Cohort profile: The Nanjing Diabetes Cohort database – a population-based surveillance cohort

Wei Feng[1,†], Yuechuchu Yin[3,1,†], Chunyan Gu[1], Yuan Mu[4], Duxiao Zhang[5], Zhenhuan Tao[6], Weidong Yin[6], Xin Zhang[3,1], Yun Yu[1,2], Jie Hu[1,2], Cheng Wan[1,2*], Yun Liu[3,1*]

[1] Department of Medical Informatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, Jiangsu, 211166, China
[2] Institute of Medical Informatics and Management, Nanjing Medical University, No.300 Guang Zhou Road, Nanjing, Jiangsu, 210029, China
[3] Department of Information, the First Affiliated Hospital, Nanjing Medical University, No.300 Guang Zhou Road, Nanjing, Jiangsu, 210029, China
[4] Department of Laboratory Medicine, the First Affiliated Hospital, Nanjing Medical University, No.300 Guang Zhou Road, Nanjing, Jiangsu, 210029, China
[5] Department of Pharmacy, the First Affiliated Hospital, Nanjing Medical University, No.300 Guang Zhou Road, Nanjing, Jiangsu, 210029, China
[6] Nanjing Health Information Center, No. 3, Zizhulin Xufu Lane, Gulou District, Nanjing, Jiangsu, China
[†] Wei Feng and Yuechuchu Yin contributed equally.
[*] Corresponding Author:
Yun Liu[5,7], Cheng Wan[1,2]
Email address: yun_liu@njmu.edu.cn, chengwan@njmu.edu.cn

Abstract

Purpose: To study epidemiology, complications, risk factors, clinical course and treatment patterns of diabetes, the Nanjing Diabetes Cohort (NDC) was established using anonymized electronic medical records from 650 hospitals and primary cares since 2020. This cohort provides important opportunities for healthcare professionals treating patients with diabetes.
Participants: Diabetes was defined as having inpatient or outpatient encounters with diagnosis of diabetes ICD-9/10 codes, any use of insulin or oral hypoglycemic drugs, or one encounter with hemoglobin A1C > 4.8mmol/mol or 6.5%. Patients with diabetes have been continuously enrolled from hospitals and primary cares in Nanjing since 2020. Demographic, medications, and comorbidities data were abstracted using clinical notes, diagnostic codes, labs, medications, and vital signs among different types of diabetes.
Findings to date: The NDC consisted of 1033904 patients from Jan 1st, 2020 to Dec 31th, 2022, the most part of whom were male (50.62%) and from district Gulou (30.79%). The clinical characteristics and medications usage of patients with type 1 diabetes, type 2 diabetes, gestational diabetes and other diabetes were assessed. The prevalence of hypertension, heart failure, and cerebrovascular disease were 31.20%, 21.46%, and 13.74% respectively.
Future plans: NDC will enroll eligible patients continuously each year. The data of NDC is

maintained by Department of Medical Informatics, Nanjing Medical University, and available on request after ethical approval.

Strengths and limitations of this study

1. The NDC is an ongoing population-base, large-sized cohort that contains over 1000000 patients with type 1 diabetes, type 2 diabetes, gestational diabetes, and other diabetes with follow-up records.
2. NDC coverts the diabetes patients from Nanjing with urban and suburban areas, guaranteeing the generalizability of coverage.
3. The NDC captures patients with diabetes continuously from the Nanjing healthcare information platform, and provides exhaustive clinical data on clinical report, laboratory results, glucose-lowering drugs, diabetes-related comorbidities, and death report.
4. Patients who had done no measurement of glycemic indexes or suffered prediabetes with normal HbAlc would not be included in the NDC.

Data availability statement

Data are available upon reasonable request.

Funding declaration

This study was supported by the industry prospecting and common key technology key projects of Jiangsu Province Science and Technology Department (Grant no. BE2020721), the industrial and information industry transformation and upgrading special fund of Jiangsu Province in 2021 (Grant no. [2021]92), the key project of smart Jiangsu in 2020 (Grant no. [2021]1), Jiangsu Province Engineering Research Center of big data application in chronic disease and intelligent health service (Grant no. [2020]1460).

