMA definite; MA absent 14Z=Venous loops ≥ D/1; MA absent 15=Hemorrhages(s) definite; MA absent 20=MA definite; other characteristics absent 35A=Venous loops ≥ D/1; 35B=SE, IRMA or VB=Q 35C=Retinal hemorrhages present 35D=HE ≥ D/1; 35E=HE ≥ M/1; 35F=SE ≥ D/1 43A=H/Ma=M/4-5 or S/1; 43B=IRMA=D/1-3 47A=Both 43A and 43B; 47B=IRMA=D/4-5 47C=H/Ma=S/2-3; 47D=VB=D/1 53A=Two or more of 47A, 47B and 47C

		53B=H/Ma ≥ s/4-5; 53C=IRMA ≥ M/1 53D=VB ≥ D/2-3 53E=Two or more of 53B, 53C, and 53D 60=Panretinal photocoagulation or local Rx of NV 61A=FPD or FPE present with NVD and NVe absent 61B=NVE=D/1-5 65A=NVE ≥ N/1; and VHJ and PRH=A or Q 65B=NVD=D; and Vh and PRH=A or Q 65C=VH or PRH=D and NVE<M/1 and NVD absent 71A=VH or PRH ≥ M/1 71B=NVE ≥ M/1 and VH or PRH ≥ D/1 71C=NVD=2 and Vh or PRH ≥ D/1; 71D=NVD ≥ M 75=NVD ≥ M and VH or PRH ≥ D/1 81=NVD=CG, or NVD<D and NVE=CG in ≥ 1 field and absent in all others; and RDCM<D 85A=VH=VS in Field 1 or 2; 85B=RDCM=D 90=Cannot grade
DRSEVERITYRECODE	Diabetic retinopathy severity categorized on eye level (categories 1-10, 90 not gradable)	1=DRSEVERITY levels 10, 12 2=DRSEVERITY levels 14A, 14B, 14C, 14Z, 15, 20 3=DRSEVERITY levels 35A, 35B, 35C, 35D, 35E, 35F 4= DRSEVERITY levels 43A, 43B 5= DRSEVERITY levels 47A, 47B, 47C, 47D 6= DRSEVERITY levels 53A, 53B, 53C, 53D, 53E 7= DRSEVERITY levels 60, 61A, 61B 8= DRSEVERITY levels 65A, 65B, 65C 9= DRSEVERITY levels 71A, 71B, 71C, 71D 10= DRSEVERITY level 75 90= DRSEVERITY level 90
Participant level variables		

DRSEVERITYRECODESUBJECT	Diabetic retinopathy severity grading categories on participant level	1= DRSEVERITYRECODE levels 10=10, 10=19, 12<12, 12=12, 12=90 2=DRSEVERITYRECODE levels 14<14, 15<14, 20<14 3=DRSEVERITYRECODE levels 14=14, 14=90, 15=14, 15=15, 15=90, 20=14, 20=15, 20=20, 20=90 4=DRSEVERITYRECODE level 35<35 5=DRSEVERITYRECODE levels 35=35, 35=90 6=DRSEVERITYRECODE level 43<43 7=DRSEVERITYRECODE levels 43=43, 43=90 8=DRSEVERITYRECODE level 47<47 9=DRSEVERITYRECODE levels 47=47, 47=90 10=DRSEVERITYRECODE level 53<53 11=DRSEVERITYRECODE levels 53=53, 53=90 12=DRSEVERITYRECODE levels 61<60, 60<60 13=DRSEVERITYRECODE levels 61=61, 61=60, 60=60, 60=90, 61=90 14=DRSEVERITYRECODE level 65<65 15=DRSEVERITYRECODE levels 65=65, 65=90 16=DRSEVERITYRECODE level 71<71 17=DRSEVERITYRECODE levels71=71, 71=90 18=DRSEVERITYRECODE level 75<75 19=DRSEVERITYRECODE levels 75=75, 75=90 20=DRSEVERITYRECODE level 81<81 21=DRSEVERITYRECODE levels 81=81, 81=90 22=DRSEVERITYRECODE level 85<85 23=DRSEVERITYRECODE levels 85=85, 85=90 90=DRSEVERITYRECODE level 90=90
DRSEVERITYSUBJECTCAT	Diabetic retinopathy severity summarized (i.e., grouped DRSEVERITYRECODE SUBJECT categories) on participant level	Early or stable, treated PDR= Early or stable, treated PDR High risk PDR=High risk PDR Images not gradable/available=Images not gradable/available Mild NPDR=Mild NPDR Moderate NPDR=Moderate NPDR Moderately severe NPDR=Moderately severe NPDR No definitive diabetic retinopathy=No definitive diabetic retinopathy Severe NPDR=Severe NPDR Very mild NPDR=Very mild NPDR

