BMJ Open

Validating the standardized and ICD-9 code of type 2 diabetes mellitus in a common data model

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-065468
Article Type:	Original research
Date Submitted by the Author:	13-Jun-2022
Complete List of Authors:	Althunian, Turki; Saudi Food and Drug Authority, Research and Studies Alrasheed, Meshael M.; Saudi Food and Drug Authority Alnofal, Fatemah ; Saudi Food and Drug Authority, Research and Studies Tafish, Rawan ; Kingdom Hospital & Consulting Clinics, Orthopedic and Spinal Surgery Mira, Mahmood; Kingdom Hospital & Consulting Clinics, Orthopedic and Spinal Surgery Alroba, Raseel ; Saudi Food and Drug Authority, Research and Studies Kirdas, Mohammed ; Kingdom Hospital & Consulting Clinics, Orthopedic and Spinal Surgery Alshammari, Thamir; University of Hail, Department of Clinical Pharmacy; King Saud University College of Pharmacy, 2 Medication Safety Research Chair
Keywords:	Epidemiology < TROPICAL MEDICINE, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, DIABETES & ENDOCRINOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Validating the standardized and ICD-9 code of type 2 diabetes mellitus in a common data model

Turki A. Althunian, [1] Meshael M. Alrasheed, [1] Fatemah A. Alnofal, [1] Rawan T. Tafish, [2] Mahmood A. Mira, [2] Raseel A. Alroba, [1] Mohammed W. Kirdas, [2] Thamir M. Alshammari [1, 3*]

[1]Executive Directorate for Research and Studies, Saudi Food and Drug Authority, Riyadh, Saudi Arabia

[2]Kingdom Hospital and Consulting Clinics, Riyadh, Saudi Arabia

[3]College of Pharmacy, Riyadh Elm University, Riyadh, Saudi Arabia

[*]Corresponding author: thamer.alshammary@gmail.com

Word count: 1,927

Abstract

Objective: No studies have been published to assess the validity of recording diagnostic codes in general, and that of type 2 diabetes mellitus in particular, in the Saudi electronic health records (EHRs). This study was conducted to assess the validity of diagnostic codes (original and standardized) of type 2 diabetes mellitus in the centralized Saudi National Pharmacoepidemiologic Database (NPED).

Design: A retrospective validation study.

Setting: The study was conducted using the EHRs that were imported from a Saudi hospital to the NPED (a standardized common data model [CDM]).

Participants: A total of 437 random diabetic patients (≥18 years old) were extracted from the CDM, between 01 January 2013 and 01 July 2018, and matched with 437 controls.

Primary outcome: Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of comparing the standardized coding of type 2 diabetes in the CDM vs. the original electronic records at the hospital (International Statistical Classification of Diseases and Related Health Problems 9th Revision (ICD-9), vs. the paper-based medical records at the hospital, and vs. the physician re-assessment of diabetes.

Results: PPVs of NPED vs. original EHRs, paper-based records and physician re-assessment were 1.0 (95% confidence interval [CI] 0.99 to 1.0), 0.54 (95%CI 0.47 to 0.61), and 1.0 (95%CI 0.99 to 1.0); respectively. Sensitivities were 0.95 (95%CI 0.93 to 0.97), 0.93 (95%CI 0.86 to 0.97), and 0.95 (95%CI 0.93 to 0.97); respectively. NPVs were 0.95 (95%CI 0.92 to 0.97), 0.96 (0.92 to 0.98), and 0.95 (0.92 to 0.97); respectively. Specificities were 1.0 (95%CI 0.99 to 1.0), 0.68 (95%CI 0.62 to 0.73), and 1.0 (95%CI 0.99 to 1.0); respectively.

Conclusions: The results of our study substantiate the validity of coding, extracting, and standardizing type 2 diabetes mellitus in the NPED. A future multi-center study would help adding more emphasis to the study findings.

Strength and limitations of the study

- The validity of disease diagnostic codes in general, and type 2 diabetes mellitus in particular, has not been assessed in Saudi Arabia
- This study showed that the majority of the pairwise comparison validation estimates (sensitivity, specificity, positive and negative predictive values) were above 90%
- The results of our study substantiate the validity of coding, extracting, and standardizing type 2 diabetes mellitus in the NPED
- Our study was included only one center
- The validity of only ICD-9 code was assessed in the study (most hospitals have started using ICD-10 code)

INTRODUCTION

Data collected electronically from the provision of routine clinical care (i.e. real-world data [RWD]) have been used to generate evidence (Real-world evidence [RWE]) on benefits, risks, and the utilization of pharmaceuticals. [1-10] In Saudi Arabia, the electronic recording of health data in hospital settings has increased at the major tertiary hospitals during the last decade. [11-14] In 2018, the Saudi Food and Drug Authority (SFDA) established the National Pharmacoepidemiologic Database (NPED) to integrate and standardize electronic health records (EHRs) from different hospitals in Saudi Arabia.[11] The NPED was initiated to maximize the utilization of RWE in supporting drug regulatory decision-making processes.[11] The NPED will also be utilized in determining disease natural histories and trends in Saudi Arabia.[11] A standardization was performed for the EHRs that were imported from the first hospital using the Observational Health Data Sciences and Informatics (OHDSI) Common Data Model (CDM).[11] The standardization process was followed by an initial data quality assessment (no alarming concerns were identified). However, this quality assessment did not include assessing the validity of the recorded data.[11] Additionally, and up to our knowledge, no study has been published to assess the validity of the health recording practice at any of the Saudi hospitals (especially those of disease diagnostic codes).

The validity of RWD is integral in conducting pharmacoepidemiologic research studies.[8,15,16] Conducting validation studies in the Saudi health care system would assist not only in improving the quality of the generated RWE but also in supporting stakeholders in implementing their quality improvement initiatives. Validating the diagnostic codes of diabetes

(especially type 2 diabetes mellitus) is a priority given its high prevalence in Saudi Arabia (25% prevalence among the adult population), and given the lack of well-designed and large-scale pharmacoepidemiologic studies in the Saudi diabetic population.[17,18] This study was conducted to assess the validity of the original, the extracted, and the standardized diagnostic codes of type 2 diabetes mellitus of the EHRs that were imported from the first hospital to the NPED. The validity of the original diagnosis of type 2 diabetes mellitus at that hospital was also assessed. Finally, the study was aimed at assessing whether the diagnostic code of type 2 diabetes can be used to identify patients with type 2 diabetes (or diabetes as an outcome) in the standardized EHRs of that hospital.

RESEARCH DESIGN AND METHODS

Study design, data source, and patient population

This study was a retrospective validation study. The study was carried out using the EHRs that were imported and mapped from a private hospital in Riyadh to the NPED. A random sample of type 2 diabetic patients, who visited the hospital in the period between 01 January 2013 and 01 July 2018, was extracted from the standardized EHRs of the hospital and matched with a control group (non-diabetic patients) based on age and sex. The included participants were required to be ≥18 years and have at least one health record.

Validation methods

The standardized diagnostic code of diabetes in the CDM was validated using a 3-step validation approach (Figure 1). The first validation step was aimed at confirming the presence (in the included cases) and the absence (in the included controls) of type 2 diabetes in the sample that was extracted from the CDM by reviewing the patients' original EHRs at the hospital (the first reference). Diseases were coded during the study period at the hospital using the International Statistical Classification of Diseases and Related Health Problems 9th Revision (ICD-9) code (ICD-9 code of type 2 diabetes mellitus: 250.00). This validation step will help in assessing the accuracy of the CDM standardization process at the NPED, the completeness of the EHR extraction process, and the validity of the original coding ICD coding of type 2 diabetes mellitus at the hospital.

The second validation assessment was performed using a different reference: the patients' paper-based medical records (Figure 1). This validation step also included a comparison between the original EHRs (the first reference) vs. paper-based medical records (the second reference) as an additional step to validate the former standard. The final validation step, which was also a step to validate the original diagnosis of type 2 diabetes mellitus at the hospital, all study patients (both cases and controls) were re-assessed for the presence of type 2 diabetes by one of the study physicians (the third reference) based on the hospital criteria which are adopted from the

American Diabetes Association (AD) classification and diagnosis of diabetes (Table 1).[19] The physician was allowed to use all resources at the hospital that are necessary to complete the diagnosis. The diabetic code of patients in the standardized CDM was compared to the third reference. The findings from the assessment of the third reference were also compared with those of the first (original EHRs) and the second (paper-based medical records) references as an additional validation step of the latter two standards (Figure 1).