Collaborators

The data of the NDC are not open access but can be shared under conditions of collaboration and endowment. Collaborative research projects are encouraged. For more detailed information on the NDC, please contact: Prof. Yun Liu (yun_liu@njmu.edu.cn)

Contributors

Study concept and design: Wei Feng, Yuechuchu Yin, Chen Wan; acquisition of data: Wei Feng, Yuechuchu Yin, Chunyan Gu, Chen Wan, Yuan Mu, Zhenhuan Tao, Weidong Yin; data analysis and verification: Chunyan Gu, Xin Zhang, Yuzhuo Wang, Yun Yu, Yuan Mu, Duxiao Zhang; drafting of the manuscript: Wei Feng, Yuechuchu Yin, Chen Wan; acquisition of funding: Yun Liu. All authors have read and approved the final version.

Competing interests

None declared.

Patient consent for publication

Not required.

Ethics approval

The study was approved by the ethical review board of the First Affiliated Hospital, Nanjing Medical University (IRB#2019-SR-153).

Acknowledgements
We thank all NDC participants, the former and present researchers in this study. We also thank all the members severed in the Nanjing Health Information Center.

Introduction

Diabetes can cause a series of complications and comorbidities including high blood pressure, depression, stroke, heart attack, retinopathy, and infections[1,2]. The population-based cohorts from true-world can help to study epidemiology, complications, risk factors, clinical course and treatment patterns of diabetes[3]. While the population-based studies regarding prevalence and burden of diabetes and diabetes-related complications are less concentrated in low-income and mid-income countries[4].

Large diabetes databases can help researchers improving the knowledge on strategies of diabetes prevention and diabetes epidemiological research. In China, the number of people with diabetes is 116.4 million and will reach 140.5 million in 2030[5]. According to a national cross-sectional study, the diabetes prevalence was increased from 10.9% (2013) to 12.4% (2018) in China[6]. There are several large population-based cohorts launched in China. China Health and Nutrition Survey (CHNS) is a nationally ongoing open cohort that was initialed in 1989[7]. It provides information of numerous changes in socioeconomic areas of families throughout 12 provinces in China. Risk Evaluation of Cancers in Chinese Diabetic Individuals, a Longitudinal (REACTION) study is aimed to explore the association of type 2 diabetes and pre-diabetes with the risk of cancer since 2011[8]. The Hong Kong Diabetes Surveillance Database (HKDSD) focus on the epidemiology of gestational diabetes, prediabetes and diabetes in Hong Kong, China[9]. It captured clinical information from electronic medical system since 2000.

Nanjing is the capital of Jiangsu Province, with 9.31 million residents according to data from China's seventh national census in 2021[10]. Nanjing is one of the most comprehensive industrial production bases and an important transportation hub in East China[11]. From 2011 to 2013, a survey on total of 4918 adult residents in Nanjing showed that the prevalence rate of metabolic syndrome was 26.3%[12]. The prevalence rates in 2 districts of Nanjing were estimated to be 6.0% for diabetes in 2017[13], which was lower than the national average (11.2%)[14]. Therefore, it is crucial to identify the diabetes risks and improve knowledge on strategies and epidemiological status for diabetes in Nanjing using large population-based cohorts.

The Nanjing Diabetes Cohort (NDC) is an electronic record of medical encounters, including all patients who meet the diabetes criteria since 2020 in health services providers of Nanjing. The NDC was established with the major aims of: (1) integrating EHR data for diabetes and complications in Nanjing; (2) exploring the clinical factors on diabetes progression; (3) developing individualized intervention strategies for the treatment of diabetes and comorbidities.

Cohort Description

Study setting and participants

In 2017, the Nanjing healthcare information platform was started to routinely collect clinical information for all people attending hospitals or primary healthcare institutions. As of 2020, there

were 650 medical institutions from 11 districts in NDC (Figure 1). A unique identification number is used in the platform to link EHR and other data sets, such as mortality records. To analyze the data anonymously, a unique number generated by platform is marked as clinical identifier for each patient.

We defined diabetes using following definition comprised of any of three criteria, according to the "Standards of medical care for type 2 diabetes in China 2019" by Chinese Diabetes Society (CDS)[15] and the "Standards of Medical Care in Diabetes" by American Diabetes Association (ADA)[16]. Patients who has (1) at least one encounter with haemoglobin A1c (HbA1c) level ≥6.5%[16] or (2) any encounters (inpatient or outpatient) with diagnosis of diabetes or (3) any encounters with prescriptions of diabetes medications. According to the recommendations on medications from CDS[15] and ADA[17], the diabetes medications include: dipeptidyl peptidase 4 (DPP-4) inhibitor, glucagon-like peptide 1 (GLP-1) receptor agonist (RA), α-glucosidase inhibitors, insulin, biguanides, sodium–glucose cotransporter 2 (SGLT2) inhibitor, sulphonylureas, thiazolidinedione, and traditional Chinese medicine (TCM).

