BMJ Open Diabetes Research & Care

Using the IMEDS distributed database for epidemiological studies in type 2 diabetes mellitus

Ting-Ying Huang ![ORCID],[1] Carla Rodriguez-Watson,[2] Tongtong Wang ![ORCID],[3] Shawna R Calhoun,[3] James Marshall,[4] Jillian Burk,[4] Young Hee Nam,[1] Aaron B Mendelsohn,[1] Aziza Jamal-Allial,[5] Robert T Greenlee,[6] Mano Selvan,[7] Pamala A Pawloski,[8] Cheryl N McMahill Walraven,[9] Ashish Rai,[1] Sengwee Toh ![ORCID],[1] Jeffery S Brown[1]

Correspondence to
Dr Sengwee Toh;
Darren_Toh@harvardpilgrim.org

ABSTRACT

Introduction This study aimed to assess data relevancy and data quality of the Innovation in Medical Evidence Development and Surveillance System Distributed Database (IMEDS-DD) for diabetes research and to evaluate comparability of its type 2 diabetes cohort to the general type 2 diabetes population.

Research design and methods A retrospective study was conducted using the IMEDS-DD. Eligible members were adults with a medical encounter between April 1, 2018 and March 31, 2019 (index period). Type 2 diabetes and co-existing conditions were determined using all data available from April 1, 2016 to the most recent encounter within the index period. Type 2 diabetes patient characteristics, comorbidities and hemoglobin A_{1c} (HbA_{1c}) values were summarized and compared with those reported in national benchmarks and literature.

Results Type 2 diabetes prevalence was 12.6% in the IMEDS-DD. Of 414 672 patients with type 2 diabetes, 52.8% were male, and the mean age was 65.0 (SD 13.3) years. Common comorbidities included hypertension (84.5%), hyperlipidemia (82.8%), obesity (45.3%), and cardiovascular disease (44.7%). Moderate-to-severe chronic kidney disease was observed in 20.2% patients. The most commonly used antihyperglycemic agents included metformin (35.7%), sulfonylureas (14.8%), and insulin (9.9%). Less than one-half (48.9%) had an HbA_{1c} value recorded. These findings demonstrated the notable similarity in patient characteristics between type 2 diabetes populations identified within the IMEDS-DD and other large databases.

Conclusions Despite the limitations related to HbA_{1c} data, our findings indicate that the IMEDS-DD contains robust information on key data elements to conduct pharmacoepidemiological studies in diabetes, including member demographic and clinical characteristics and health services utilization.

INTRODUCTION

The Innovation in Medical Evidence Development and Surveillance Distributed Database (IMEDS-DD) is a subset of the Food and Drug Administration (FDA) Sentinel Distributed Database. The IMEDS is a public-private

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Distributed database network has become a new type of data source for observational diabetes research.
⇒ The Sentinel Distributed Database (SDD) of the Sentinel System, a national electronic active surveillance system for medical product safety established under the US Food and Drug Administration's Sentinel Initiative, is one successful example.

WHAT DOES THIS STUDY ADD

⇒ As a subset of SDD, the Innovation in Medical Evidence Development and Surveillance System Distributed Database (IMEDS-DD) reliably captures key data elements to conduct pharmacoepidemiological studies in patients with type 2 diabetes, including demographics, comorbidities, and health services utilization—substantial overlap was found in patient characteristics between such population from the IMEDS-DD and those included in published benchmarks or within other large databases.

HOW MIGHT THIS STUDY AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The IMEDS-DD, accessible through the IMEDS program, is a new data source for epidemiological studies related to type 2 diabetes and its management.

partnership launched in 2017 by the Reagan-Udall Foundation for the FDA, an independent, non-profit organization created by the United States Congress, to advance the FDA's mission by promoting regulatory science. The IMEDS provides a framework for private-sector entities (eg, regulated industry, academic institutes) to leverage the FDA Sentinel System, a national electronic system established under the Sentinel Initiative[1 2] for active safety surveillance of medical products including drugs, biologics, vaccines, and medical devices. The policies and procedures for using the IMEDS-DD for

observational research are available on the foundation's website (https://reaganudall.org/programs/research/about-imeds).

The IMEDS-DD comprises selected network partners and uses routinely collected administrative claims and laboratory result data from the Sentinel Distributed Database, including data elements commonly available in database studies such as demographics, health plan enrollment, diagnoses, procedures, outpatient pharmacy dispensing records, and laboratory results. As of early 2021, the IMEDS-DD had access to data available for over 111 million person-lives across 9 health plan partners and is expected to be largely representative of the commercially insured population, including employer-sponsored health plan and Medicare Advantage members, in the USA. The IMEDS-DD shares the same data management, common data model, privacy protection methods, and quality assurance procedures with the Sentinel Distributed Database,[3–7] and the same secure distributed querying approach with the Sentinel System . As such, the IMEDS-DD inherited the data curation process and infrastructure standards, as well as the privacy-preserving techniques and analytic tools, of the FDA Sentinel System.