Judging whether the diagnostic code of diabetes is sufficient to identify patients with type 2 diabetes mellitus (or to identify type 2 diabetes as an outcome) in the CDM is conditional on the values of the validation assessment. If the estimates of the validation assessment (in particular sensitivity and positive predictive value [PPV]) vs. one of the references were deemed high (i.e. a chosen cut-off value of >85% based on the team clinical judgement), we would conclude that using only the diagnostic code is sufficient to identify diabetic patients in the CDM or to identify diabetes as an outcome. Otherwise, an assessment of the validity of additional identification

algorithms will be would have been performed. The following algorithms would have been examined

• FPG => 126 mg/dL (7.0 mmol/L)
• 2-h PG => 200 mg/dL (11.1 mmol/L) during an OGTT
• A1C => 6.5% (48 mmol/mol).
• Symptoms of hyperglycemia or hyperglycemic crisis (polyuria, polydipsia, and unexplained weight loss), AND a random plasma glucose => 200 mg/dL (11.1 mmol/L).
• On therapy for Diabetes mellitus (Anti-diabetic medications) and previous diagnosis of Diabetes Mellitus in medical records.

vs. the standardized code in the CDM:

- First algorithm: type 2 diabetes code + a prescription of an ant-diabetic medication
- Second algorithm: type 2 diabetes code + a prescription of an anti-diabetic medication + a blood measurement reflective of diabetes (Table 1)

Table 1. Criteria for diagnosing type 2 diabetes mellitus

* In the absence of unequivocal hyperglycemia, these criteria should be confirmed by repeat testing on a different day

These algorithms were chosen based on clinical judgment and based on other previously published algorithms.[20,21]

Statistical analysis

Sensitivity, positive predictive value (PPV), specificity, and negative predictive value (NPV) were estimated for each pairwise comparison (Figure 1) with their corresponding 95% confidence intervals (CIs). To demonstrate a sensitivity of 85% and an expected width of 95%CI of 10% (taking into account the 25% prevalence of type 2 diabetes mellitus in Saudi Arabia), a minimum sample size of 196 patients (98 cases and 98 controls) was needed.[17,18,22]

All statistical analyses were performed using RStudio Version 1.4.1103.

Patient and public involvement

Patients and/or public were not involved in the study.

RESULTS

Figure 1 shows a description of the number of patients who were included in each pairwise comparison. A total of 437 random diabetic patients were identified and were matched with 437 controls. Of these matched pairs, only 190 (43.0 %) had paper-based medical records. The median age of the included patients (both the cases and controls) was 56 years (interquartile

range=21), and 522 of the included patients (60.0 %) were male. Type 2 diabetes mellitus (among the cases) was diagnosed between 2007 and 2018. The majority of the cases (83.6%) had abnormal hemoglobin A1c levels at the time of (or the nearest time to) the diagnosis of type 2 diabetes mellitus.

The estimates of validating the standardized diabetic code in the CDM vs. two references (EHRs and the re-diagnosis) were all above 90% (Table 2). The validation estimates for EHRs vs. re-diagnosis were also above 90%. The PPV and specificity for CDM vs. paper-based documentation were 46% and 31% lower compared with those vs. EHRs and re-diagnosis (Table 3). Of 190 cases that were included in the validation assessment with the paper-based documentation as a reference, 87 (46%) did not have any records for type 2 diabetes mellitus in their medical charts. Type 2 diabetes mellitus among these 87 was mostly diagnosed after 2013 (the year of the large-scale utilization of EHRs at the hospital), and only 8 patients were diagnosed with diabetes before 2013. This may justify the absence of diabetes recording in their paper-based medical charts. The sensitivity and PPV of CDM vs. both EHRs and physician re-diagnosis were above 85%. Therefore, an assessment of additional diabetes-identification algorithms was deemed unnecessary.

Table 2. Validation estimates of the pairwise comparison among all study patients (excluding the paper-based comparison)

Comparisons	Validation estimates (%)			
	PPV (95%CI)	Sensitivity (95%CI)	NPV (95%CI)	Specificity (95%CI)
CDM vs. EHRs	100 (99 to 100)	95 (93 to 97)	95 (92 to 97)	100 (99 to 100)
CDM vs. re-diagnosis	100 (99 to 100)	95 (93 to 97)	95 (92 to 97)	100 (99 to 100)
EHRs vs. re-diagnosis	100 (99 to 100)	100 (99 to 100)	100 (99 to 100)	100 (99 to 100)

Table 3. Validation estimates of the pairwise comparison among patients with paper-based medical records (n=380)

Comparisons	Validation estimates (%)			
	PPV (95%CI)	Sensitivity (95%CI)	NPV (95%CI)	Specificity (95%CI)
CDM vs. EHRs	100 (98 to 100)	90 (85 to 93)	88 (83 to 93)	100 (98 to 100)
CDM vs. paper-based records	54 (47 to 61)	93 (86 to 97)	96 (92 to 98)	68 (62 to 73)
CDM vs. re-diagnosis	100 (98 to 100)	90 (85 to 93)	88 (83 to 93)	100 (98 to 100)
EHRs vs. paper-based records	51 (44 to 58)	98 (94 to 100)	99 (96 to 100)	62 (56 to 68)
EHRs vs. re-diagnosis	100 (97 to 100)	100 (97 to 100)	100 (98 to 100)	100 (98 to 100)
Paper-based records vs. re-diagnosis	51 (44 to 58)	98 (94 to 100)	99 (96 to 100)	62 (56 to 68)

DISCUSSION

The results of our study showed a high level of agreement in the validation estimates between the standardized diagnostic code of type 2 diabetes mellitus in the NPED (from the first hospital) and two references: the original ICD-9 coding of the EHRs and the physician re-assessment of type 2 diabetes mellitus. The results also showed a high sensitivity and NPV vs. paper-based records, but lower PPV and specificity compared with those vs. other references. Finally, the study showed that using only the standardized diagnostic code is sufficient to identify patients with type 2 diabetes mellitus or to identify type 2 diabetes mellitus as an outcome.

In our assessment of the validity of the standardized code of type diabetes mellitus in the CDM, the minimum value of sensitivity (93% vs. paper-based records) was higher than the average minimum value observed in the previous diabetes-case definition validation studies (26.9%).[20,21,23-25] On the other hand, the minimum value of specificity (62%) was lower compared with the average minimum value observed in the published studies (88%).[20,21,23-25] The minimum values of PPV and NPV were almost comparable with the average minimum values that were observed in the published studies (51% vs. 54%, and 95% vs. 90.8%; respectively).[20,21,23-25] Two of the references in our study (EHRs and re-diagnosis) were used as

references in the previous diabetes-case validation studies.[20,21,23-25] In 44.4% (8 of 18) of the published diabetes-case definition validation studies, the original EHRs were used as a reference (the other references were the physician re-diagnosis, self-reported or telephone surveys, and a multisource approach).[20,21,23,24] Type 2 diabetes mellitus was confirmed in 100% of the cases in our study, which is almost comparable to the confirmation results in a previous study in which the re-diagnosis was used as a reference.[25] Observing the high levels of sensitivity and PPV in our study meant that using only the diagnostic code of diabetes would be sufficient to identify patients with type 2 diabetes mellitus as cohorts or to identify diabetes as an outcome in the standardized EHRs that were imported from the first hospital.

Our study was the first diabetes-case definition validation study (and the first validation study for a diagnostic code) in the region. Three gold standards were used in our study, and validity was assessed at different three levels (code extraction, code standardization, and the original diagnosis of diabetes). Our study has two limitations. Firstly, the study was a single-center study. The generalizability may improve by conducting a multi-center study that takes the variability of hospital coding systems into account. Secondly, ICD-9 was used to code type 2 diabetes mellitus at the hospital during the study period; however, the hospital (and other hospitals) started upgrading their coding system to ICD-10. Including the same hospital in a future single or multi-

center with an updated sample of ICD-10 coded type2 diabetes mellitus would help in adding more emphasis to the study findings.

The results of our study substantiate the validity of coding, extracting, and standardizing type 2 diabetes mellitus in the Saudi National Pharmacoepidemiologic Database. The results also assured the process of diagnosing type 2 diabetes mellitus and the use of the standardized code to identify patients with type 2 diabetes mellitus or ascertain type 2 diabetes mellitus as an outcome in the NPED. A future multi-center that includes an updated sample from the hospital with ICD-10 coded type 2 diabetes mellitus would help in adding more emphasis to the study findings.

Ethics approval: Not applicable.

Patient and public involvement: Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

Funding: This research received no specific grant from any funding agency in the public, commercial or non-profit sectors.

Data availability statement: The datasets of this study cannot be shared without the permission of the research team and from the hospital.

Acknowledgments: We thank the information technology (IT) and the medical records (MR) departments in the Kingdom Hospital for the help in accessing the patient records and retrieving the data required for this research.

Contributors: T.A.A. and T.M.A. designed the study. M.M.A., R.T.T., and FAA contributed to the data collection process. All authors were involved were involved in designing, testing, and conducting the study. T.A.A. drafted the report. All authors critically reviewed the manuscript. T.A.A. and T.M.A. accept full responsibility for the work and conduct of the study, had full access to the data and controlled the decision to publish.

Competing interest: None declared.

Patient consent for publication: None applicable.

References:

(1) Barda N, Dagan N, Ben-Shlomo Y, et al. Safety of the BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Setting. *N Engl J Med* 2021; **385:** 1078-90.