Data collection

Baseline of cohort was defined as encounters for patients after 1st January 2020 from 11 districts (Figure 1). Top 3 districts which most data came from where hospitals or health services providers located are Gulou (30.79%), Qinhuai (21.99%), and Jiangning (11.95%). The cohort is ongoing with contiguous enrolment annually.

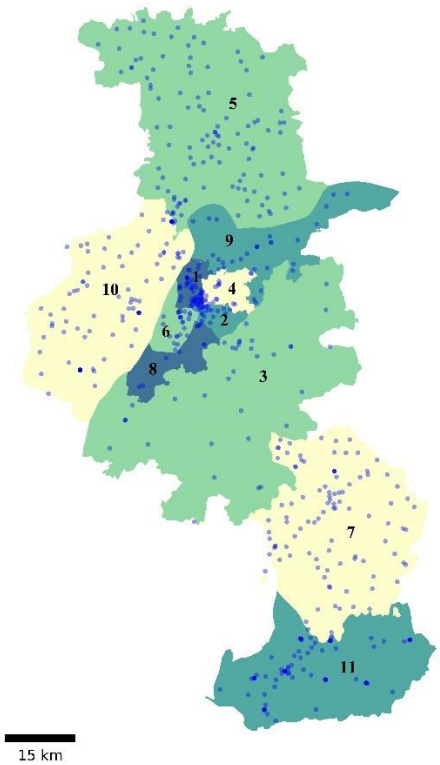

Figure 1 Distribution of medical institutions in NDC from 11 districts of Nanjing. Districts (sorted by the number of patients in each district): 1, Gulou; 2, Qinhuai; 3, Jiangning; 4, Xuanwu; 5, Luhe; 6, Jianye; 7, Lishui; 8, Yuhuatai; 9, Qixia; 10, Pukou; 11, Gaochun.

Each participant has unique clinical identifier. In Table 1, features for individual-level data in NDC including demographics, clinic visits for outpatients, admissions and discharges for inpatients, diagnoses, laboratory measurements and medications were obtained from the platform. Demographics including sex and date of birth are static variables. Laboratory measurements, medications, principal and secondary diagnosis codes are captured from both outpatients and inpatients. All patients are followed until death. The cause of death is stored as death report.

Table 1 Summary of measurements for patients in NDC.

Category	Measurements
Demographics	sex, date of birth
Outpatient clinic attendance	principal/secondary diagnosis codes (ICD-10)
Hospital admission	admission report, principal/secondary diagnosis codes of admission (ICD-10)
Hospital discharge	admission report, discharge report, principal/secondary diagnosis codes of discharge (ICD-10), death report
Glycaemic indices	FPG, HbAlc, random and after-meal glucose
Lipid profile	TC, LDL, HDL
Kidney function	ALB, eGFR, Urine ACR
Medications	DPP-4 inhibitor,GLP-1 RA, α-glucosidase inhibitors, insulin , biguanides, SGLT2 inhibitor, sulphonylureas, thiazolidinedione, glinide, and TCM
Clinical notes	complaint, history of illness
General information for diabetes	classification of diabetes, BMI

FPG, fasting plasma glucose; TC, total cholesterol; LDL, low-density-lipoprotein cholesterol; HDL, high-density-lipoprotein cholesterol; ALB, urine trace albumin; eGFR, estimated glomerular filtration rate; ACR, ratio of urinary albumin and creatinine; BMI, body mass index; TCM, traditional Chinese medicine.

Laboratory measurements used in clinical encounters from the platform are automatically captured via clinical identifier number. These measurements include glycaemic indices (FPG, HbAlc, random and after-meal glucose), lipid profile (TC, LDL, and HDL), kidney function (ALB, eGFR, and Alb/Cr). Diagnosis codes during the encounter of hospitalizations or clinic visits are coded using ICD-10. Each encounter contains a principal diagnosis code and less than 9 secondary diagnosis code.

The NDC also captures diabetes comorbidities or complications, including hypertension, cerebrovascular disease (CVD), angina pectoris, myocardial infarction, heart failure, mental disorders (depression and anxiety), and other diabetes-related conditions (diabetic foot, diabetic retinopathy, kidney failure, ketoacidosis, and severe hypoglycaemia).

The medications for each encounter are also captured by NDC. Data are available on the type, dose, and fee of treatment with glucose-lowering drugs, including DPP-4 inhibitor, GLP-1 RA, α-

glucosidase inhibitors, insulin, biguanides, SGLT2 inhibitor, sulphonylureas, thiazolidinedione, glinide, and TCM. The names and categories of each drug can be found in Supplementary Material.