DR_PROG	Diabetic retinopathy ETDRS grading scale change between the 2 fundus photograph assessments	numeric
DR_PROG3S	Diabetic retinopathy progression, defined as a 3-step or more progression in the ETDRS grading scale between the 2 fundus photograph assessments	0=No 1=Yes

5.5.4 TODAY2.LIPO: Liposcience Data

This dataset includes one record per participant per data collection visit. Liposcience data were collected during TODAY (2004-2011) and during the first three years of TODAY2 (2011-2014) per study protocol. Liposcience data were not available at the time of the June 2014 TODAY data release (assays were performed in 2016). Therefore, results for both TODAY and TODAY2 visits are included in this data release.

NMR LipoProfile® testing was performed using serum from participants who provided consent for use of archived frozen samples and performed at LabCorp (formerly LipoScience, Raleigh, NC) on the 400 MHz Profiler platform using the LP4 deconvolution algorithm. VLDL, LDL, and HDL subclass particle concentrations were quantified on the basis of the amplitudes of their spectroscopically-distinct lipid methyl group NMR signals. Total particle concentrations of VLDL (VLDL-P), LDL (LDL-P), and HDL (HDL-P) are the sums of the particle concentrations of their respective subclasses. The absolute concentrations of LDL and HDL particles generated by the LP4 algorithm are calibrated to agree more closely with those assessed by their apolipoprotein compositions. Mean VLDL, LDL, and HDL particle sizes (nm diameter) are weighted averages derived from the sum of the diameters of each subclass multiplied by its relative mass percentage. Also calculated is the Lipoprotein Insulin Resistance Index (LP-IR), a multi-marker score ranging from 0 (most insulin sensitive) to 100 (most insulin resistant) derived by combining appropriately-weighted subclass particle concentrations of large VLDL, small LDL, and large HDL plus mean VLDL, LDL, and HDL particle sizes.

Information about the individual variables, their descriptions, and any associated coding are provided in the table below. The LPPHASE variable indicates whether the sample was taken during a TODAY or TODAY2 visit. The variable LPMONTH, indicating the visit month when the lipoprotein measurement was obtained, can be recoded and renamed as needed to merge records using the variable PVISIT in the TODAY2.CBL dataset. For example, a value for 48 for LPMONTH corresponds to PVISIT=P48 if the visit occurred during TODAY2.

Variable Name	Description	Units / Coding
RELEASEID	Participant ID for NIDDK Database Repository	
LPMONTH	Liposcience visit month	0=baseline (M00) 12=month 12 (M12) 24=month 24 (M/P24) 36=month 36 (M/P36) 48=month 48 (M/P48) 60=month 60 (M/P60)

LPPHASE	Study phase of liposcience measure	0=TODAY 1=TODAY2
DAYS	Days from randomization to liposcience assessment	Days
ALA	Alanine Concentrations	µmol/L
APOA1	ApoA-1 Concentrations	mg/dL
APOB	ApoB Concentrations	mg/dL
BCAA	Total BCAC Concentrations	µmol/L
CTR	Citrate	µmol/L
DRF5	5-Year Diabetes Risk Factor Index	score 1-100
GLU	Glucose	mg/dL
GLYCA	GlycA	µmol/L
GXH13_I	Short-term Diabetes Risk Factor (H1&3)	score 1-100
GXH5_I	Short-term Diabetes Risk Factor (H5)	score 1-100
GXH6_I	Short-term Diabetes Risk Factor (H6)	score 1-100
H1P	H1P cHDLP Concentrations, 7.4	µmol/L
H2P	H2P cHDLP Concentrations, 7.8	µmol/L
H3P	H3P cHDLP Concentrations, 8.7	µmol/L
H4P	H4P cHDLP Concentrations, 9.5	µmol/L
H5P	H5P cHDLP Concentrations, 10.3	µmol/L
H6P	H6P cHDLP Concentrations, 10.8	µmol/L
H7P	H7P cHDLP Concentrations, 12.0	µmol/L
HDLZ	HDL Size, range 7.4-13	nm
IRDRF	Insulin Resistance Diabetes Risk Factor	score 1-100
ILEU	Isoleucine Concentrations	µmol/L
LDLP	Total cLDLP Concentrations, range 19-23	nmol/L
LDLZ	LDL Size, range 19-22.5	nm
LGVX	NMR Longevity Index	score 1-100
LPIR	Lipoprotein Insulin Resistance Index	score 1-100
L_TRLP	Large TRLP Particle Concentrations, range 50-89	nmol/L
L_CHDLP	Large cHDLP Concentrations, range 9.6-13	µmol/L
L_CLDLP	Large cLDLP Concentrations, range 21.5-23	nmol/L
LEU	Leucine Concentrations	µmol/L
M_TRLP	Medium TRLP Particle Concentrations, range 37-49	nmol/L
M_CHDLP	Medium cHDLP Concentrations, range 8.1-9.5	µmol/L
M_CLDLP	Medium cLDLP Concentrations, range 20.5-21.4	nmol/L
NHDLC	HDLC Concentrations	mg/dL
NLDLC	LDLC Concentrations	mg/dL
NTC	Total Chol Concentrations	mg/dL
NTG	Total TG Concentrations	mg/dL
NTRLC	TRL C Concentrations	mg/dL
NTRLTG	TRL TG Concentrations	mg/dL
SDRF_I	Short-term Diabetes Risk Factor (I)	score 1-100
SDRF_M	Short-term Diabetes Risk Factor (M)	score 1-100
S_TRLP	Small TRLP Particle Concentrations, range 30-36	nmol/L
S_CHDLP	Small cHDLP Concentrations, range 7.4-8.0	µmol/L
S_CLDLP	Small cLDLP Concentrations, range 19-20.4	nmol/L
TRLP	Total TRLP Particle Concentrations, range 24-240	nmol/L
TRLZ	TRL Size, range 30-100	nm