(2) Hernán MA. Methods of Public Health Research - Strengthening Causal Inference from Observational Data. *N Engl J Med* 2021;385:1345-8.

(3) Li M, Chen S, Lai Y, et al. Integrating Real-World Evidence in the Regulatory Decision-Making Process: A Systematic Analysis of Experiences in the US, EU, and China Using a Logic Model. *Front Med (Lausanne)* 2021; **8:** 669509.

(4) Beaulieu-Jones BK, Finlayson SG, Yuan W, et al. Examining the Use of Real-World Evidence in the Regulatory Process. *Clin Pharmacol Ther* 2020; **107:** 843-52.

(5) Wu J, Wang C, Toh S, Pisa FE, Bauer L. Use of real-world evidence in regulatory decisions for rare diseases in the United States-Current status and future directions. *Pharmacoepidemiol Drug Saf* 2020; **29:** 1213-8.

(6) Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the Use of Nonrandomized Real-World Data Analyses for Regulatory Decision Making. *Clin Pharmacol Ther* 2019; **105:** 867-77.

(7) Danaei G, García Rodríguez LA, Cantero OF, Logan RW, Hernán MA. Electronic medical records can be used to emulate target trials of sustained treatment strategies. *J Clin Epidemiol* 2018; **96:** 12-22.

(8) Toh S. Pharmacoepidemiology in the era of real-world evidence. *Curr Epidemiol Rep* 2017; **4:** 262-5.

(9) Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-World Evidence - What Is It and What Can It Tell Us? *N Engl J Med* 2016; **375:** 2293-7.

(10) Dreyer NA. Making observational studies count: shaping the future of comparative effectiveness research. *Epidemiology* 2011; **22:** 295-7.

(11) Alnofal FA, Alrwisan AA, Alshammari TM. Real-world data in Saudi Arabia: Current situation and challenges for regulatory decision-making. *Pharmacoepidemiol Drug Saf* 2020; **29:** 1303-6.

(12) Samra H, Li A, Soh B, Zain MA. Utilisation of hospital information systems for medical research in Saudi Arabia: A mixed-method exploration of the views of healthcare and IT professionals involved in hospital database management systems. *Health Inf Manag* 2020; **49:** 117-26.

(13) Alsulame K, Khalifa M, Househ M. eHealth in Saudi Arabia: Current Trends, Challenges and Recommendations. *Stud Health Technol Inform* 2015; **213:** 233-6.

(14) El Mahalli A. Adoption and Barriers to Adoption of Electronic Health Records by Nurses in Three Governmental Hospitals in Eastern Province, Saudi Arabia. *Perspect Health Inf Manag* 2015; **12:** 1f.

(15) Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010; **69:** 4-14.

(16) Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. *Br J Gen Pract* 2010; **60:** e128-36.

(17) Robert AA, Al Dawish MA. The Worrying Trend of Diabetes Mellitus in Saudi Arabia: An Urgent Call to Action. *Curr Diabetes Rev* 2020; **16:** 204-10.

(18) Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9(th) edition. *Diabetes Res Clin Pract* 2019; **157:** 107843.

(19) American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2020. *Diabetes Care* 2020; **43:** S14-31.

(20) Chen G, Khan N, Walker R, Quan H. Validating ICD coding algorithms for diabetes mellitus from administrative data. *Diabetes Res Clin Pract* 2010; **89:** 189-95.

(21) Lipscombe LL, Hwee J, Webster L, Shah BR, Booth GL, Tu K. Identifying diabetes cases from administrative data: a population-based validation study. *BMC Health Serv Res* 2018; **18:** 316,018-3148-0.

(22) Buderer NM. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emerg Med* 1996; **3:** 895-900.

(23) Khokhar B, Jette N, Metcalfe A, et al. Systematic review of validated case definitions for diabetes in ICD-9-coded and ICD-10-coded data in adult populations. *BMJ Open* 2016; **6:** e009952,2015-009952.

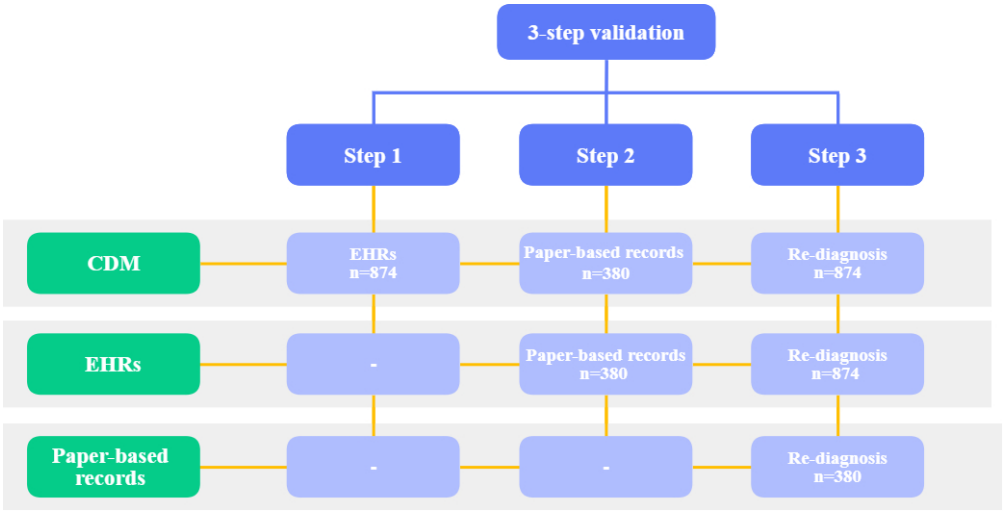
(24) Moreno-Iribas C, Sayon-Orea C, Delfrade J, et al. Validity of type 2 diabetes diagnosis in a population-based electronic health record database. *BMC Med Inform Decis Mak* 2017; **17:** 34,017-0439-z.

(25) de Burgos-Lunar C, Salinero-Fort MA, Cárdenas-Valladolid J, et al. Validation of diabetes mellitus and hypertension diagnosis in computerized medical records in primary health care. *BMC Med Res Methodol* 2011; **11:** 146,2288-11-146.

Figures

Title of Figure 1: A description of the study groups.

Legend/captions: none.



366x185mm (72 x 72 DPI)

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	2, 10, 11
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	2
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	5, 6
	4	Study objectives and hypotheses	5, 6
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	6
Participants	**6**	Eligibility criteria	6
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	6
	8	Where and when potentially eligible participants were identified (setting, location and dates)	6
	9	Whether participants formed a consecutive, random or convenience series	6
Test methods	**10a**	Index test, in sufficient detail to allow replication	6
	10b	Reference standard, in sufficient detail to allow replication	7-9
	11	Rationale for choosing the reference standard (if alternatives exist)	7-9
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	7-9
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	7-9
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	7-9
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	7-9
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	9, 10
	15	How indeterminate index test or reference standard results were handled	9, 10
	16	How missing data on the index test and reference standard were handled	9, 10
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	9, 10
	18	Intended sample size and how it was determined	
RESULTS			
Participants	**19**	Flow of participants, using a diagram	18
	20	Baseline demographic and clinical characteristics of participants	10
	21a	Distribution of severity of disease in those with the target condition	10
	21b	Distribution of alternative diagnoses in those without the target condition	10
	22	Time interval and any clinical interventions between index test and reference standard	10
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	11, 12
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	10-12
	25	Any adverse events from performing the index test or the reference standard	-
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	12-14
	27	Implications for practice, including the intended use and clinical role of the index test	12-14
OTHER INFORMATION			
	28	Registration number and name of registry	-
	29	Where the full study protocol can be accessed	-
	30	Sources of funding and other support; role of funders	15

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.

BMJ Open

Recording type 2 diabetes mellitus in a standardized central Saudi database: A retrospective validation study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-065468.R1
Article Type:	Original research
Date Submitted by the Author:	20-Oct-2022
Complete List of Authors:	Althunian, Turki; Saudi Food and Drug Authority, Research and Studies Alrasheed, Meshael M.; Saudi Food and Drug Authority Alnofal, Fatemah ; Saudi Food and Drug Authority, Research and Studies Tafish, Rawan ; Kingdom Hospital & Consulting Clinics, Orthopedic and Spinal Surgery Mira, Mahmood; Kingdom Hospital & Consulting Clinics, Orthopedic and Spinal Surgery Alroba, Raseel ; Saudi Food and Drug Authority, Research and Studies Kirdas, Mohammed ; Kingdom Hospital & Consulting Clinics, Orthopedic and Spinal Surgery Alshammari, Thamir; University of Hail, Department of Clinical Pharmacy; King Saud University College of Pharmacy, 2 Medication Safety Research Chair
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Health informatics, Health services research, Diabetes and endocrinology, Diagnostics
Keywords:	Epidemiology < TROPICAL MEDICINE, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, DIABETES & ENDOCRINOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Recording type 2 diabetes mellitus in a standardized central Saudi database: A retrospective validation study

Turki A. Althunian, [1] Meshael M. Alrasheed, [1] Fatemah A. Alnofal, [1] Rawan T. Tafish, [2] Mahmood A. Mira, [2] Raseel A. Alroba, [1] Mohammed W. Kirdas, [2] Thamir M. Alshammari [1, 3*]

[1]Executive Directorate for Research and Studies, Saudi Food and Drug Authority, Riyadh, Saudi Arabia

[2]Kingdom Hospital and Consulting Clinics, Riyadh, Saudi Arabia

[3]College of Pharmacy, Riyadh Elm University, Riyadh, Saudi Arabia

[*]Corresponding author: thamer.alshammary@gmail.com

Word count: 2,133

Abstract

Introduction: The validity of disease diagnostic codes has not been assessed in Saudi Arabia. This study was conducted to assess the validity of recording (and the original diagnostic practice) of type 2 diabetes mellitus at a hospital whose records were integrated to a centralized database (the standardized common data model [CDM] of the Saudi National Pharmacoepidemiologic Database [NPED]).