The classification of diabetes is processed through diagnosis, present history of diseases, and medication usage from platform. Total of 554296 participants with diabetes were included from platform after criteria of diabetes applied. Participants are classified into: type 1 diabetes mellitus (T1DM), type 2 diabetes mellitus (T2DM), gestational diabetes mellitus (GDM), and other diabetes, according to classification approach from CDS[15] and ADA[16]. GDM, T1DM, and other diabetes (monogenic diabetes, pancreatic diabetes, steroid diabetes, or treated with glucocorticoid/interferon α) were checked through diagnosis and history of diseases. To identify all patients with T2DM, patients aged >= 30 who diagnosed with diabetes related disease, has diabetes-related history, or prescribed at least one non-insulin medications before insulin were also classified as T2DM. At last, 49580 participants were classified as unclear.

Findings to date

The characteristics of the cohort recruited between 1st Jan, 2020 and 31th Dec, 2022 are summarized in Table 2. Overall, 94.58% (n=977880) of patients had T2DM, 0.47% (n=4818) had T1DM, 2.16% (n=22331) had GDM, 0.83% (8534) had other diabetes, and 1.97% (n=20341) is unknown due to the lack of diabetes-related information. Notably, 80.03% of patients in unclear group were female, and the median of FBG was 5.15 (IQR: 4.60-6.11). Most of patients in unclear group could suffer from GDM. The overall description of the data is detailed in Supplemental Table 1.

Table 2 Baseline characteristics of patients according to diabetes type in NDC, 2020-2022

	T1DM	T2DM	GDM	Other	Unclear	Total
Number of patients, N (%)	4818 (0.47)	977880 (94.58)	22331 (2.16)	8534 (0.83)	20341 (1.97)	1033904 (100.00)
Sex, N (%)						
Female	2311 (47.97)	465189 (47.57)	22331 (100.00)	4473 (52.41)	16279 (80.03)	510583 (49.38)
Male	2507 (52.03)	512691 (52.43)	0 (0.00)	4061 (47.59)	4062 (19.97)	523321 (50.62)
Age, median years (IQR)	37 (22-59)	65 (55-74)	33 (30-35)	57 (42-68)	26 (23-28)	64 (53-73)
Age group, years (%)						
<=40	2612 (54.21)	76132 (7.79)	21094 (94.46)	1842 (21.58)	20341 (100.00)	122021 (11.80)
[40, 60)	1119 (23.23)	317181 (32.44)	1214 (5.44)	3219 (37.72)	0 (0.00)	322733 (31.21)
[60, 80)	886 (18.39)	470809 (48.15)	14 (0.06)	2950 (34.57)	0 (0.00)	474659 (45.91)
>=80	201 (4.17)	113758 (11.63)	9 (0.04)	523 (6.13)	0 (0.00)	114491 (11.07)
Examination, median (IQR)						
Temperature, °C	36.39 (36.10-36.50)	36.43 (36.12-36.50)	36.49 (36.26-36.52)	36.45 (36.10-36.54)	36.50 (36.18-36.55)	36.43 (36.13-36.50)
SBP, mmHg	123.50 (116.00-130.00)	126.00 (120.00-131.67)	118.20 (110.00-121.56)	124.00 (115.50-130.79)	120.00 (110.00-123.67)	125.83 (120.00-131.46)