Although the Sentinel Distributed Database and data from many of its network partners have been used widely in epidemiological studies for type 2 diabetes, the IMEDS-DD has not previously been used for the same purpose. This study aimed to serve as a feasibility assessment of using the IMEDS-DD for observational diabetes research. Specifically, the study evaluated data relevancy and data quality of the database and examined availability of key data elements in epidemiological studies for type 2 diabetes as well as comparability of such to those provided by other healthcare databases.

RESEARCH DESIGN AND METHODS
Study design and study population
This observational study adopted a non-interventional, retrospective design and examined demographic and clinical characteristics of patients with type 2 diabetes identified from the IMEDS-DD. This study included data from seven national and regional health insurers participating in the IMEDS-DD in the USA: CVS Health/Aetna, Harvard Pilgrim Health Care, HealthCore/Elevance Health, HealthPartners, Humana, Marshfield Clinic Health System, and TennCare. Eligible members had to meet the following criteria up until the date of their most recent medical encounter (referred to as the 'index date', regardless of care settings) between April 1, 2018 and March 31, 2019 (index period): 18 years of age or older, at least 6 months continuous enrollment in medical and prescription drug insurance plans (maximum allowable enrollment gap of 45 days), and at least one *E11.x Type 2 Diabetes Mellitus* plus no *E10.x Type 1 Diabetes Mellitus* or *O24.4x Gestational Diabetes Mellitus in Pregnancy* International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) diagnosis recorded in claims

any time since April 1, 2016 through the end of the index period (see design diagram in figure 1).

Patient characteristics assessment
The data availability was first assessed via characterization of the above cohort by key dimensions in diabetes research, including demographics, antihyperglycemic treatment by drug class (metformin, insulin, sulfonylurea, thiazolidinediones, dipeptidyl peptidase-4 (DPP-4) inhibitors, glucagon-like peptide-1 (GLP-1) receptor agonist; sodium-glucose cotransporter-2 (SGLT-2) inhibitors, and others (alpha-glucosidase inhibitors, meglitinides), comorbidities, general health services utilization, and hemoglobin A_{1c} (HbA_{1c}) values. The evaluation period was from April, 2016 to the index date except for demographics and antihyperglycemic treatment by drug class, which were evaluated on the index date (figure 1). For antihyperglycemic treatments, drug class utilization was determined by any dispensing with overlapping days supply on the index date, including dispensing on the index date.

This study used outpatient pharmacy dispensings to define drug utilization and medical encounter claims to define existing conditions or medical history. Specifically, individual drugs were identified using the National Drug Codes; medical conditions were identified using algorithms based on diagnosis and procedure codes encoded in the following systems: ICD-9-CM/ICD-10-CM, ICD-10-Procedure Coding System, Healthcare Common Procedure Coding System, and Current Procedural Terminology codes.

Presence of comorbid conditions was assessed using all available data prior to and including the index date. The following conditions were evaluated: cardiovascular disease (CVD), moderate-to-severe chronic kidney disease (CKD, stages 3–5, assumed estimated glomerular filtration rate $<60\,mL/min/1.73\,m^{2}$[8]), retinopathy, nephropathy, neuropathy, amputation, hypertension, hypoglycemia, hyperlipidemia, obesity, and pancreatitis. CVD was categorized based on diagnoses for cerebrovascular disease, coronary heart disease, heart failure, myocardial infarction, peripheral artery disease, or stroke.

An exploratory analysis focused on a subset of the type 2 diabetes cohort members identified from the five (of the seven total) participating network partners who provided HbA_{1c} results. To explore the quality of HbA_{1c} data in the IMEDS-DD, the study summarized the total number patients with at least one HbA_{1c} value recorded during the study period, as well as average testing intervals among those who had at least two HbA_{1c} test results. The study further examined ranges of the most recent test per patient.