VL_TRLP	Very Large TRLP Particle Concentrations, range 90-240	nmol/L
VS_TRLP	Very Small TRLP Particle Concentrations, range 24-29	nmol/L
VAL	Valine Concentrations	µmol/L
CHDLP	Total cHDLP Concentrations, range 7.4-13	µmol/L

5.5.5 TODAY2.OCT: OCT Photography

This dataset includes one record per participant per eye and per data collection visit. Time Domain Optical Coherence Tomography (TD-OCT) and Spectral Domain Optical Coherence Tomography (SD-OCT) scans were completed in the last year of TODAY (2010-2011). SD-OCT scans were obtained once more, approximately 7 years later (2017-2018) during TODAY2. OCT scans were obtained at the same time as the Fundus photography assessments (Section 5.5.5) and were collected using standard clinic procedures to obtain optimal quality scans.

Participants in the TODAY study were imaged using either TD-OCT technology (Stratus, Carl Zeiss Meditec, Inc., Dublin, CA) or SD-OCT technology (Spectralis, Heidelberg Engineering, Inc., Heidelberg, Germany and Cirrus, Carl Zeiss, Dublin, California, USA), whereas all imaging in TODAY2 was in the form of SD-OCT, with 6 x 6 mm macular volume scans being acquired from each eye. All OCTs were graded centrally at the Wisconsin Reading Center (Madison, WI) by graders masked to treatment and other clinical characteristics. Data from the OCT scans collected in the last year of TODAY were included in the June 2014 TODAY data release. In October 2018, the reading center regraded the SD-OCT images from TODAY using a different grading mechanism. The data provided in this dataset include the newly regraded OCT data from TODAY as well as data collected at the second OCT assessment in TODAY2; both (regraded TODAY scans and TODAY2 scans) were evaluated via the same methodology.

Total retinal thickness, measured from internal limiting membrane (ILM) to retinal pigment epithelium (RPE) in the central subfield, was considered within normal range if the thickness was < 300 microns for males and < 285 microns for females. The SD-OCT volume scans underwent a custom-built semi-automated segmentation software (independent of the manufacturer) for ILM, outer plexiform layer (OPL), ellipsoid zone (EZ), and RPE layers. Any segmentation errors were reviewed and manually corrected to generate total retinal thickness (ILM to RPE), inner retinal thickness (ILM to OPL), outer retinal thickness (OPL to RPE), and photoreceptor thickness (EZ to RPE). The OCT thicknesses were calculated for the central 1-mm and the inner 3-mm circle diameter regions. The thickness in the central 1-mm is referred to as the central subfield thickness and the inner subfield thickness in the 3-mm circle is the average thickness of the central and four inner subfields. Morphological abnormalities graded on OCT were subretinal fluid, intraretinal cystoid spaces, posterior vitreous detachment (PVD), epiretinal membrane (ERM), retinal traction and distortion (RTD), and macular hole (MH).

Since the OCT studies were all performed around the same time during the study, they are not associated with a study visit. Not all participants with a follow-up OCT assessment collected during TODAY2 had an initial one performed during TODAY, and vice-versa. OCT scans for which the measurement of central subfield thickness was deemed unreliable and did not pass quality control procedures were excluded. Information about the individual variables, their descriptions, and any associated coding are provided in the table below.