DBP, mmHg	77.50 (70.42-80.00)	78.33 (73.33-81.00)	74.33 (69.00-80.00)	78.00 (71.00-81.50)	77.83 (69.00-80.00)	78.25 (73.01-80.91)
Pulse rate, times/min	73.33 (70.00-81.00)	72.00 (70.00-77.50)	80.00 (72.00-89.00)	74.00 (68.75-82.25)	77.50 (72.00-88.00)	72.00 (70.00-78.00)
FBG, mmol/L	7.70 (6.10-11.72)	7.34 (6.30-8.89)	4.85 (4.50-5.25)	6.80 (5.80-8.40)	5.15 (4.60-6.11)	7.10 (5.90-8.62)
HDL cholesterol, mmol/L	1.32 (1.06-1.58)	1.23 (1.01-1.48)	1.68 (1.36-2.01)	1.21 (0.95-1.49)	1.30 (1.10-1.52)	1.23 (1.02-1.49)
LDL cholesterol, mmol/L	2.60 (2.04-3.26)	2.63 (2.04-3.24)	2.61 (2.12-3.18)	2.56 (1.95-3.19)	2.70 (2.19-3.28)	2.63 (2.04-3.24)
Total cholesterol, mmol/L	4.74 (3.93-5.54)	4.59 (3.85-5.36)	5.46 (4.71-6.33)	4.66 (3.81-5.56)	4.55 (3.94-5.26)	4.62 (3.87-5.39)
eGFR, mL/min/1.73 m^2	121.95 (94.59-151.00)	102.50 (75.00-124.60)	164.18 (145.59-185.06)	124.37 (99.07-151.16)	137.50 (119.03-157.85)	104.73 (77.20-128.19)
Urine ACR, mg/g	12.03 (6.50-25.37)	19.00 (8.25-78.30)	17.00 (8.96-36.11)	16.40 (8.60-45.68)	12.40 (6.00-76.75)	18.90 (8.20-75.40)
Urine albumin, mg/L	12.61 (5.66-28.27)	9.40 (5.72-20.57)	3.45 (2.72-4.69)	5.42 (4.09-7.89)	11.14 (4.33-17.42)	9.08 (5.60-20.30)
HbA1c, %	7.80 (6.70-9.60)	6.90 (6.10-8.10)	5.30 (5.07-5.67)	6.73 (6.00-8.17)	5.60 (5.30-6.60)	6.81 (6.04-8.08)
Random glucose, mmol/L	9.09 (6.39-13.45)	7.13 (5.76-9.13)	5.17 (4.59-5.91)	6.97 (5.66-9.20)	5.20 (4.72-6.08)	7.00 (5.63-9.00)
Glucose-lowering drugs, N (%)						
Insulin	1928 (40.02)	167543 (17.13)	1127 (5.05)	1818 (21.30)	589 (2.90)	173005 (16.73)
Biguanid (metformin)	489 (10.15)	276446 (28.27)	794 (3.56)	1923 (22.53)	5901 (29.01)	285553 (27.62)
α-glucosidase inhibitors	330 (6.85)	138024 (14.11)	74 (0.33)	877 (10.28)	117 (0.58)	139422 (13.49)
Sulfonylurea	203 (4.21)	179976 (18.40)	66 (0.30)	942 (11.04)	175 (0.86)	181362 (17.54)
SGLT-2 inhibitors	137 (2.84)	108050 (11.05)	140 (0.63)	783 (9.18)	339 (1.67)	109449 (10.59)
DPP-4 inhibitors	153 (3.18)	95676 (9.78)	75 (0.34)	686 (8.04)	118 (0.58)	96708 (9.35)
Thiazolidinedione	58 (1.20)	44948 (4.60)	101 (0.45)	307 (3.60)	425 (2.09)	45839 (4.43)
Glinide	42 (0.87)	29880 (3.06)	12 (0.05)	173 (2.03)	30 (0.15)	30137 (2.91)
GLP-1 RA	12 (0.25)	14792 (1.51)	123 (0.55)	154 (1.80)	963 (4.73)	16044 (1.55)
TCM	7 (0.15)	2023 (0.21)	6 (0.03)	24 (0.28)	1 (0.00)	2061 (0.20)

HDL, high-density lipoprotein; LDL, low-density lipoprotein; SBP, systolic blood pressure; DBP, diastolic systolic blood pressure; eGFR, estimated glomerular filtration rate; ACR, albumin creatinine ratio; FPG, fasting plasma glucose; DPP4, dipeptidyl peptidase-4; SGLT-2, sodium/glucose cotransporter 2; GLP, glucagon-like peptide-1; TCM, traditional Chinese medicine.

For patients with T1DM or T2DM, over 50% were male. The percentage of male was higher among patients with T2DM (52.43%) in the cohort than the rest types of diabetes, which is comparable to a Chinese cross-sectional study in 2018[14]. Median age of patients in GDM (33 years) and T1DM (37 years) group were lower than T2DM (65 years). This likely reflects the classification process of

patients for T2DM. About 7.79% (n=76132) of T2DM patients were under 40, which is lower than the early onset of diabetes in Beijing (10.9%)[18]. This might result from environmental and socioeconomic of different regions and cities[6,14,19].