Statistical analyses
All analyses were completed using the Sentinel Routine Querying System version 9.4.0[9] with additional custom programming. Patient characteristics, comorbid
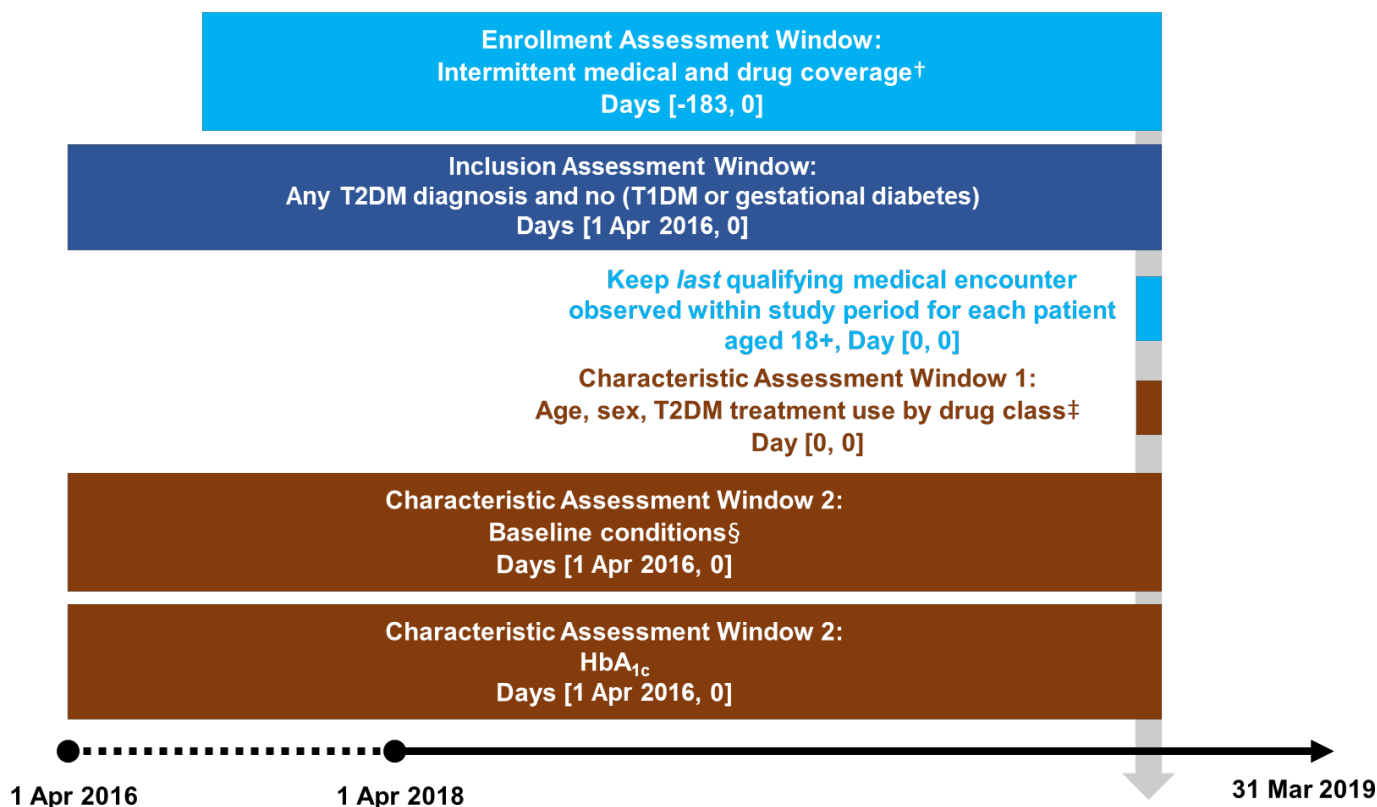
Figure 1 Cohort identification strategy used in this study. (*) Cohort entry date was defined by date of the *last* medical encounter during the period from April 1, 2018 to March 31, 2019. Members were only allowed to enter the analytic cohort *once*. (†) Health plan enrollment might start before April 1, 2018. Up to 45-day gap(s) in medical or prescription drug plan enrollment was allowed. (‡) Evidenced by dispensing date or days supply (indicator of ongoing drug use). T2DM treatment included metformin, insulin, sulfonylurea, thiazolidinediones, dipeptidyl peptidase-4 inhibitors, glucagon-like peptide-1 receptor agonist; sodium-glucose cotransporter-2 inhibitors, and others (alpha-glucosidase inhibitors, meglitinides). (§) Comorbid conditions included: combined comorbidity score, cardiovascular disease (cerebrovascular disease, coronary heart disease, heart failure, myocardial infarction, stroke, peripheral artery disease), hypertension, hyperlipidemia, hypoglycemia, hypovolemia, hypoxemia, obesity, pancreatitis, renal insufficiency (moderate-to-severe chronic kidney disease), diabetic complications (nephropathy, neuropathy, retinopathy, amputation), number of unique antihyperglycemics/non-antihyperglycemics/physician visits/hospitalizations. HbA_{1c}, hemoglobin A_{1c}; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus.

conditions, and HbA_{1c} availability were summarized via descriptive analyses. Continuous variables were reported as means and SDs, and categorical variables were summarized as number and proportion of the total study population in each cohort.

The study then evaluated the comparability of measured characteristics of patients with type 2 diabetes in the IMEDS-DD with those identified in other populations or data sources in the USA and other countries.

RESULTS
Characteristics of patients with type 2 diabetes in the IMEDS-DD

A total of 414 672 (12.6%) patients with type 2 diabetes were identified from the 3 280 646 active members in the IMEDS-DD who had a medical encounter between April 1, 2018 and March 31, 2019 at the time the data were accessed for this study. In these patients, 52.8% were

male, and the mean age was 65.0 years (SD 13.3 years). Persons aged 75 years or older comprised 24.1% of this cohort (table 1). The race composition was 40.4% white, 9.9% black, and 46.9% unknown.

Common comorbid conditions were hypertension (84.5%) and hyperlipidemia (82.8%). Around two in every five patients had a history of obesity (45.3%) or existing CVD (44.7%), including coronary heart disease (31.2%), peripheral artery disease (18.1%), cerebrovascular disease (17.4%), heart failure (16.9%), and stroke (15.3%). Nearly one-half (47.7%) of patients experienced at least one diabetic complication, with nephropathy (31.8%) and neuropathy (25.4%) being the most prevalent. Prevalence of moderate-to-severe CKD was 20.2%.