Research design and methods: A random sample of patients with type 2 diabetes mellitus (≥18 years old and with a code of type 2 diabetes mellitus) was extracted from the CDM (only the records of one tertiary care hospital were integrated at the time of the study) between 01 January 2013 and 01 July 2018 and matched with a control group (patients without diabetes) based on age and sex. The standardized coding of type 2 diabetes in the CDM was validated by comparing the presence of diabetes in the CDM vs. the original electronic records at the hospital, the recording in paper-based medical records, and the physician re-assessment of diabetes in the included cases and controls; respectively. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) values were estimated for each pairwise comparison using RStudio 1.4.1103.

Results: A total of 437 random sample of patients with type 2 diabetes mellitus was identified and matched with 437 controls. Only 190 of 437 (43.0 %) had paper-based medical records. All estimates were above 90% except for sensitivity and specificity of CDM vs. paper-based records (54%; 95% confidence interval [CI] 47 to 61%, and 68%; 95%CI 62 to 73%; respectively).

Conclusions: This study provided an assessment to the extent of which only type 2 diabetes mellitus code can be used to identify patients with this disease at Saudi centralized database. A future multi-center study would help adding more emphasis to the study findings.

Strengths and limitations

- We examined the validity of using only the code of type 2 diabetes mellitus in cohort and outcome identification vs. three reference standards in the standardized EHRs of a single hospital
- We furtherly examined the validity of two additional algorithms to identify type 2 diabetes in cohort and outcome identification in the original EHRs of the hospital
- To our knowledge, our study was the first of its kind in the region
- Our study was limited by including only one center
- Not including all type-2-diabetes-mellitus-related diagnostic codes was another limitation

INTRODUCTION

Data collected electronically from the provision of routine clinical care (i.e. real-world data [RWD]) have been used to generate evidence (Real-world evidence [RWE]) on benefits, risks, and the utilization of pharmaceuticals. [1-10] In Saudi Arabia, the electronic recording of health data in hospital settings has increased at the major tertiary hospitals during the last decade. [11-14] In 2018, the Saudi Food and Drug Authority (SFDA) established the National Pharmacoepidemiologic Database (NPED) to integrate and standardize electronic health records (EHRs) from different hospitals in Saudi Arabia.[11] The NPED was initiated to maximize the utilization of RWE in supporting drug regulatory decision-making processes.[11] The NPED will also be utilized in determining disease natural histories and trends in Saudi Arabia.[11] A standardization was performed for the EHRs that were imported from the first hospital using the Observational Health Data Sciences and Informatics (OHDSI) Common Data Model (CDM).[11] The standardization process was followed by an initial data quality assessment (no alarming concerns were identified). However, this quality assessment did not include assessing the validity of the recorded data.[11] Additionally, and up to our knowledge, no study has been published to assess the validity of the health recording practice at any of the Saudi hospitals (especially those of disease diagnostic codes).

The validity of RWD is integral in conducting pharmacoepidemiologic research studies.[8,15,16] Conducting validation studies in the Saudi health care system would assist not only in improving the quality of the generated RWE but also in supporting stakeholders in implementing their quality improvement initiatives. Validating the diagnostic codes of diabetes

(especially type 2 diabetes mellitus) is a priority given its high prevalence in Saudi Arabia (up to 25% of the Saudi population was estimated to have diabetes with an increased prevalence of 51% among the 70 to 79 year-old population), and given the lack of well-designed and large-scale pharmacoepidemiologic studies in the Saudi population with diabetes.[17-19] Studies have shown that the validity of recording diabetes mellitus in the context of RWD has been assessed in different health records using different types of data sources (e.g. physician claims, hospital discharge data, electronic health records), with different reference standards (mostly medical records, self-reported or telephone surveys), and different case definitions (e.g. using one diagnostic code or one claim for diabetes mellitus [and/or another indicator of diabetes mellitus such as high glucose levels], two or more codes/claims). [20-24]

This study was conducted to assess the validity of the original, the extracted, and the standardized diagnostic codes of type 2 diabetes mellitus of the EHRs that were imported from the first hospital to the NPED. The validity of the original diagnosis of type 2 diabetes mellitus at that hospital was also assessed. Finally, the study was aimed at assessing whether the diagnostic code of type 2 diabetes can be used to identify patients with type 2 diabetes (or diabetes as an outcome) in the standardized EHRs of that hospital.

RESEARCH DESIGN AND METHODS

Study design, data source, and patient population

This study was a retrospective single-center validation study. The study was carried out using the EHRs that were imported and mapped from a 129-bed private tertiary care hospital in Riyadh (the imported EHRs included a record of at least 500,000 patients) to the NPED. A sample of patients with type 2 diabetes mellitus (one code of type 2 diabetes mellitus), who visited the hospital in the period between 01 January 2013 and 01 July 2018, was randomly selected from the standardized EHRs of the hospital (i.e. CDM), then (using the standardized EHRs)a control group (patients without type 2 diabetes mellitus) was randomly matched based on age and sex (a control group was included for the estimation of specificities and negative predictive values). The included participants were required to be ≥18 years and have at least one health record.

Validation methods

The standardized diagnostic code of diabetes in the CDM was validated using a 3-step validation approach (Figure 1). The first validation step was aimed at confirming the presence (in the included cases) and the absence (in the included controls) of type 2 diabetes in the sample that was extracted from the CDM by reviewing the patients' original EHRs at the hospital (the first reference). Diseases were coded during the study period at the hospital using the International

Statistical Classification of Diseases and Related Health Problems 9th Revision (ICD-9) code (ICD-9 code of type 2 diabetes mellitus: 250.00). This validation step will help in assessing the accuracy of the CDM standardization process at the NPED, the completeness of the EHR extraction process, and the validity of the original coding ICD coding of type 2 diabetes mellitus at the hospital.

The second validation assessment was performed using a different reference: the patients' paper-based medical records (Figure 1). This validation step also included a comparison between the original EHRs (the first reference) vs. paper-based medical records (the second reference) as an additional step to validate the former standard. The final validation step, which was also a step to validate the original diagnosis of type 2 diabetes mellitus at the hospital, all study patients (both cases and controls) were re-assessed for the presence of type 2 diabetes by one of the study physicians (the third reference) based on the hospital criteria which are adopted from the American Diabetes Association (AD) classification and diagnosis of diabetes (Table 1).[25] The physician was allowed to use all resources at the hospital that are necessary to complete the diagnosis. The code of patients with type 2 diabetes in the standardized CDM was compared to the third reference. The findings from the assessment of the third reference were also compared with those of the first (original EHRs) and the second (paper-based medical records) references as an additional validation step of the latter two standards (Figure 1). An additional step to

confirm the diagnosis of type 2 diabetes mellitus in the original EHRs was performed using two algorithms:

- First algorithm: type 2 diabetes code + a prescription of an anti-diabetic medication
- Second algorithm: type 2 diabetes code + a prescription of an anti-diabetic medication + a blood measurement reflective of diabetes

The degree to which the results of these assessment agree with the code-only analysis will also provide more information on whether only the code can be used to identify this population in both the CDM and EHRs.

Sensitivity, specificity, the positive predictive value (PPV), and the negative predictive value (NPV) were the targeted estimates in this study. The values of these validation estimates might give an indication of the extent to which only the diagnostic code of type 2 diabetes mellitus can be used to identify type 2 diabetes mellitus as an outcome in the CDM.

Table 1. Criteria for diagnosing type 2 diabetes mellitus

• FPG => 126 mg/dL (7.0 mmol/L)
• 2-h PG => 200 mg/dL (11.1 mmol/L) during an OGTT
• Hemoglobin A1c (HbA1c) => 6.5% (48 mmol/mol).
• Symptoms of hyperglycemia or hyperglycemic crisis (polyuria, polydipsia, and unexplained weight loss), AND a random plasma glucose => 200 mg/dL (11.1 mmol/L).
• On therapy for Diabetes mellitus (Anti-diabetic medications) and previous diagnosis of Diabetes Mellitus in medical records.