Median DBP levels were similar among patients with T2DM (78.33 mm Hg, IQR: 73.33-81.00 mm Hg), T1DM (77.50 mm Hg, IQR: 70.42-80.00 mm Hg), and other diabetes (78.00 mm Hg, IQR: 71.00-81.50 mm Hg). And SBP showed the similar pattern, while it was much lower in GDM (118.20 mm Hg, IQR: 110.00-121.56 mm Hg). Among patients with GDM, Median HDL (1.68 mmol/L, IQR: 1.36-2.01 mmol/L) and Total cholesterol level (5.46 mmol/L, IQR: 4.71-6.33 mmol/L) were higher than other types of diabetes. The median HbA1c were higher in T1DM (7.6%, IQR: 6.6-9.1%) compared to T2DM (7.1%, IQR: 6.3-8.4%). Markedly elevated glucose levels are often showed in patients with T1DM[20]. Also, median random glucose level was higher in patients with T1DM (9.09 mmol/L, IQR: 6.39-13.45 mmol/L) than T2DM (7.13 mmol/L, IQR: 5.76-9.13 mmol/L). In addition to insulin treatment (16.73%), majority of prescriptions were on first-line and second-line medications (metformin and sulfonylurea) in T2DM.

As shown in Table 3, the prevalence rate per 1000 person-years of hypertension is higher among patients with T2DM (297.68) than the other comorbidities, which is comparable to a Chinese cohort[21]. The prevalence rates of angina pectoris (58.43) and CVD (147.46) were higher in T2DM compared to T1DM. While the prevalence rates of diabetic retinopathy and ketoacidosis were higher in T1DM.

Table 3 Number of and incidence rates for diabetes-related conditions during follow-up according to diabetes type, 2020-2022.

	T1DM		T2DM		GDM		Total	
	No. of events (%)	Rate per 1000 person-years	No. of events (%)	Rate per 1000 person-years	No. of events (%)	Rate per 1000 person-years	No. of events (%)	Rate per 1000 person-years
Angina Pectoris	195 (4.05)	22.60	101605 (10.39)	58.43	19 (0.09)	1.65	102604 (9.92)	56.92
CVD	485 (10.07)	55.80	255055 (26.08)	147.46	85 (0.38)	4.97	257423 (24.90)	143.62
IHD	358 (7.43)	41.96	182683 (18.68)	106.48	42 (0.19)	3.89	184528 (17.85)	104.53
Hypertension	1051 (21.81)	122.47	507493 (51.90)	297.68	1154 (5.17)	58.68	514027 (49.72)	291.22
Hypoglycaemia	12 (0.25)	1.35	687 (0.07)	0.38	6 (0.03)	0.33	710 (0.07)	0.38
Ketoacidosis	308 (6.39)	33.92	1966 (0.20)	1.08	9 (0.04)	0.89	2435 (0.24)	1.29
Kidney disease	426 (8.84)	48.50	99259 (10.15)	57.27	140 (0.63)	7.40	100992 (9.77)	56.64
Diabetic retinopathy	374 (7.76)	42.42	23673 (2.42)	13.26	50 (0.22)	2.86	24271 (2.35)	13.09
Diabetic foot	35 (0.73)	3.93	5827 (0.60)	3.22	7 (0.03)	0.20	5890 (0.57)	3.13

The Table 4 show the prevalence of patients in NDC from years 2020 to 2022. The number of diabetic patients in NDC had kept growing in the past 3 years. Among them, the number of T2DM patients increased from 529265 in 2020 to 660968. However, due to the uncertainty about the types of some diabetic patients, we assumed patients over 30 years old to have T2DM. Therefore, the number of T2DM patients may be further specified more accurately in the future.