Antihyperglycemic treatments utilization included metformin (35.7%), followed by sulfonylureas (14.8%), insulin (9.9%), and DPP-4 inhibitors (6.0%). Between

Table 1 Characteristics of patients with type 2 diabetes identified in the IMEDS Distributed Database between April 1, 2018 and March 31, 2019

	N/Mean	%/SD
Number of patients	414 672	100.0%
Demographics		
Age, years	65.0	13.3
18–44	35 361	8.5%
45–64	168 013	40.5%
65–74	111 366	26.9%
≥75	99 932	24.1%
Male sex	218 759	52.8%
Ethnicity		
Hispanic	10 843	2.6%
Non-Hispanic	156 168	37.7%
Unknown	247 661	59.7%
Comorbid conditions		
Combined comorbidity score[23]	3.6	4.0
Cardiovascular disease	185 527	44.7%
Cerebrovascular disease	72 350	17.4%
Coronary heart disease	129 293	31.2%
Heart failure	70 105	16.9%
Myocardial infarction	48 488	11.7%
Peripheral artery disease	75 018	18.1%
Stroke	63 489	15.3%
Hypertension	350 295	84.5%
Hyperlipidemia	343 350	82.8%
Hypoglycemia	19 579	4.7%
Obesity	187 841	45.3%
Pancreatitis	7092	1.7%
Moderate-to-severe chronic kidney disease	83 653	20.2%
Diabetic complications	197 853	47.7%
Nephropathy	131 991	31.8%
Neuropathy	105 188	25.4%
Retinopathy	45 380	10.9%
Amputation	2 976	0.7%
Ongoing use of antihyperglycemic agents		
Metformin	147 954	35.7%
Sulfonylureas	61 395	14.8%
Insulin	41 148	9.9%
DPP-4 inhibitors	24 797	6.0%
SGLT-2 inhibitors	16 231	3.9%
GLP-1 agonists	16 366	3.9%
Thiazolidinediones	10 426	2.5%
Others (alpha-glucosidase inhibitors, meglitinides)	1 431	0.3%
Health services utilization		
Number of antihyperglycemics	1.3	1.4
Number of non-antihyperglycemics	12.4	9.0
Number of inpatient hospital encounters	0.7	1.7
Number of ambulatory encounters	36.6	41.1
Most recent HbA$_{1c}$ test result*		
No HbA$_{1c}$ test results	212 049	51.1%

Continued

Table 1 Continued

	N/Mean	%/SD
Most recent HbA$_{1c}$ test result, %	7.0	3.2
<7.0	129 071	63.7%†
7.0–7.9	34 617	17.1%†
8.0–8.9	16 660	8.2%†
≥9.0	22 275	11.0%†

Row values represent mean and SD where no % follows.
*Exploratory analyses using data from five network partners with HbA$_{1c}$ values available.
†Percentage calculated from patients with most recent HbA$_{1c}$ results.
DPP-4, dipeptidyl peptidase-4; GLP-1, glucagon-like peptide-1; HbA$_{1c}$, hemoglobin A$_{1c}$; IMEDS, Innovation in Medical Evidence and Development Surveillance; SGLT-2, sodium-glucose cotransporter-2.

April 1, 2016 and their most recent medical encounter, patients with type 2 diabetes used an average of one antihyperglycemic drug (mean 1.3 drugs) and had frequent ambulatory visits (mean 36.6 visits).

Of the 414 672 patients with type 2 diabetes identified, 48.9% had at least one HbA$_{1c}$ value recorded. Among these, the mean of the most recent value was 7.0% (SD 3.2%), and 36.3% of these results were ≥7.0%. Of 146 418 patients with two or more HbA$_{1c}$ results recorded, 8.6% had an average testing interval within 90 days, and 48.3% within 91–183 days.

Comparability of patients with type 2 diabetes in the IMEDS-DD versus in the other data sources commonly used for diabetes research

Comparability of the IMEDS-DD patients with type 2 diabetes to those in the general population was summarized by population prevalence and patient demographic (table 2), by comorbidity and health services utilization (table 3), as well as by antihyperglycemic utilization and HbA$_{1c}$ value (table 4). The 12.6% type 2 diabetes prevalence in the IMEDS-DD population is generally consistent with estimates by the most recent Centers for Disease Control and Prevention (CDC) National Diabetes Statistics Report[10] and the International Diabetes Federation (IDF) Diabetes Atlas[11]: the CDC estimated a 13.0% diabetes prevalence in the US population using data from the National Health and Nutrition Examination Survey (NHANES) and a survey respondent self-report approach (ie, 'being told by a doctor or health professional that they had diabetes'), whereas the IDF estimated a 13.3% diabetes prevalence in the US population using data gathered for the IDF Diabetes Atlas.