VARIABLE	Description	Coding
RELEASEID	Participant ID for the NIDDK Database Repository	
TIMEPOINT	Time point of OCT assessment (first assessment vs. second assessment)	OCT1=First OCT2=Second
DAYS	Number of days from randomization to OCT completion	Days
EYE	Eye photographed	L=Left; R=Right
OCT_TYPE	OCT type	SD-OCT=SD-OCT TD-OCT=TD-OCT
MODEL	OCT manufacturer – model	Cirrus=Cirrus Spectralis=Spectralis Stratus=Stratus
LAYER	OCT layer	ILM/OPL=Inner retinal thickness ILM/RPE_ON=Total retinal thickness ISe/RPE_ON=Photoreceptor retinal thickness
SECTOR	OCT sector	CSF=Center II=Inferior IN=Nasal IS=Superior IT= Temporal
RETINALTHICKNESS	Retinal thickness value across the different layers and sectors, at the center point measured by SD-OCT	µm
SSR_PRESENCE	Presence of subretinal fluid	0=Absent; 1=Questionable; 2=Definite, SSR only present outside central 1mm 3=Definite, SSR only present within central 1mm 4=Definite, SSR present within the central 1mm and outside the central 1mm 7=Definite, unable to determine location; 8=Cannot grade; 9=Not applicable

CYSTOIDSPACES	Presence intraretinal cystoid spaces	0=Absent 1=Questionable 2=Definite, cystoid spaces present only outside central 1mm 3=Definite, cystoid spaces present only within central 1mm 4=Definite, cystoid spaces present within and outside central 1mm 7=Definite, unable to determine location 8=Cannot grade 9=Not applicable
VITAB_MH	Vitreoretinal interface abnormalities – macular hole	0=Absent 1=Questionable 2=Pseudohole or lamellar hole 3=Definite, stage 1 4=Definite, stage 2, stage 3 or stage 4 8=Cannot grade 9=Not applicable
VITAB_ERM	Presence of Vitreoretinal Interface Abnormalities - Epiretinal Membrane	0=Absent 1=Questionable 2=Definite, outside central 1mm 3=Definite, central 1mm involved – questionable 4=Definite, central 1mm involved – definite 8=Cannot grade 9=Not applicable
VITAB_PVD	Presence of Vitreoretinal Interface Abnormalities - Posterior Vitreous Detachment	0=Absent 1=Questionable 2=Definite, non-adherent 3=Definite, questionably adherent 4=Definite, partially adherent 8=Cannot grade 9=Not applicable
VITAB_RTD	Presence of Vitreoretinal Interface Abnormalities - Retinal Traction and/or Distortion	0=Absent 1=Questionable 2=Definite, outside central 1mm 3=Definite, central 1mm involved – questionable 4=Definite, central 1mm involved – definite 8=Cannot grade 9=Not applicable

5.5.6 TODAY2.PSG: Polysomnogram

This dataset includes one record per participant. Per study protocol, polysomnograms (PSGs) were obtained once on a subset of TODAY participants (n=115) at 6 clinical centers during 2015-2016 to match the racial/ethnic distribution of the study cohort. Males were oversampled in order to provide a 1:1 female to male ratio. Any subject who was pregnant or congested due to a cold was not eligible to participate. Additionally, any participant who had previously been diagnosed with obstructive sleep apnea (OSA) was also ineligible. Otherwise, all participants enrolled in TODAY2 were eligible to participate. The PSG included an EEG (2 central leads, 2 frontal lead and 2 occipital leads), right and left electro-oculogram or EOG, chin EMG, 2 measures of airflow (nasal pressure transducer and oronasal thermistor), abdominal and thoracic measures of effort using respiratory inductance plethysmography, snoring microphone, finger pulse oximetry, right and left leg EMG, and 2-lead ECG. The participant was also monitored by a sleep technologist to document observations including body position, biocalibration notes, artifact recognition, lights off/on times, and any unusual behaviors. All data were transferred to and graded by the University of Chicago Sleep Metabolism and Health Center (SMAHC).

Since the polysomnogram studies were performed around the same time during TODAY2 (2015-2016), they are not associated with a study visit. Information about the individual variables, their descriptions, and any associated coding are provided in the table below.