* In the absence of unequivocal hyperglycemia, these criteria should be confirmed by repeat testing on a different day

These algorithms were chosen based on clinical judgment and based on other previously published algorithms.[20,21]

Statistical analysis

The validation metrics were estimated for each pairwise comparison (Figure 1) with their corresponding 95% confidence intervals (CIs). To demonstrate a sensitivity of 85% and an expected width of 95%CI of 10% (taking into account the 25% prevalence of type 2 diabetes mellitus in Saudi Arabia), a minimum sample size of 196 patients (98 cases and 98 controls) was needed.[17,18,26] The minimum total sample sizes for validating the first and second algorithms in the original EHRs were 138 and 73; respectively. All statistical analyses were performed using RStudio Version 1.4.1103.

Patient and public involvement

No patients or public involved.

RESULTS

Figure 1 shows a description of the number of patients who were included in each pairwise comparison. A total of 437 random patients with type 2 diabetes mellitus (427 [98.0%] were on anti-diabetics and/or had HbA1c measurement of >6.5%) were identified and were matched with 437 controls. Almost one- third of the cases were identified before 2016 (141 of 437 [32.3%]). Of the totally matched pairs, only 190 (43.0 %) had paper-based medical records. The median age of the included patients (both the cases and controls) was 56 years (interquartile range=21), and 522 of the included patients (60.0 %) were male. Type 2 diabetes mellitus (among the cases) was diagnosed between 2007 and 2018. The majority of the cases (83.6%) had abnormal HbA1c levels at the time of (or within 6 months) the index date (the date of diagnosing type 2 diabetes mellitus).

The estimates of validating the standardized code of type 2 diabetes mellitus in the CDM vs. two references (EHRs and the re-diagnosis) were all above 90% (Table 2). The validation estimates for EHRs vs. re-diagnosis were also above 90%. The PPV and specificity for CDM vs. paper-based documentation were 46% and 32% lower compared with those vs. EHRs and re-diagnosis (Table2). Of 190 cases that were included in the validation assessment with the paper-based documentation as a reference, 87 (46%) did not have any records for type 2 diabetes mellitus in

their medical charts. Type 2 diabetes mellitus among these 87 was mostly diagnosed after 2013 (the year of the large-scale utilization of EHRs at the hospital), and only 8 patients were diagnosed with diabetes before 2013. This may justify the absence of diabetes recording in their paper-based medical charts. The results of the validation assessment of the first and second algorithms in the EHRs were comparable with that of the type 2 diabetes mellitus code-only analysis (Table 3).

Table 2. Validation estimates of the pairwise comparisons among all study patients

Comparisons	**Validation estimates (%)**			
	PPV (95%CI)	**Sn (95%CI)**	**NPV (95%CI)**	**Sp (95%CI)**
CDM vs. EHRs (n=874)	1.00 (0.99, 1.00)	0.93 (0.90, 0.95)	0.92 (0.89, 0.95)	1.00 (0.99, 1.00)
CDM vs. paper-based records	0.54 (0.47, 0.61)	0.93 (0.86, 0.97)	0.96 (0.92, 0.98)	0.68 (0.62, 0.73)
CDM vs. re-diagnosis (n=874)	1.00 (0.99, 1.00)	0.93 (0.90, 0.95)	0.92 (0.89, 0.95)	1.00 (0.99, 1.00)
EHRs vs. re-diagnosis (n=806*)	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)

EHRs vs. paper-based records (n=336*)	0.53 (0.45, 0.61)	0.98 (0.92, 1.00)	0.99 (0.96, 1.00)	0.68 (0.62, 0.74)

*Controls in the CDM that were identified as cases in EHRs were excluded from this analysis with their corresponding cases

Table 3. Validation estimates of the assessed algorithms

Comparisons	**Validation estimates (%)**			
	PPV (95%CI)	**Sn (95%CI)**	**NPV (95%CI)**	**Sp (95%CI)**
EHRs vs. re-diagnosis (n=784) Code and anti-diabetic(s) (1st algorithm)	0.97 (0.95, 0.98)	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)	0.97 (0.95, 0.98)
EHRs vs. re-diagnosis (n=598) Code and anti-diabetic(s) and HbA1c>6.5% (2nd algorithm)	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)	1.00 (0.98, 1.00)	1.00 (0.98, 1.00)

DISCUSSION

This study assessed an approach to the population of type 2 diabetes mellitus using the disease-only code in standardized EHRs of a Saudi hospital. With the exception of the assessment vs. paper-based records, all validation estimates of the standardized and the original codes of type 2 diabetes mellitus were above 90% (the estimates of the algorithms were almost comparable with these estimates). These findings might be supportive of using only the standardized diagnostic code to identify type 2 diabetes mellitus as an outcome in these records.

In our assessment of the validity of the standardized code of type diabetes mellitus in the CDM, the minimum value of sensitivity (93% vs. paper-based records) was higher than the average minimum value observed in the previous diabetes-case definition validation studies (26.9%).[20,21,23-25] On the other hand, the minimum value of specificity (68%) was lower compared with the average minimum value observed in the published studies (88%).[20-24] The minimum values of PPV and NPV were almost comparable with the average minimum values that were observed in the published studies (54% vs. 54%, and 92% vs. 90.8%; respectively).[20-24] Two of the references in our study (EHRs and re-diagnosis) were used as references in the previous diabetes-case validation studies.[20-24] In 44.4% (8 of 18) of the published diabetes-case definition validation studies, the original EHRs were used as a reference (the other references

were the physician re-diagnosis, self-reported or telephone surveys, and a multisource approach).[20,21,22,23] Type 2 diabetes mellitus was confirmed in 100% of the cases in our study, which is almost comparable to the confirmation results in a previous study in which the re-diagnosis was used as a reference.[24] The value of validation estimates in our study might be supportive of using only the diagnostic code to identify patients with type 2 diabetes mellitus as cohorts or to identify diabetes as an outcome in the standardized EHRs that were imported from the first hospital.

Our study was the first diabetes-case definition validation study (and the first validation study for a diagnostic code) in the region. Three reference standards were used in our study, and validity was assessed at different three levels (code extraction, code standardization, and the original diagnosis of diabetes) and was compared to those of algorithms. Our study has two limitations. Firstly, the study was a single-center study. The generalizability may improve by conducting a multi-center study that takes the variability of hospital coding systems into account. Secondly, ICD-9 was used to code type 2 diabetes mellitus at the hospital during the study period; however, the hospital (and other hospitals) started upgrading their coding system to ICD-10. Additionally, we did not include other type-2-diabetes-related ICD-9 codes (codes for uncontrolled diabetes and diabetic complications) in our assessment. Including the same hospital in a future single or multi-

center with an updated sample of ICD-10 codes of type 2 diabetes mellitus and/or its complications would help in adding more emphasis to the study findings.

The validity of the (standardized) diagnostic code of type 2 diabetes mellitus at different levels and provided an indication to the extent to which this code can be used to identity this disease as an outcome. A future multi-center study that includes an updated sample from the hospital with ICD-10 codes of type 2 diabetes mellitus and/or its complications would help in adding more emphasis to the study findings.

Ethics approval: The study was approved by the SFDA ethics committee (ethics approval number: 2020_012).

Funding: This study received no specific funding.

Data availability statement: The datasets of this study cannot be shared without the permission of the research team and from the hospital.

Acknowledgments: We thank the information technology (IT) and the medical records (MR) departments in the Kingdom Hospital for the help in accessing the patient records and retrieving the data required for this research.

Contributors: T.A.A. and T.M.A. designed the study. M.M.A., R.T.T., R.A.A. and FAA contributed to the data collection, data extraction and data analysis processes. All authors were involved in designing, testing, and conducting the study. M.A.M and M.W.K. were involved in confirming the re-diagnosis processes. T.A.A. drafted the report. All authors critically reviewed the manuscript. T.A.A. and T.M.A. accept full responsibility for the work and conduct of the study, had full access to the data and controlled the decision to publish.

Competing interest: None declared.

Patient consent for publication: None applicable.

References:

(1) Barda N, Dagan N, Ben-Shlomo Y, et al. Safety of the BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Setting. *N Engl J Med* 2021; **385:** 1078-90.

(2) Hernán MA. Methods of Public Health Research - Strengthening Causal Inference from Observational Data. *N Engl J Med* 2021;385:1345-8.

(3) Li M, Chen S, Lai Y, et al. Integrating Real-World Evidence in the Regulatory Decision-Making Process: A Systematic Analysis of Experiences in the US, EU, and China Using a Logic Model. *Front Med (Lausanne)* 2021; **8:** 669509.

(4) Beaulieu-Jones BK, Finlayson SG, Yuan W, et al. Examining the Use of Real-World Evidence in the Regulatory Process. *Clin Pharmacol Ther* 2020; **107:** 843-52.

(5) Wu J, Wang C, Toh S, Pisa FE, Bauer L. Use of real-world evidence in regulatory decisions for rare diseases in the United States-Current status and future directions. *Pharmacoepidemiol Drug Saf* 2020; **29:** 1213-8.