Patients with type 2 diabetes in the IMEDS-DD had a mean age over 60 years with similar proportions of female and male patients. On average, these estimates align with findings from both nationwide surveys such as NHANES[10 12] and large databases frequently used for diabetes research including the IBM MarketScan databases in the USA,[13] the Canadian Network for Observational Drug Effect Studies system,[14] the Clinical Practice Research Datalink in the UK,[15] and the Scottish Care Information-Diabetes system (SCI-Diabetes).[16]

Table 2 Comparability of patients with type 2 diabetes between the IMEDS Distributed Database and other data sources commonly used for diabetes research: prevalence and demographics

	IMEDS-DD	Other databases	References Source	Data	Study period
Prevalence					
Population prevalence of type 2 diabetes	12.6%	13.0%	Centers for Disease Control and Prevention (CDC) National Diabetes Statistics Report[10]	National Health and Nutrition Examination Survey (NHANES)	2013–2016
		13.3%	International Diabetes Federation (IDF) website[11]	IDF Diabetes Atlas, 9th edition, 2019	2019
Demographics of patients with type 2 diabetes					
Age, years (mean, SD)	65.0 (13.3)	63.9 (0.2)	Bailey et al[12]	NHANES	1999–2012
		60.9 (12.8)	Weng et al[13]	IBM MarketScan Research Databases	2015
		63.8 (9.5)	Douros et al[14]	Canadian Network for Observational Drug Effect Studies: 7 Canadian provincial databases and Clinical Practice Research Datalink (CPRD)	2013–2018
		66.4 (12.3)	Tang et al[15]	CPRD	2012–2014
18–44	8.5%	14.4%	CDC National Diabetes Statistics Report[10]	NHANES	2013–2016
45–64	40.5%	43.4%			
≥65	51.0%	41.9%			
Sex: male	52.8%	52.5%	CDC National Diabetes Statistics Report[10]	NHANES	2013–2016
		56.4%	McGurnaghan et al[16]	Scottish Care Information-Diabetes system (SCI-Diabetes)	2014–2016

IMEDS, Innovation in Medical Evidence and Development Surveillance.

Over 40% of patients with in the IMEDS-DD had CVD, comparable to 45.2% in finding of Weng et al from the IBM MarketScan databases.[13] The percentage of study patients with heart failure was within the range of estimates established by the American Heart Association for patients with type 2 diabetes[17] (9%–22%). The prevalences of moderate-to-severe CKD,[10 12 16 18] neuropathy,[19 20] and retinopathy[10 13] were also in line with prior characterization of the type 2 diabetes population in literature. Similarly, consistency was observed in health services utilization,[21] including the number of antihyperglycemic agents used.[18 22]

In general, the most commonly observed antihyperglycemic treatment in the IMEDS-DD patients with type 2 diabetes—metformin, sulfonylurea, insulin, and DPP-4 inhibitors—largely match published findings based on SCI-Diabetes[16] and data from various electronic health record (EHR) systems, such as the National Patient-Centered Clinical Research Network (PCORnet)[22] and the Cleveland Clinic EHR system.[19] The utilization proportions of individual drug class, however, vary by data source, evaluation period, or definition of patients with type 2 diabetes. Antihyperglycemic treatment utilization observed in the study was generally lower than possible observations restricted to treated patients with type 2 diabetes only.[18]

Both the availability and mean values of the most recent HbA_{1c} test result in the IMEDS-DD for patients with type 2 diabetes were comparable to findings in published studies. Specifically, Bachmann et al also found that around one-half of patients with type 2 diabetes in the PCORnet had one or more documented HbA_{1c} test result,[22] and Pantalone et al, McGurnaghan et al, Iglay et al, respectively observed an average HbA_{1c} value of 7.0% in data from Cleveland Clinic EHR system,[19] SCI-Diabetes,[16] and Quintiles Electronic Medical Record research database.[18]

CONCLUSIONS

This study assessed data quality of the IMEDS-DD and demonstrated feasibility of epidemiological studies for type 2 diabetes in this database. Availability of key data elements was generally high in the IMDES-DD for patients with type 2 diabetes. A descriptive summary of the study cohort suggests that basic demographics, comorbidities, and diabetes treatment or other health services utilization were reliably captured.

Findings from this comparability evaluation indicate notable similarities between the patients with type 2 diabetes in the IMEDS-DD and those in the general US patient populations. Despite slightly lower type 2 diabetes

Table 3 Comparability of patients with type 2 diabetes between the IMEDS Distributed Database and other data sources commonly used for diabetes research: comorbidities and health services utilization