VARIABLE NAME	Description	Units / Coding
RELEASEID	Participant ID for NIDDK Database Repository	
DAYS	Days from randomization to polysomnogram assessment	Days
OSA_SEVERITY_LEVEL	OSA Severity Level	0=None 1=Mild 2=Moderate 3=Severe
CENTRAL_AHI	Central Apnea Hypopnea Index (AHI)	events/hour
LATENCY_SLEEP_ONSET	Latency Sleep Onset	min
LIGHTS_OFF	Lights off	clock time
LIGHTS_ON	Lights on	clock time
LOWEST_O2_NREM	Lowest oxygen saturation during NREM sleep	%
LOWEST_O2_REM	Lowest oxygen saturation during REM sleep	%
LOWEST_O2_ALL	Lowest oxygen saturation during sleep	%
N1	N1 min	min
N2	N2 min	min
N1PCT	N1 %	%
N2PCT	N2 %	%
NREM3_ODI	NREM 3% oxygen desaturation index	events/hour
NREM_MA_INDEX	NREM microarousal index	events/hour
NREM_OBSTRUCTIVE_AHI	NREM Obstructive AHI	events/hour
NONSUPINE_AHI	NonSupine AHI	events/hour
NONSUPINE_SLEEP	Nonsupine Sleep (min)	min
ODI4_NREM	NREM 4% oxygen desaturation index	events/hour

ODI4_TOTAL	4% oxygen desaturation index Total	events/hour
OBSTRUCTIVE_AHI	Obstructive AHI	events/hour
REM	REM min	min
REM3_ODI	REM 3% oxygen desaturation index	events/hour
REM4_ODI	REM 4% oxygen desaturation index	events/hour
REMPCT	REM %	%
REM_MA_INDEX	REM microarousal index	events/hour
REM_OBSTRUCTIVE_AHI	REM Obstructive AHI	events/hour
SWS	N3 or slow wave sleep min	min
SWSPCT	N3 or slow wave sleep %	%
SLEEP_EFFICIENCY	Sleep Efficiency	%
SUPINE_AHI	Supine AHI	events/hour
SUPINE_SLEEP	Supine Sleep (min)	min
T90	% of total sleep time below 90% oxygen saturation	%
T90_NREM	% of NREM sleep time below 90% oxygen saturation	%
T90_REM	% of REM sleep time below 90% oxygen saturation	%
TOT_MA_INDEX	Total microarousal index	events/hour
TRT	Total recording time (min)	min
TST	Total sleep time (min)	min
TOTAL3_ODI	Total 3% oxygen desaturation index	events/hour
WASO	Wake after sleep onset (min)	min
WAKE	Wake min	min
WAKEPCT	Wake %	%

5.5.7 TODAY2.PWV: Pulse Wave Velocity and Heart Rate Variability

This dataset includes one per participant per assessment. Arterial stiffness assessments (pulse wave velocity [PWV] and heart rate variability [HRV]) were assessed twice during TODAY2, once in 2013-2014 and again five years later (2018-2019).

Pulse wave velocity, augmentation index and heart rate variability measurements were obtained using the SphygmoCor CPV system (AtCor Medical, Lisle, IL). Brachial distensibility was measured using the DynaPulse 2000 (PulseMetric, San Diego, CA). All measurements were conducted fasting and after the participant rested for at least 10 minutes. All prescriptions and over-the-counter medications were held on the day of testing until both tests were complete. The arterial stiffness protocol was identical at both assessments. Study staff who conducted the assessments were certified for performance by a central Vascular Reading Center located in Cincinnati, Ohio.

Since the arterial stiffness studies were all performed around the same time during TODAY2 (2013-2014 and 2018-2019), they are not associated with a study visit. Not all participants with a follow-up arterial stiffness assessment collected had an initial one performed, and vice-versa. Records from participants were not included if the participant was pregnant or if their cardiovascular risk factor measurements were not collected within 3 months of the arterial stiffness assessment. Information about the individual variables, their descriptions, and any associated coding are provided in the table below.