(6) Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the Use of Nonrandomized Real-World Data Analyses for Regulatory Decision Making. *Clin Pharmacol Ther* 2019; **105:** 867-77.

(7) Danaei G, García Rodríguez LA, Cantero OF, Logan RW, Hernán MA. Electronic medical records can be used to emulate target trials of sustained treatment strategies. *J Clin Epidemiol* 2018; **96:** 12-22.

(8) Toh S. Pharmacoepidemiology in the era of real-world evidence. *Curr Epidemiol Rep* 2017; **4:** 262-5.

(9) Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-World Evidence - What Is It and What Can It Tell Us? *N Engl J Med* 2016; **375:** 2293-7.

(10) Dreyer NA. Making observational studies count: shaping the future of comparative effectiveness research. *Epidemiology* 2011; **22:** 295-7.

(11) Alnofal FA, Alrwisan AA, Alshammari TM. Real-world data in Saudi Arabia: Current situation and challenges for regulatory decision-making. *Pharmacoepidemiol Drug Saf* 2020; **29:** 1303-6.

(12) Samra H, Li A, Soh B, Zain MA. Utilisation of hospital information systems for medical research in Saudi Arabia: A mixed-method exploration of the views of healthcare and IT professionals involved in hospital database management systems. *Health Inf Manag* 2020; **49:** 117-26.

(13) Alsulame K, Khalifa M, Househ M. eHealth in Saudi Arabia: Current Trends, Challenges and Recommendations. *Stud Health Technol Inform* 2015; **213:** 233-6.

(14) El Mahalli A. Adoption and Barriers to Adoption of Electronic Health Records by Nurses in Three Governmental Hospitals in Eastern Province, Saudi Arabia. *Perspect Health Inf Manag* 2015; **12:** 1f.

(15) Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010; **69:** 4-14.

(16) Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. *Br J Gen Pract* 2010; **60:** e128-36.

(17) Robert AA, Al Dawish MA. The Worrying Trend of Diabetes Mellitus in Saudi Arabia: An Urgent Call to Action. *Curr Diabetes Rev* 2020; **16:** 204-10.

(18) Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9(th) edition. *Diabetes Res Clin Pract* 2019; **157:** 107843.

(19) World health survey Saudi Arabia: 2019 final report. Ministry of Health. Available: https://www.moh.gov.sa/en/Ministry/Statistics/Population-Health-Indicators/Documents/World-Health-Survey-Saudi-Arabia.pdf [Accessed 10 Oct 2022].

(20) Chen G, Khan N, Walker R, Quan H. Validating ICD coding algorithms for diabetes mellitus from administrative data. *Diabetes Res Clin Pract* 2010; **89:** 189-95.

(21) Lipscombe LL, Hwee J, Webster L, Shah BR, Booth GL, Tu K. Identifying diabetes cases from administrative data: a population-based validation study. *BMC Health Serv Res* 2018; **18:** 316,018-3148-0.

(22) Khokhar B, Jette N, Metcalfe A, et al. Systematic review of validated case definitions for diabetes in ICD-9-coded and ICD-10-coded data in adult populations. *BMJ Open* 2016; **6:** e009952,2015-009952.

(23) Moreno-Iribas C, Sayon-Orea C, Delfrade J, et al. Validity of type 2 diabetes diagnosis in a population-based electronic health record database. *BMC Med Inform Decis Mak* 2017; **17:** 34,017-0439-z.

(24) de Burgos-Lunar C, Salinero-Fort MA, Cárdenas-Valladolid J, et al. Validation of diabetes mellitus and hypertension diagnosis in computerized medical records in primary health care. *BMC Med Res Methodol* 2011; **11:** 146,2288-11-146.

(25) American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2020. *Diabetes Care* 2020; **43:** S14-31.

(26) Buderer NM. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emerg Med* 1996; **3:** 895-900.

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	2
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	5, 6
	4	Study objectives and hypotheses	5, 6
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	7
Participants	**6**	Eligibility criteria	7
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	7
	8	Where and when potentially eligible participants were identified (setting, location and dates)	7
	9	Whether participants formed a consecutive, random or convenience series	7
Test methods	**10a**	Index test, in sufficient detail to allow replication	7-10
	10b	Reference standard, in sufficient detail to allow replication	7-10
	11	Rationale for choosing the reference standard (if alternatives exist)	7-10
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	7-10
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	7-10
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	7-10
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	7-10
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	10
	15	How indeterminate index test or reference standard results were handled	7- 10
	16	How missing data on the index test and reference standard were handled	N/A
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	N/A
	18	Intended sample size and how it was determined	10
RESULTS			
Participants	**19**	Flow of participants, using a diagram	N/A
	20	Baseline demographic and clinical characteristics of participants	11
	21a	Distribution of severity of disease in those with the target condition	11
	21b	Distribution of alternative diagnoses in those without the target condition	11
	22	Time interval and any clinical interventions between index test and reference standard	11
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	11-13
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	11-13
	25	Any adverse events from performing the index test or the reference standard	-
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	13-15
	27	Implications for practice, including the intended use and clinical role of the index test	13-15
OTHER INFORMATION			
	28	Registration number and name of registry	-
	29	Where the full study protocol can be accessed	-
	30	Sources of funding and other support; role of funders	15

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.

BMJ Open

Recording type 2 diabetes mellitus in a standardized central Saudi database: A retrospective validation study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-065468.R2
Article Type:	Original research
Date Submitted by the Author:	07-Feb-2023
Complete List of Authors:	Althunian, Turki; Saudi Food and Drug Authority, Research and Studies Alrasheed, Meshael M.; Saudi Food and Drug Authority Alnofal, Fatemah ; Saudi Food and Drug Authority, Research and Studies Tafish, Rawan ; Kingdom Hospital & Consulting Clinics, Orthopedic and Spinal Surgery Mira, Mahmood; Kingdom Hospital & Consulting Clinics, Orthopedic and Spinal Surgery Alroba, Raseel ; Saudi Food and Drug Authority, Research and Studies Kirdas, Mohammed ; Kingdom Hospital & Consulting Clinics, Orthopedic and Spinal Surgery Alshammari, Thamir; University of Hail, Department of Clinical Pharmacy; King Saud University College of Pharmacy, 2 Medication Safety Research Chair
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Health informatics, Health services research, Diabetes and endocrinology, Diagnostics
Keywords:	Epidemiology < TROPICAL MEDICINE, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, DIABETES & ENDOCRINOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Recording type 2 diabetes mellitus in a standardized central Saudi database: A retrospective validation study

Turki A. Althunian, [1, 2] Meshael M. Alrasheed, [3] Fatemah A. Alnofal, [1] Rawan T. Tafish, [4] Mahmood A. Mira, [4] Raseel A. Alroba, [1] Mohammed W. Kirdas, [4] Thamir M. Alshammari [1, 5*]

[1]Research Informatics Department, the Saudi Food and Drug Authority, Riyadh, Saudi Arabia

[2]College of Medicine, Alfaisal University, Riyadh, Saudi Arabia

[3]Executive Directorate for Research and Studies, the Saudi Food and Drug Authority, Riyadh, Saudi Arabia

[4]Kingdom Hospital and Consulting Clinics, Riyadh, Saudi Arabia

[5]College of Pharmacy, Almaarefa University, Riyadh, Saudi Arabia

[*]Corresponding author: thamer.alshammary@gmail.com , tshammari.c@mcst.edu.sa

Word count: 2,100

Abstract

Objectives: This study was conducted to assess the validity of recording (and the original diagnostic practice) of type 2 diabetes mellitus at a hospital whose records were integrated to a centralized database (the standardized common data model [CDM] of the Saudi National Pharmacoepidemiologic Database [NPED]).

Design: A retrospective single-center validation study

Settings: Data of the study participants were extracted from the CDM of the NPED (only records of one tertiary care hospital were integrated at the time of the study) between 01 January 2013 and 01 July 2018.

Participants: A random sample of patients with type 2 diabetes mellitus (≥18 years old and with a code of type 2 diabetes mellitus) matched with a control group (patients without diabetes) based on age and sex.

Outcome measures: The standardized coding of type 2 diabetes in the CDM was validated by comparing the presence of diabetes in the CDM vs. the original electronic records at the hospital, the recording in paper-based medical records, and the physician re-assessment of diabetes in the included cases and controls; respectively. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) values were estimated for each pairwise comparison using RStudio 1.4.1103.

Results: A total of 437 random sample of patients with type 2 diabetes mellitus was identified and matched with 437 controls. Only 190 of 437 (43.0 %) had paper-based medical records. All estimates were above 90% except for sensitivity and specificity of CDM vs. paper-based records (54%; 95% confidence interval [CI] 47 to 61%, and 68%; 95%CI 62 to 73%; respectively).

Conclusions: This study provided an assessment to the extent of which only type 2 diabetes mellitus code can be used to identify patients with this disease at a Saudi centralized database. A future multi-center study would help adding more emphasis to the study findings.