	IMEDS-DD	Other databases	References Source	Data	Study period
Comorbid conditions in patients with type 2 diabetes					
Combined comorbidity score[23] (mean, SD)	3.6 (4.0)	2.4 (2.1)*	Weng et al[13]	IBM MarketScan Research Databases	2015
		2.5 (1.7)*	Tang et al[15]	CPRD	2012–2014
Cardiovascular disease	44.7%	45.2%	Weng et al[13]	IBM MarketScan Research Databases	2015
Heart failure	16.9%	9%–22%	Dunlay et al[17]	Type 2 Diabetes Mellitus and Heart Failure: A Scientific Statement from the American Heart Association and the Heart Failure Society of America	2019
		11.2%	Bailey et al[12]	NHANES	1999–2012
Moderate-to-severe chronic kidney disease	20.2%	24.9%	CDC National Diabetes Statistics Report[10]	NHANES	2013–2016
		24.1%	Iglay et al[18]	Quintiles Electronic Medical Record (EMR) research database (Q-EMR)	2014–2015
		22% (20.4%–23.5%)	Bailey et al[12]	NHANES	1999–2012
		22.3%–23.5%	McGurnaghan et al[16]	SCI-Diabetes	2014–2016
		18.9%	Tang et al[15]	CPRD	2012–2014
Neuropathy	25.4%	16.2%	Pantalone et al[19]	Cleveland Clinic electronic health record system	2008–2013
		28.5%	Gregg et al[20]	NHANES	1999–2000
Retinopathy	10.9%	11.7%	CDC National Diabetes Statistics Report[10]	Behavioral Risk Factor Surveillance System	2018
		12.1%	Weng et al[13]	IBM MarketScan Research Databases	2015
Health services utilization in patients with type 2 diabetes					
Number of unique antihyperglycemics (mean, SD)	1.3 (1.4)	42.2% on monotherapy	Bachmann et al[22]	National Patient-Centered Clinical Research Network (PCORnet)	2012–2017
		40.6% on monotherapy	Iglay et al[18]	Q-EMR	2014–2015
Number of unique non-antihyperglycemics (mean, SD)	12.4 (9.0)	8.3 (7.0)–10.1 (8.3)	Mehta et al[21]	Q-EMR linked to administrative claims from the Truven Health MarketScan database	2012
Number of inpatient hospital encounters (mean, SD)	0.7 (1.7)	0.2 (0.5)–0.4 (0.8)			
Number of ambulatory encounters (mean, SD)	36.6 (41.1)	17.8 (18.6)–27.5 (25.0)			

*Charlson Comorbidity Index.[24]
CDC, Centers for Disease Control and Prevention; CPRD, Clinical Practice Research Datalink; IMEDS, Innovation in Medical Evidence and Development Surveillance; IMEDS-DD, IMEDS Distributed Database; NHANES, National Health and Nutrition Examination Survey; SCI-Diabetes, Scottish Care Information-Diabetes system.

prevalence estimate in the IMEDS-DD than the CDC and IDF benchmarks, the difference was expected and may be attributed to the broader diabetes definitions used in the references (ie, mixed diabetes definition of types 1 and 2 diabetes). Given that patients with type 2 diabetes typically account for over 90% of the diabetes patients in total,[10] the 12.6% prevalence estimate in the IMEDS-DD population remains comparable to that of the US population. With regard to the overall age difference, the older patients with type 2 diabetes in the IMEDS-DD may be a database strength that can be leveraged to enhance the generalizability of future IMEDS studies. In particular, the high proportion of older patients could be helpful to address study questions focusing on patients treated with second-line antihyperglycemics, a subpopulation which is often characterized by advanced age.[14]

While administrative health insurance claims databases provide comprehensive and longitudinal records of encounters within the healthcare system and outpatient pharmacy dispensings, they may not contain sufficient clinical details. For example, while the performance of a laboratory test, such as HbA_{1c}, would generate a claim, the test result is not always available within the claims database. This is no exception to the IMEDS-DD. This feasibility assessment shows that the most recent HbA_{1c} result was only available among 48.9% of the IMEDS-DD study population. Although HbA_{1c} was not available for half of the study population, for epidemiological studies of type 2 diabetes, other detectable comorbidities/sequelae such as history of CVD, diabetic complications, and utilization of antihyperglycemic drugs may alternatively be considered as proxies for diabetes control or severity.

Table 4 Comparability of patients with type 2 diabetes between the IMEDS Distributed Database and other data sources commonly used for diabetes research: treatment and clinical characteristics

	IMEDS-DD	Other databases	References Source	Data	Study period
Ongoing use of antihyperglycemic agents in patients with type 2 diabetes					
Metformin	35.7%	32.6%	Bachmann et al[22]	PCORnet	2012–2017
Insulin	9.9%	11.3%	Bachmann et al[22]	PCORnet	2012–2017
		10.9%	McGurnaghan et al[16]	SCI-Diabetes	2014–2016
Sulfonylureas	14.8%	14.6%	Bachmann et al[22]	PCORnet	2012–2017
Thiazolidinediones	2.5%	0.6%	Bachmann et al[22]	PCORnet	2012–2017
		4.8%	McGurnaghan et al[16]	SCI-Diabetes	2014–2016
		0.6%	Pantalone et al[19]	Cleveland Clinic electronic health record system	2008–2013
SGLT-2 inhibitors	3.9%	0.1%	Bachmann et al[22]	PCORnet	2012–2017
		2.0%	McGurnaghan et al[16]	SCI-Diabetes	2014–2016
		0.1%	Pantalone et al[19]	Cleveland Clinic electronic health record system	2008–2013
DPP-4 inhibitors	6.0%	7.0%	Bachmann et al[22]	PCORnet	2012–2017
		10.1%	McGurnaghan et al[16]	SCI-Diabetes	2014–2016
GLP-1 agonists	3.9%	2.2%	Bachmann et al[22]	PCORnet	2012–2017
		2.7%	McGurnaghan et al[16]	SCI-Diabetes	2014–2016
		2.2%	Pantalone et al[19]	Cleveland Clinic electronic health record system	2008–2013
Most recent HbA$_{1c}$ value* in patients with type 2 diabetes					
% with ≥1 HbA$_{1c}$ test results	48.9%*	54.0%	Bachmann et al[22]	PCORnet	2012–2017
HbA$_{1c}$, % (mean, SD)	7.0 (3.2)*†	7.0 (5.4, 8.6)‡	Pantalone et al[19]	Cleveland Clinic electronic health record system	2008–2013
		7.2 (6.5, 8.2)‡	McGurnaghan et al[16]	SCI-Diabetes	2014–2016
		6.8 (1.7)‡	Iglay et al[18]	Q-EMR	2014–2015