Table 4 Data characteristics of patients in the NDC dataset, 2020-2022

	2020	2021	2022
Patients, N	544421	638319	689340
Gender, N (%)			
Female	260759 (47.90)	311475 (48.80)	332506 (48.24)
Male	283662 (52.10)	326844 (51.20)	356834 (51.76)
Age	66 (56-74)	65 (55-74)	65 (55-74)
Age Group, N (%)			
(0, 40]	39484 (7.25)	55476 (8.69)	59344 (8.61)
(40, 60]	160746 (29.53)	189227 (29.64)	203771 (29.56)
(60, 80]	276838 (50.85)	321656 (50.39)	345126 (50.07)
(80, 1000]	67353 (12.37)	71960 (11.27)	81099 (11.76)
Diabetes Type, N (%)			
T1DM	2693 (0.49)	3055 (0.48)	3266 (0.47)
T2DM	529265 (97.22)	610005 (95.56)	660968 (95.88)
GDM	1768 (0.32)	12549 (1.97)	11069 (1.61)
Other	4413 (0.81)	5156 (0.81)	5398 (0.78)
Unknown	6282 (1.15)	7554 (1.18)	8639 (1.25)
Laboratory records, median (IQR)			
Temperature, C	36.50 (36.25-36.50)	36.43 (36.10-36.50)	36.45 (36.20-36.50)
SBP, mmHg	126.00 (120.00-131.29)	126.14 (120.00-131.89)	126.25 (120.00-132.00)
DBP, mmHg	78.75 (73.33-81.00)	78.72 (73.75-81.00)	78.67 (73.67-80.77)
Pulse rate, times/min	72.00 (71.82-74.00)	72.00 (71.00-77.00)	72.00 (72.00-77.31)
FBG, mmol/L	7.30 (6.30-8.80)	7.06 (5.90-8.49)	7.10 (5.96-8.53)
HDL cholesterol, mmol/L	1.28 (1.06-1.55)	1.25 (1.04-1.51)	1.20 (0.99-1.45)
LDL cholesterol, mmol/L	2.65 (2.07-3.27)	2.62 (2.03-3.24)	2.59 (1.99-3.21)
Total cholesterol, mmol/L	4.55 (3.81-5.33)	4.63 (3.88-5.42)	4.61 (3.85-5.39)
eGFR, mL/min/1.73 m2	111.73 (90.20-134.00)	103.30 (76.10-126.80)	100.40 (70.00-124.07)
Urine ACR, mg/g	18.00 (8.00-68.60)	20.80 (9.40-90.00)	19.00 (8.10-80.00)
Urine albumin, mg/L	6.08 (4.77-8.21)	8.80 (5.57-20.10)	14.04 (6.39-23.59)
HbA1c, %	6.70 (5.90-7.83)	6.83 (6.10-8.00)	7.00 (6.27-8.20)
Random glucose, mmol/L	6.70 (5.40-8.68)	7.07 (5.76-9.00)	7.30 (5.99-9.28)
Diabetes-related conditions，N (%)			
Angina Pectoris	46781 (8.59)	55690 (8.72)	43442 (6.30)
CVD	124905 (22.94)	142007 (22.25)	105000 (15.23)
Heart Failure	3754 (0.69)	5055 (0.79)	3472 (0.50)
IHD	90743 (16.67)	124808 (19.55)	89872 (13.04)

Hypertension	289760 (53.22)	322623 (50.54)	246383 (35.74)
Hypoglycaemia	223 (0.04)	275 (0.04)	213 (0.03)
Ketoacidosis	635 (0.12)	876 (0.14)	588 (0.09)
Kidney disease	43306 (7.95)	68175 (10.68)	52214 (7.57)
Diabetic retinopathy	7763 (1.43)	11253 (1.76)	8332 (1.21)
Diabetic foot	1498 (0.28)	2563 (0.40)	1709 (0.25)

Strengths and limitations

There are several features of the NDC that reflect the strengths. First, the NDC is an ongoing population-base, large-sized cohort that contains over 550000 patients with type 1 diabetes, type 2 diabetes, gestational diabetes, and other diabetes with follow-up records. It would provide data for future studies of diabetes duration with medication evaluation for a large population. Second, the NDC coverts the diabetes patients from Nanjing with urban and suburban areas, guaranteeing the generalizability of coverage. Each diabetic patient has longitudinal reliable medical records, including medical record data that had been verified manually. Third, the NDC captures patients with diabetes continuously from the Nanjing healthcare information platform, and provides exhaustive clinical data on clinical report, laboratory results, glucose-lowering drugs, diabetes-related comorbidities, and death report.

The NDC also has several limitations. First, patients who had done no measurement of glycemic indexes or suffered prediabetes with normal HbAlc would not be included in the NDC. Second, given the nature of administrative data, not everyone can complete all measurements. The missing variables are inevitable in EHR. Third, NDC captured data from health services providers in Nanjing only. Patients who have emigrated out of Nanjing could censor the follow-up.

Follow-up schedule

The cohort follow-ups will consist of annual checks of Nanjing healthcare information platform. The baseline data was linked to the platform. All encounters from the same participants and new participants who meet the criteria of diabetes would be captured. In order to keep track patients' medical events and medication usage, the NDC will censor only at death.

Patient and public involvement

No patient involved.

References

1 Chinese Diabetes Society. Guidelines for the prevention and control of type 2 diabetes in China (2017 Edition). *Chin J Pract Intern Med* 2018;**38**:292–344.