Strengths and limitations

- We examined the validity of using only the code of type 2 diabetes mellitus in cohort and outcome identification vs. three reference standards in the standardized EHRs of a single hospital
- We furtherly examined the validity of two additional algorithms to identify type 2 diabetes in cohort and outcome identification in the original EHRs of the hospital
- To our knowledge, our study was the first of its kind in the region
- Our study was limited by including only one center
- Not including all type-2-diabetes-mellitus-related diagnostic codes was another limitation

INTRODUCTION

Data collected electronically from the provision of routine clinical care (i.e. real-world data [RWD]) have been used to generate evidence (Real-world evidence [RWE]) on benefits, risks, and the utilization of pharmaceuticals. [1-10] In Saudi Arabia, the electronic recording of health data in hospital settings has increased at the major tertiary hospitals during the last decade. [11-14] In 2018, the Saudi Food and Drug Authority (SFDA) established the National Pharmacoepidemiologic Database (NPED) to integrate and standardize electronic health records (EHRs) from different hospitals in Saudi Arabia.[11] The NPED was initiated to maximize the utilization of RWE in supporting drug regulatory decision-making processes.[11] The NPED will also be utilized in determining disease natural histories and trends in Saudi Arabia.[11] A standardization was performed for the EHRs that were imported from the first hospital using the Observational Health Data Sciences and Informatics (OHDSI) Common Data Model (CDM).[11] The standardization process was followed by an initial data quality assessment (no alarming concerns were identified). However, this quality assessment did not include assessing the validity of the recorded data.[11] Additionally, and up to our knowledge, no study has been published to assess the validity of the health recording practice at any of the Saudi hospitals (especially those of disease diagnostic codes).

The validity of RWD is integral in conducting pharmacoepidemiologic research studies.[8,15,16] Conducting validation studies in the Saudi health care system would assist not only in improving the quality of the generated RWE but also in supporting stakeholders in implementing their quality improvement initiatives. Validating the diagnostic codes of diabetes (especially type 2

diabetes mellitus) is a priority given its high prevalence in Saudi Arabia (up to 25% of the Saudi population was estimated to have diabetes with an increased prevalence of 51% among the 70 to 79 year-old population), and given the lack of well-designed and large-scale pharmacoepidemiologic studies in the Saudi population with diabetes.[17-19] Studies have shown that the validity of recording diabetes mellitus in the context of RWD has been assessed in different health records using different types of data sources (e.g. physician claims, hospital discharge data, electronic health records), with different reference standards (mostly medical records, self-reported or telephone surveys), and different case definitions (e.g. using one diagnostic code or one claim for diabetes mellitus [and/or another indicator of diabetes mellitus such as high glucose levels], two or more codes/claims). [20-24]

This study was conducted to assess the validity of the original, the extracted, and the standardized diagnostic codes of type 2 diabetes mellitus of the EHRs that were imported from the first hospital to the NPED. The validity of the original diagnosis of type 2 diabetes mellitus at that hospital was also assessed. Finally, the study was aimed to assess whether the diagnostic code of type 2 diabetes can be used to identify patients with type 2 diabetes (or diabetes as an outcome) in the standardized EHRs of that hospital.

METHODS

Study design, data source, and patient population

This study was a retrospective single-center validation study. The study was carried out using the EHRs that were imported and mapped from a 129-bed private tertiary care hospital in Riyadh (the imported EHRs included a record of at least 500,000 patients) to the NPED. A sample of patients with type 2 diabetes mellitus (one code of type 2 diabetes mellitus), who visited the hospital in the period between 01 January 2013 and 01 July 2018, was randomly selected from the standardized EHRs of the hospital (i.e. CDM), then (using the standardized EHRs)a control group (patients without type 2 diabetes mellitus) was randomly matched based on age and sex (a control group was included for the estimation of specificities and negative predictive values). The included participants were required to be ≥18 years and have at least one health record.

Validation methods

The standardized diagnostic code of diabetes in the CDM was validated using a 3-step validation approach. The first validation step was aimed to confirm the presence (in the included cases) and the absence (in the included controls) of type 2 diabetes in the sample that was extracted from the CDM by reviewing the patients' original EHRs at the hospital (the first reference). Diseases were coded during the study period at the hospital using the International Statistical

Classification of Diseases and Related Health Problems 9th Revision (ICD-9) code (ICD-9 code of type 2 diabetes mellitus: 250.00). This validation step will help in assessing the accuracy of the CDM standardization process at the NPED, the completeness of the EHR extraction process, and the validity of the original coding ICD coding of type 2 diabetes mellitus at the hospital.

The second validation assessment was performed using a different reference: the patients' paper-based medical records. This validation step also included a comparison between the original EHRs (the first reference) vs. paper-based medical records (the second reference) as an additional step to validate the former reference. In the final validation step, which was also a step to validate the original diagnosis of type 2 diabetes mellitus at the hospital, all study patients (both cases and controls) were re-assessed for the presence of type 2 diabetes by one of the study physicians (the third reference) based on the hospital criteria which are adopted from the American Diabetes Association (AD) classification and diagnosis of diabetes (Table 1).[25] The physician was allowed to use all resources at the hospital that are necessary to complete the diagnosis. The code of patients with type 2 diabetes in the standardized CDM was compared to the third reference. The findings from the assessment of the third reference were also compared with those of the first (original EHRs) and the second (paper-based medical records) references as an additional validation step of the latter two references. An additional step to confirm the diagnosis of type 2 diabetes mellitus in the original EHRs was performed using two algorithms:

- First algorithm: type 2 diabetes code + a prescription of an anti-diabetic medication

- Second algorithm: type 2 diabetes code + a prescription of an anti-diabetic medication + a blood measurement reflective of diabetes

These algorithms were chosen based on clinical judgment and based on other previously published algorithms.[20,21]

The degree to which the results of the algorithm assessments agree with the code-only analysis will also provide more information on whether only the code can be used to identify this population in both the CDM and EHRs. Sensitivity, specificity, the positive predictive value (PPV), and the negative predictive value (NPV) were the targeted parameters in this study. The values of these validation estimates might give an indication of the extent to which only the diagnostic code of type 2 diabetes mellitus can be used to identify type 2 diabetes mellitus as an outcome in the CDM.

• FPG => 126 mg/dL (7.0 mmol/L)
• 2-h PG => 200 mg/dL (11.1 mmol/L) during an OGTT
• Hemoglobin A1c (HbA1c) => 6.5% (48 mmol/mol).
• Symptoms of hyperglycemia or hyperglycemic crisis (polyuria, polydipsia, and unexplained weight loss), AND a random plasma glucose => 200 mg/dL (11.1 mmol/L).
• On therapy for Diabetes mellitus (Anti-diabetic medications) and previous diagnosis of Diabetes Mellitus in medical records.

Table 1. Criteria for diagnosing type 2 diabetes mellitus

* In the absence of unequivocal hyperglycemia, these criteria should be confirmed by repeat testing on a different day

Statistical analysis

The validation parameters were estimated for each pairwise comparison with their corresponding 95% confidence intervals (CIs). To demonstrate a sensitivity of 85% and an expected width of 95%CI of 10% (taking into account the 25% prevalence of type 2 diabetes mellitus in Saudi Arabia), a minimum sample size of 196 patients (98 cases and 98 controls) was needed.[17,18,26]

The minimum total sample sizes for validating the first and second algorithms in the original

EHRs were 138 and 73; respectively. All statistical analyses were performed using RStudio Version 1.4.1103.

Patient and public involvement

Patients and the public were not involved in the design, conduct or reporting/dissemination of this study.

RESULTS

Table 2 shows the number of patients who were included in each pairwise comparison. A total of 437 random patients with type 2 diabetes mellitus (427 [98.0%] were on anti-diabetics and/or had HbA1c measurement of >6.5%) were identified and were matched with 437 controls. Almost one- third of the cases were identified before 2016 (141 of 437 [32.3%]). Of the totally matched pairs, only 190 (43.0 %) had paper-based medical records. The median age of the included patients (both the cases and controls) was 56 years (interquartile range=21), and 522 of the included patients (60.0 %) were male. Type 2 diabetes mellitus (among the cases) was diagnosed between 2007 and 2018. The majority of the cases (83.6%) had abnormal HbA1c levels at the time of (or within 6 months) the index date (the date of diagnosing type 2 diabetes mellitus).

The estimates of validating the standardized code of type 2 diabetes mellitus in the CDM vs. two references (EHRs and the re-diagnosis) were all above 90% (Table 2). The validation estimates for EHRs vs. re-diagnosis were also above 90%. The PPV and specificity for CDM vs. paper-based documentation were 46% and 32% lower compared with those vs. EHRs and re-diagnosis (Table 2). Of 190 cases that were included in the validation assessment with the paper-based documentation as a reference, 87 (46%) did not have any records for type 2 diabetes mellitus in their medical charts. Type 2 diabetes mellitus among these 87 was mostly diagnosed after 2013 (the year of the large-scale utilization of EHRs at the hospital), and only 8 patients were diagnosed with diabetes before 2013. This may justify the absence of diabetes recording in their paper-based medical charts. The results of the validation assessment of the first and second algorithms in the EHRs were comparable with that of the type 2 diabetes mellitus code-only analysis (Table 3).