*Exploratory analyses using data from five network partners with HbA$_{1c}$ values available.
†Calculation based on patients who had a most recent HbA$_{1c}$ value.
‡Median (25th percentile, 75th percentile).
DPP-4, dipeptidyl peptidase-4; EMR, Electronic Medical Record; GLP-1, glucagon-like peptide-1; HbA$_{1c}$, hemoglobin A$_{1c}$; IMEDS, Innovation in Medical Evidence and Development Surveillance; IMEDS-DD, IMEDS Distributed Database; PCORnet, National Patient-Centered Clinical Research Network; Q-EMR, Quintiles EMR research databaseSCI-Diabetes, Scottish Care Information-Diabetes system ; SGLT-2, sodium-glucose cotransporter-2.

As with any other health insurance administrative and claims database, the IMEDS-DD provides comprehensive and longitudinal records of encounters within the healthcare system and of pharmacy dispensing records. Yet, the IMEDS-DD also shares general limitations of these databases. First, misclassification is possible due to the use of diagnosis, procedure, and drug codes for identification of specific medical conditions or drug exposure. For example, rule-out diagnoses are possible and may partially explain the low antihyperglycemic agent use among IMEDS-DD patients with type 2 diabetes, as the study might include patients with prediabetes and lower HbA_{1c} who did not yet need treatment. Second, limited clinical details were available to verify personal history and severity of the measured medical conditions. The sizable missingness of HbA_{1c} data may be due to the patient not having an HbA_{1c} measurement or the health plan not being provided the test result. Similarly, a high degree of missingness might exist in other laboratory data pertinent to diabetes research, such as kidney function and lipid profile. Third, most data were tailored to administrative or billing needs and thus reflected selected information and the use of covered health services only. Race/ethnicity information is sourced through health insurance demographic data and does not have complete capture. Utilization of the over-the-counter medications and free drug samples is unknown. Records may be incomplete for health services subject to bundle payment (eg, hospital admissions) due to lack of recording incentive. Substantial underestimation for lifestyle factors may exist. Dispensing records do not reflect the actual drug ingestion. Fourth, since the data presented in table 2 come from disparate sources and from studies with different protocols, detailed statistical assessments were not possible. As such, this study provides an overview of the IMEDS-DD data elements relevant to diabetes research and a broad comparison of the estimates of these elements in the IMEDS-DD with data from other sources. Insofar as the estimates for certain key parameters are broadly similar between the IMEDS-DD and other databases of different types, the IMEDS-DD serves as a potential data source for observational diabetes research. Lastly, study results are only generalizable to the commercial health insurance population from which the study population was derived as well as other populations sharing similar characteristics.

In conclusion, the IMEDS-DD contains key data elements to conduct pharmacoepidemiological studies in patients with type 2 diabetes, including demographics, clinical characteristics, and health services utilization. This study found substantial overlap in patient characteristics between the patients with type 2 diabetes identified from the IMEDS-DD population and those included in published benchmarks or within other large databases. Despite the general limitations related to claims data such as the lack of HbA_{1c} data, this study supports the IMEDS-DD as an appropriate data source for epidemiological studies related to type 2 diabetes and its management.

Author affiliations
[1]Department of Population Medicine, Harvard Medical School, Boston, Massachusetts, USA
[2]Reagan-Udall Foundation for the Food and Drug Administration, Washington, District of Columbia, USA
[3]Epidemiology, Merck & Co, Inc, Rahway, New Jersey, USA
[4]Department of Population Medicine, Harvard Pilgrim Health Care Institute, Boston, Massachusetts, USA
[5]HealthCore, Inc, Watertown, Massachusetts, USA
[6]Marshfield Clinic Research Institute, Marshfield, Wisconsin, USA
[7]Humana Healthcare Research, Louisville, Kentucky, USA
[8]HealthPartners, Bloomington, Minnesota, USA
[9]CVS Health Clinical Trial Services, Blue Bell, Pennsylvania, USA

ORCID iDs
Ting-Ying Huang http://orcid.org/0000-0002-8896-6138
Tongtong Wang http://orcid.org/0000-0001-7621-7781
Sengwee Toh http://orcid.org/0000-0002-5160-0810

REFERENCES
1 Platt R, Carnahan RM, Brown JS, *et al*. The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction. *Pharmacoepidemiol Drug Saf* 2012;21 Suppl 1:1–8.

2 Behrman RE, Benner JS, Brown JS, *et al*. Developing the Sentinel System--a national resource for evidence development. *N Engl J Med* 2011;364:498–9.

3 McGraw D, Rosati K, Evans B. A policy framework for public health uses of electronic health data. *Pharmacoepidemiol Drug Saf* 2012;21 Suppl 1:18–22.