2 Poudel A, Zhou JY, Story D, *et al.* Diabetes and Associated Cardiovascular Complications in American Indians/Alaskan Natives: A Review of Risks and Prevention Strategies. *J Diabetes Res* 2018;**2018**:2742565. doi:10.1155/2018/2742565

3 Wild S, Fischbacher C, McKnight J. Using Large Diabetes Databases for Research. *J Diabetes Sci Technol* 2016;**10**:1073–8. doi:10.1177/1932296816645120

4 Harding JL, Pavkov ME, Magliano DJ, *et al.* Global trends in diabetes complications: a review of current evidence. *Diabetologia* 2019;**62**:3–16. doi:10.1007/s00125-018-4711-2

5 International Diabetes Federation. IDF Diabetes Atlas, 9th edn. Brussels Belgium: 2019. https://www.diabetesatlas.org

6 Wang L, Peng W, Zhao Z, *et al.* Prevalence and Treatment of Diabetes in China, 2013-2018. *JAMA* 2021;**326**:2498–506. doi:10.1001/jama.2021.22208

7 Zhang B, Zhai FY, Du SF, *et al.* The China Health and Nutrition Survey, 1989-2011: China Health and Nutrition Survey. *Obes Rev* 2014;**15**:2–7. doi:10.1111/obr.12119

8 Ning G, Group TRS. Risk Evaluation of cAncers in Chinese diabeTic Individuals: A lONgitudinal (REACTION) study. *J Diabetes* 2012;**4**:172–3. doi:10.1111/j.1753-0407.2012.00182.x

9 Wu H, Lau ESH, Yang A, *et al.* Data Resource Profile: The Hong Kong Diabetes Surveillance Database (HKDSD). *Int J Epidemiol* 2022;**51**:e9–17. doi:10.1093/ije/dyab252

10 Nanjing Municipal Bureau of Statistics. Communiqué of the seventh National Population Census of Nanjing. 2021.http://tjj.nanjing.gov.cn/bmfw/njsj/202105/t20210524_2945571.html (accessed 29 Sep 2022).

11 Li H, Qian X, Hu W, *et al.* Chemical speciation and human health risk of trace metals in urban street dusts from a metropolitan city, Nanjing, SE China. *Sci Total Environ* 2013;**456–457**:212–21. doi:10.1016/j.scitotenv.2013.03.094

12 Sun R, Guo B, Xie G, *et al.* [Prevalence and epidemic characteristics of metabolic syndrome among adult residents in Nanjing City in 2010-2013]. *Wei Sheng Yan Jiu* 2019;**48**:61–75.

13 Lu S, Bao M-Y, Miao S-M, *et al.* Prevalence of hypertension, diabetes, and dyslipidemia, and their additive effects on myocardial infarction and stroke: a cross-sectional study in Nanjing, China. *Ann Transl Med* 2019;**7**:436–436. doi:10.21037/atm.2019.09.04

14 Li Y, Teng D, Shi X, *et al.* Prevalence of diabetes recorded in mainland China using 2018 diagnostic criteria from the American Diabetes Association: national cross sectional study. *BMJ* 2020;**369**:m997. doi:10.1136/bmj.m997

15 Jia W, Weng J, Zhu D, *et al.* Standards of medical care for type 2 diabetes in China 2019. *Diabetes Metab Res Rev* 2019;**35**:e3158. doi:10.1002/dmrr.3158

16 American Diabetes Association. 2. Classification and Diagnosis of Diabetes: *Standards of Medical Care in Diabetes—2019*. *Diabetes Care* 2019;**42**:S13–28. doi:10.2337/dc19-S002

17 American Diabetes Association. 9. Pharmacologic Approaches to Glycemic Treatment: Standards of Medical Care in Diabetes-2021. *Diabetes Care* 2021;**44**:S111–24. doi:10.2337/dc21-S009

18 Wang Z, Wu Y, Wu J, *et al.* Trends in prevalence and incidence of type 2 diabetes among adults in Beijing, China, from 2008 to 2017. *Diabet Med* 2021;**38**:e14487. doi:10.1111/dme.14487

19 Yang W, Lu J, Weng J, *et al.* Prevalence of Diabetes among Men and Women in China. *N Engl J Med* 2010;**362**:1090–101. doi:10.1056/NEJMoa0908292

20 Dabelea D, Rewers A, Stafford JM, *et al.* Trends in the Prevalence of Ketoacidosis at Diabetes Diagnosis: The SEARCH for Diabetes in Youth Study. *Pediatrics* 2014;**133**:E938–45. doi:10.1542/peds.2013-2795

21 Xuebing Z, Yiming M, Wenhua Y, *et al.* Prevalence of stroke and metabolic disorders in the middle-aged and elderly Chinese with type 2 diabetes. *Chin Med J (Engl)* 2014;**127**:3537–42. doi:10.3760/cma.j.issn.0366-6999.20141018