Variable Name	Description	Units / Coding
RELEASEID	Participant ID for NIDDK Database Repository	
TIMEPOINT	Time point of arterial stiffness assessment (first assessment vs. second assessment)	PWV1=First PWV2=Second
DAYS	Days from randomization to arterial stiffness assessment	Days
SDNN	Standard Deviation of R-R Interval (termed NN interval)	ms
PNN50	Percent of consecutive NN intervals differing by more than 50ms	%
RMSSD	Root-mean-square of the difference of successive NN intervals	ms
SDANN	Standard Deviation of the Means for each R to R Segment	ms
SDNN_INDEX	Mean of the Standard Deviations for each R to R Segment	ms
LF_POWER_NORMALISED	Percentage of Low Frequency Power (0.04 – 0.4Hz)	normalized units
HF_POWER_NORMALISED	Percentage of High Frequency Power	normalized units
LF_HF_RATIO	Ratio of Low Frequency Power to High Frequency Power	
PWVF	PWV visit mean for carotid femoral	m/sec
PWVR	PWV visit mean for carotid radial	m/sec
PWVFFT	PWV visit mean for femoral to foot	m/sec
PWVD	PWV visit mean for distal	m/sec
AIX	Augmentation Index mean	%
C_SBP	Central Systolic Pressure mean	mmHg
C_DBP	Central Diastolic Pressure mean	mmHg
BRACHD	BrachD study visit mean	% change/mmHg
MAP_PM	Mean MAP measured by oscillometry	mmHg
HR_PM	measured HR	beats/min
CI	mean CO indexed to BSA	$l/min/m^2$
LVDPDT	mean maximal rate of rise of LV pressure	mmHg/sec
LV_CTX	mean LV dp/dt max/pressure	1/sec
SVC	mean Compliance of the entire systemic vascular system	mL/mmHg
SVR	mean resistance of the systemic vascular system	$dynes.sec/cm^5$

5.5.8 TODAY2.SPECKLE: Speckle Tracking

This dataset includes one record per participant per assessment. Transthoracic echocardiography was performed once during the last year of TODAY (2010-2011) and again five years later during TODAY2 (2015-2016). Speckle tracking and strain measurements were derived from echocardiograms. Details regarding the echocardiography assessment can be found in Section 5.5.2. The speckle tracking and strain data from the TODAY visits were not available at the time of the June 2014 release. Therefore, we include the speckle tracking and strain data from both study periods (TODAY and TODAY2). All speckle tracking and strain measurements were analyzed with TomTec (Unterschleissheim, Germany) for global and regional myocardial deformation and strain, volumes, mass, and ejection fraction. Measurements were made according to the American Society of Echocardiography standards.

Since the echocardiography studies were all performed around the same time during the study (2010-2011 and 2015-2016), they are not associated with a study visit. Not all participants with a follow-up echocardiography assessment collected during TODAY2 had an initial one performed during TODAY, and vice-versa. Information about the individual variables, their descriptions, and any associated coding are provided in the table below. The data include information on LV peak longitudinal 2, 3 and 4 chamber strain, and circumferential and radial strain.

Variable Name	Description	Units / Coding
RELEASEID	Participant ID for NIDDK Database Repository	
TIMEPOINT	Time point of speckle tracking assessment (first assessment vs second assessment)	SPECKLE1=First SPECKLE2=Second
DAYS	Days from randomization to speckle tracking assessment	Days
Strains		
GLS_4CH	Global longitudinal strain - four chamber view	%
GLS_4CH_T2P	Time to peak of global longitudinal strain - four chamber view	
GLS_2CH	Global longitudinal strain - two chamber view	%
GLS_2CH_T2P	Time to peak of global longitudinal strain - two chamber view	
GLS_3CH	Global longitudinal strain - three chamber view	%
GLS_3CH_T2P	Time to peak of global longitudinal strain - three chamber view	
SAX_PM_GCS	Global circumferential strain - papillary muscle level	%
SAX_PM_GCS_T2P	Time to peak - global circumferential strain - papillary muscle level	
SAX_MV_GCS	Global circumferential strain - mitral valve level	%
SAX_MV_GCS_T2P	Time to peak - global circumferential strain - mitral valve level	
SAX_AP_GCS	Global circumferential strain - apex level	%
SAX_AP_GCS_T2P	Time to peak - global circumferential strain - apex level	
SAX_GRS	Global radial strain - papillary muscle level	%
SAX_GRS_T2P	Time to peak - global radial strain - papillary muscle level	
Strain Rate		
GLSR_4CH	Global longitudinal strain rate- four chamber view	
GLSR_2CH	Global longitudinal strain rate - two chamber view	
GLSR_3CH	Global longitudinal strain rate - three chamber view	
SAX_PM_GCSR	Global circumferential strain rate - papillary muscle level	
SAX_MV_GCSR	Global circumferential strain rate - mitral valve level	
SAX_AP_GCSR	Global circumferential strain rate - apex level	
SAX_GRSR	Global radial strain rate - papillary muscle level	

Volumes and EF		
EDV_4CH	Left ventricle end-diastolic volume - four chamber view	
ESV_4CH	Left ventricle end-systolic volume - four chamber view	
EF_4CH	Left ventricle ejection fraction - four chamber view	
EDV_2CH	Left ventricle end-diastolic volume - two chamber view	
ESV_2CH	Left ventricle end-systolic volume - two chamber view	
EF_2CH	Left ventricle ejection fraction - two chamber view	
EDV_3CH	Left ventricle end-diastolic volume - three chamber view	
ESV_3CH	Left ventricle end-systolic volume - three chamber view	
EF_3CH	Left ventricle ejection fraction - three chamber view	
FAC		
SAX_PM_FAC	Fractional area change - papillary muscle level	
SAX_MV_FAC	Fractional area change - mitral valve level	
SAX_AP_FAC	Fractional area change - apex level	

6 ADDENDUM

The two datasets below include data collected during TODAY that only became available after the first TODAY data release.