4 Rosati K, Jorgensen N, Plc CB. Sentinel Initiative Principles and Policies: HIPAA and Common Rule Compliance in the Sentinel Initiative [Internet]. Available: https://www.sentinelinitiative.org/sites/default/files/communications/publications-presentations/HIPAA-Common-Rule-Compliance-in-Sentinel-Initiative.pdf [Accessed 29 Jun 2021].

5 Sentinel Operations Center. Sentinel Common Data Model [Internet]. Available: https://www.sentinelinitiative.org/methods-data-tools/sentinel-common-data-model/sentinel-common-data-model [Accessed 22 Jul 2021].

6 Sentinel Operations Center. Sentinel Data Quality Assurance Practices: Compliance With "Guidance for Industry and FDA Staff: Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data" [Internet]. Available: https://www.sentinelinitiative.org/sites/default/files/data/distributed-database/Sentinel_DataQAPractices_Memo.pdf [Accessed 29 Jun 2021].

7 Sentinel Operations Center. Data Quality Review and Characterization Programs [Internet]. Available: https://dev.sentinelsystem.org/projects/QA/repos/qa_package/browse [Accessed 29 Jun 2021].

8 Friberg L, Gasparini A, Carrero JJ. A scheme based on ICD-10 diagnoses and drug prescriptions to stage chronic kidney disease severity in healthcare administrative records. *Clin Kidney J* 2018;11:254–8.

9 Sentinel Operations Center.. Sentinel Routine Querying System version 9.4.0 [Internet]. Available: https://dev.sentinelsystem.org/projects/SENTINEL/repos/sentinel-routine-querying-tool-documentation/browse?at=refs%2Fheads%2F9.4.0 [Accessed 29 Jun 2021].

10 Centers for Disease Control and Prevention. National Diabetes Statistics Report 2020. Estimates of diabetes and its burden in the United States. [Internet], 2020. Available: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf [Accessed 29 Jun 2021].

11 International Diabetes Federation. IDF Diabetes Atlas 9th edition 2019 [Internet]. Available: https://www.diabetesatlas.org/en/ [Accessed 30 Jun 2021].

12 Bailey RA, Wang Y, Zhu V, *et al*. Chronic kidney disease in US adults with type 2 diabetes: an updated national estimate of prevalence based on kidney disease: improving global outcomes (KDIGO) staging. *BMC Res Notes* 2014;7:415.

13 Weng W, Tian Y, Kong SX, *et al*. The prevalence of cardiovascular disease and antidiabetes treatment characteristics among a large type 2 diabetes population in the United States. *Endocrinol Diabetes Metab* 2019;2:e00076.

14 Douros A, Lix LM, Fralick M, *et al*. Sodium-Glucose Cotransporter-2 Inhibitors and the Risk for Diabetic Ketoacidosis : A Multicenter Cohort Study. *Ann Intern Med* 2020;173:417–25.

15 Tang Y, Weiss T, Liu J, *et al*. Metformin adherence and discontinuation among patients with type 2 diabetes: a retrospective cohort study. *J Clin Transl Endocrinol* 2020;20:100225.

16 McGurnaghan S, Blackbourn LAK, Mocevic E, *et al*. Cardiovascular disease prevalence and risk factor prevalence in type 2 diabetes: a contemporary analysis. *Diabet Med* 2019;36:718–25.

17 Dunlay SM, Givertz MM, Aguilar D, *et al*. Type 2 diabetes mellitus and heart failure: a scientific statement from the American heart association and the heart failure Society of America: this statement does not represent an update of the 2017 ACC/AHA/HFSA heart failure guideline update. *Circulation* 2019;140:e294–324.

18 Iglay K, Hannachi H, Joseph Howie P, *et al*. Prevalence and co-prevalence of comorbidities among patients with type 2 diabetes mellitus. *Curr Med Res Opin* 2016;32:1243–52.

19 Pantalone KM, Hobbs TM, Wells BJ, *et al*. Changes in characteristics and treatment patterns of patients with newly diagnosed type 2 diabetes in a large United States integrated health system between 2008 and 2013. *Clin Med Insights Endocrinol Diabetes* 2016;9:CMED.s39761–30.

20 Gregg EW, Sorlie P, Paulose-Ram R, *et al*. Prevalence of lower-extremity disease in the US adult population >=40 years of age with and without diabetes: 1999-2000 national health and nutrition examination survey. *Diabetes Care* 2004;27:1591–7.

21 Mehta S, Ghosh S, Sander S, *et al*. Differences in all-cause health care utilization and costs in a type 2 diabetes mellitus population with and without a history of cardiovascular disease. *J Manag Care Spec Pharm* 2018;24:280–90.

22 Bachmann KN, Roumie CL, Wiese AD, *et al*. Diabetes medication regimens and patient clinical characteristics in the National patient-centered clinical research network, PCORnet. *Pharmacol Res Perspect* 2020;8:e00637.

23 Gagne JJ, Glynn RJ, Avorn J, *et al*. A combined comorbidity score predicted mortality in elderly patients better than existing scores. *J Clin Epidemiol* 2011;64:749–59.

24 Charlson ME, Pompei P, Ales KL, *et al*. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–83.