6.1.1 TODAY.ADDCBL: Additional TODAY Laboratory Data

This dataset includes one record per participant per data collection visit. This dataset includes data collected during TODAY but evaluated from storage samples only among participants who had granted permission for the use of their samples for future research. The assays were performed after the end of the TODAY randomized trial and were not included in the original TODAY data release from June 2014.

During TODAY, blood and spot urine samples were obtained after a 10-14 hour overnight fast, and processed immediately according to standardized procedures and shipped on dry ice for analysis at the central biochemical laboratory (Northwest Lipid Metabolism and Diabetes Research Laboratories, University of Washington, Seattle WA).

Per study protocol, adiponectin and high-molecular weight adiponectin were obtained at randomization (baseline visit), month 6, 24, and 36. Sex hormones (i.e., testosterone, estradiol estrone) were obtained at randomization (baseline visit), month 6, 12, and 24. Cardiac biomarkers (i.e., BNP, copeptin, selectin), uric acid, serum Cystatin C, and vascular endothelial growth factor were assessed at randomization (baseline visit) and annually during TODAY. Zinc transporter 8 was evaluated at randomization (baseline visit) and at all visits originally coded as a primary outcome or end of study TODAY visit (then recoded to a regular visit number). Estimates of glomerular filtration rates were calculated via the Full Age Spectrum (FAS) combined serum creatinine and cystatin C equation, which has been validated in children

and adults, and via the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation using serum creatinine values.

Oral glucose tolerance tests (OGTTs) were obtained after a 10-14 hour overnight fast at baseline, months 6 and 24, annually thereafter during TODAY, and at all visits originally coded as a primary outcome or end of study visit (then recoded to a regular visit number). Stimulated insulin was not collected from participants receiving exogenous insulin therapy. Measurements of insulin, glucose, and C-peptide at the different OGTT time points (0 min, 30 min, etc.) were included in the June 2014 TODAY data release. Herein, we report common OGTT-based variables (i.e., insulinogenic index, oral disposition index) that were calculated based on those time points and have been reported in prior TODAY/TODAY2 publications.

Information about the individual variables, their descriptions, and any associated coding are provided in the table below. Data included in this ADDCBL dataset can be merged with the previously released TODAY CBL dataset using the variables RELEASEID and MVISIT.

Variable Name	Description	Units / Coding
RELEASEID	Participant ID for NIDDK Database Repository	
MVISIT	Release visit number	M00=Baseline M02-M78=Month 2–Month78
DAYS	Days from randomization to CBL completion	Days
Lab analyte		
ADIPONECTIN	Adiponectin	ng/mL
HMWA	High Molecular Weight Adiponectin	ng/mL
EDOL	Estradiol Estrone	pg/mL
SHBG	Sex Hormone Binding Globulin	nmol/L
TESTOSTERONE	Testosterone	ng/mL
ZNT8	Zinc Transporter 8	
BNP	Brain Natriuretic Peptide	pg/mL
COPEPTIN	Copeptin	ng/mL
EPO	Erythropoietin	pg/mL
ESELECTIN	E-Selectin	pg/mL
FGF23	Fibroblast Growth Factor 23	pg/mL
ICAM1	Intercellular Adhesion Molecule 1	pg/mL
IGFBP1	Insulin-like Growth Factor Binding Protein 1	pg/mL
IL1	Inflammatory Cytokine IL-1	pg/mL
MCP1	Monocyte Chemoattractant Protein 1	pg/mL
SERUMCYSTC	Serum Cystatin-C	mg/L
TNFA	Tumor Necrosis Factor Alpha	pg/mL
TNFR1	Tumor Necrosis Factor Receptor 1	pg/mL
TNFR2	Tumor Necrosis Factor Receptor 2	pg/mL
TROPONIN	Cardiac Troponin I	pg/mL
UACID	Serum Uric Acid	mg/dL
VCAM1	Vascular Cell Adhesion Molecule 1	pg/mL
VEGF	Vascular Endothelial Growth Factor	pg/mL

EGFR_FAS	Estimated glomerular filtration rate, calculated using the full age spectrum combined serum creatinine and cystatin C equation	mL/min/1.73 m^2
CKD_GFR	Estimated glomerular filtration rate, calculated using the CKD-EPI serum creatinine equation	mL/min/1.73 m^2
OGTT		
INSINV	Insulin inverse (1/fasting insulin)	mL/uU
INSINDEX	Insulinogenic index (calculated as $\Delta insulin_{0\text{-}30min}/\Delta glucose_{0\text{-}30min}$)	uU/mL per mg/dL
CPEPINDEX	C-peptide index (calculated as $\Delta C\text{-}peptide_{0\text{-}30min}/\Delta glucose_{0\text{-}30min}$)	ng/mL per mg/dL
ODI	Insulin-based oral disposition index (1/fasting insulin x $\Delta insulin_{0\text{-}30min}/\Delta glucose_{0\text{-}30min}$)	mL/uU x uU/mL per mg/dL
CODI	C-peptide-based oral disposition index (1/fasting insulin x $\Delta C\text{-}peptide_{0\text{-}30min}/\Delta glucose_{0\text{-}30min}$)	mL/uU x ng/mL per mg/dL

6.1.2 TODAY.ADDDXA: Additional TODAY Dual Energy X-Ray Absorptiometry

This dataset includes one record per participant per data collection visit. Dual X-ray absorptiometry (DXA) scans were performed on all participants during TODAY at baseline (M00), 6 months (M06), 24 months (M24), and at all visits originally coded as a primary outcome or end of study visit (then recoded to a regular visit number). Whole body adiposity DXA data (including measures of percent fat, fat mass, lean mass, and total mass) collected during TODAY were previously sent to the repository at the time of the June 2014 data release. DXA data were not collected during TODAY2 per study protocol.

Herein, we include data from re-analyzed Hologic and GE scans (performed in 2015) that provided estimates of visceral adipose tissue (VAT) and subcutaneous adipose tissue (SAT) from the total abdominal fat volume where VAT = Total Abdominal Fat – SAT, and SAT is estimated from manufacturer-specific models using the fat projected outside the abdominal walls. DXA VAT has been validated against MRI or CT in both adults and children. Other measures were derived from the DXA dataset that included total mass, fat mass, lean mass, and percent fat mass of the android and gynoid regions. The pooling of results from Hologic and GE scans was done using calibrating equations and the re-analysis was performed using manufacturer-specific software (Hologic Apex 4.0 and Prodigy 14.1). DXA scans that failed quality assurance protocols were excluded.

Information about the individual variables, their descriptions, and any associated coding are provided in the table below. Data included in this ADDDXA dataset can be merged with the previously released TODAY DXA dataset using the variables RELEASEID and MVISIT.

Variable Name	Description	Units / Coding
RELEASEID	Participant ID for NIDDK Database Repository	
MVISIT	Release visit number	M00=Baseline M02-M78=Month 2–Month 78
DAYS	Days from randomization to DXA completion	Days
MACHINE	DXA machine type	HO=Hologic Delphi/A and W LU=GE Lunar Prodigy
ANDROID_FAT	ANDROID fat	grams

ANDROID_LEAN	ANDROID lean	grams
ANDROID_MASS	ANDROID mass	grams
ANDROID_PFAT	ANDROID percent fat	%
GYNOID_FAT	GYNOID fat	grams
GYNOID_LEAN	GYNOID lean	grams
GYNOID_MASS	GYNOID mass	grams
GYNOID_PFAT	GYNOID percent fat	%
ANDROID_GYNOID_RATIO	ANDROID to GYNOID ratio	
SAT_AREA	Subcutaneous adipose tissue area	cm^2
SAT_MASS	Subcutaneous adipose tissue mass	grams
SAT_VOLUME	Subcutaneous adipose tissue volume	cm^3
VFAT_AREA	Visceral adipose tissue area	cm^2
VFAT_MASS	Visceral adipose tissue mass	grams
VFAT_VOLUME	Visceral adipose tissue volume	cm^3

7 Acknowledgments

What is contained below should be used to acknowledge the source when producing manuscripts utilizing the TODAY Study data and/or specimens:

The TODAY Study was conducted by the TODAY Study Group and supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the National Heart, Lung, and Blood Institute (NHLBI), and the National Eye Institute. The data [and biospecimens] from the TODAY Study were supplied by the NIDDK Central Repository. This manuscript was not prepared under the auspices of the TODAY Study and does represent analyses or conclusions of the TODAY Study Group, the NIDDK Central Repository, or the NIH.