Table 2. Validation estimates of the pairwise comparisons among all study patients

Comparisons	**Validation estimates (%)**			
	PPV (95%CI)	**Sn (95%CI)**	**NPV (95%CI)**	**Sp (95%CI)**
CDM vs. EHRs (n=874)	1.00 (0.99, 1.00)	0.93 (0.90, 0.95)	0.92 (0.89, 0.95)	1.00 (0.99, 1.00)

CDM vs. paper-based records	0.54 (0.47, 0.61)	0.93 (0.86, 0.97)	0.96 (0.92, 0.98)	0.68 (0.62, 0.73)
CDM vs. re-diagnosis (n=874)	1.00 (0.99, 1.00)	0.93 (0.90, 0.95)	0.92 (0.89, 0.95)	1.00 (0.99, 1.00)
EHRs vs. re-diagnosis (n=806*)	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)
EHRs vs. paper-based records (n=336*)	0.53 (0.45, 0.61)	0.98 (0.92, 1.00)	0.99 (0.96, 1.00)	0.68 (0.62, 0.74)

*Controls in the CDM that were identified as cases in EHRs were excluded from this analysis with their corresponding cases

Table 3. Validation estimates of the assessed algorithms

Comparisons	**Validation estimates (%)**			
	PPV (95%CI)	**Sn (95%CI)**	**NPV (95%CI)**	**Sp (95%CI)**
EHRs vs. re-diagnosis (n=784) Code and anti-diabetic(s) (1st algorithm)	0.97 (0.95, 0.98)	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)	0.97 (0.95, 0.98)
EHRs vs. re-diagnosis (n=598)	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)	1.00 (0.98, 1.00)	1.00 (0.98, 1.00)

Code and anti-diabetic(s) and HbA1c>6.5% (2nd algorithm)				

DISCUSSION

This study assessed an approach to the population of type 2 diabetes mellitus using the disease-only code in standardized EHRs of a Saudi hospital. With the exception of the assessment vs. paper-based records, all validation estimates of the standardized and the original codes of type 2 diabetes mellitus were above 90% (the estimates of the algorithms were almost comparable with these estimates). These findings might be supportive of using only the standardized diagnostic code to identify type 2 diabetes mellitus as an outcome in these records.

In our assessment of the validity of the standardized code of type diabetes mellitus in the CDM, the minimum value of sensitivity (93% vs. paper-based records) was higher than the average minimum value observed in the previous diabetes-case definition validation studies (26.9%).[20,21,23-25] On the other hand, the minimum value of specificity (68%) was lower compared with the average minimum value observed in the published studies (88%).[20-24] The minimum values of PPV and NPV were almost comparable with the average minimum values

that were observed in the published studies (54% vs. 54%, and 92% vs. 90.8%; respectively).[20-24] Two of the references in our study (EHRs and re-diagnosis) were used as references in the previous diabetes-case validation studies.[20-24] In 44.4% (8 of 18) of the published diabetes-case definition validation studies, the original EHRs were used as a reference (the other references were the physician re-diagnosis, self-reported or telephone surveys, and a multisource approach).[20,21,22,23] Type 2 diabetes mellitus was confirmed in 100% of the cases in our study, which is almost comparable to the confirmation results in a previous study in which the re-diagnosis was used as a reference.[24] The value of validation estimates in our study might be supportive of using only the diagnostic code to identify patients with type 2 diabetes mellitus as cohorts or to identify diabetes as an outcome in the standardized EHRs that were imported from the first hospital.

Our study was the first diabetes-case definition validation study (and the first validation study for a diagnostic code) in the region. Three reference standards were used in our study, and validity was assessed at different three levels (code extraction, code standardization, and the original diagnosis of diabetes) and was compared to those of algorithms. Our study has two limitations. Firstly, the study was a single-center study. The generalizability may improve by conducting a multi-center study that takes the variability of hospital coding systems into account. Secondly,

ICD-9 was used to code type 2 diabetes mellitus at the hospital during the study period; however, the hospital (and other hospitals) started upgrading their coding system to ICD-10. Additionally, we did not include other type-2-diabetes-related ICD-9 codes (codes for uncontrolled diabetes and diabetic complications) in our assessment. Including the same hospital in a future single or multi-center with an updated sample of ICD-10 codes of type 2 diabetes mellitus and/or its complications would help in adding more emphasis to the study findings.

We assessed the validity of the (standardized) diagnostic code of type 2 diabetes mellitus at different recording levels and provided an indication to the extent to which this code can be used to identity this disease as an outcome. A future multi-center study that includes an updated sample from the hospital with ICD-10 codes of type 2 diabetes mellitus and/or its complications would help in adding more emphasis to the study findings.

Ethics approval: The study was approved by the SFDA ethics committee (ethics approval number: 2020_012).

Funding: This study received no specific funding.

Data availability statement: The datasets of this study cannot be shared without the permission of the research team and from the hospital.

Disclaimer: The views expressed in this paper are those of the author(s) and do not necessarily reflect those of the SFDA or its stakeholders. Guaranteeing the accuracy and the validity of the data is a sole responsibility of the research team

Acknowledgments: We thank the information technology (IT) and the medical records (MR) departments in the Kingdom Hospital for the help in accessing the patient records and retrieving the data required for this research.

Contributors: T.A.A., M.W.K., and T.M.A. designed the study. M.M.A., R.T.T., R.A.A., and F.A.A. contributed to the data collection process. All authors were involved in designing, analyzing, and conducting the study. M.A.M. was involved in confirming the diagnosis. T.A.A. drafted the report. All authors critically reviewed the manuscript. T.A.A. and T.M.A. accept full

responsibility for the work and conduct of the study, had full access to the data and controlled the decision to publish.

Competing interest: None declared.

Patient consent for publication: None applicable.

References:

(1) Barda N, Dagan N, Ben-Shlomo Y, et al. Safety of the BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Setting. *N Engl J Med* 2021; **385:** 1078-90.

(2) Hernán MA. Methods of Public Health Research - Strengthening Causal Inference from Observational Data. *N Engl J Med* 2021;385:1345-8.

(3) Li M, Chen S, Lai Y, et al. Integrating Real-World Evidence in the Regulatory Decision-Making Process: A Systematic Analysis of Experiences in the US, EU, and China Using a Logic Model. *Front Med (Lausanne)* 2021; **8:** 669509.

(4) Beaulieu-Jones BK, Finlayson SG, Yuan W, et al. Examining the Use of Real-World Evidence in the Regulatory Process. *Clin Pharmacol Ther* 2020; **107:** 843-52.

(5) Wu J, Wang C, Toh S, Pisa FE, Bauer L. Use of real-world evidence in regulatory decisions for rare diseases in the United States-Current status and future directions. *Pharmacoepidemiol Drug Saf* 2020; **29:** 1213-8.

(6) Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the Use of Nonrandomized Real-World Data Analyses for Regulatory Decision Making. *Clin Pharmacol Ther* 2019; **105:** 867-77.

(7) Danaei G, García Rodríguez LA, Cantero OF, Logan RW, Hernán MA. Electronic medical records can be used to emulate target trials of sustained treatment strategies. *J Clin Epidemiol* 2018; **96:** 12-22.

(8) Toh S. Pharmacoepidemiology in the era of real-world evidence. *Curr Epidemiol Rep* 2017; **4:** 262-5.

(9) Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-World Evidence - What Is It and What Can It Tell Us? *N Engl J Med* 2016; **375:** 2293-7.

(10) Dreyer NA. Making observational studies count: shaping the future of comparative effectiveness research. *Epidemiology* 2011; **22:** 295-7.

(11) Alnofal FA, Alrwisan AA, Alshammari TM. Real-world data in Saudi Arabia: Current situation and challenges for regulatory decision-making. *Pharmacoepidemiol Drug Saf* 2020; **29:** 1303-6.

(12) Samra H, Li A, Soh B, Zain MA. Utilisation of hospital information systems for medical research in Saudi Arabia: A mixed-method exploration of the views of healthcare and IT

professionals involved in hospital database management systems. *Health Inf Manag* 2020; **49:** 117-26.

(13) Alsulame K, Khalifa M, Househ M. eHealth in Saudi Arabia: Current Trends, Challenges and Recommendations. *Stud Health Technol Inform* 2015; **213:** 233-6.

(14) El Mahalli A. Adoption and Barriers to Adoption of Electronic Health Records by Nurses in Three Governmental Hospitals in Eastern Province, Saudi Arabia. *Perspect Health Inf Manag* 2015; **12:** 1f.

(15) Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010; **69:** 4-14.

(16) Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. *Br J Gen Pract* 2010; **60:** e128-36.

(17) Robert AA, Al Dawish MA. The Worrying Trend of Diabetes Mellitus in Saudi Arabia: An Urgent Call to Action. *Curr Diabetes Rev* 2020; **16:** 204-10.

(18) Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9(th) edition. *Diabetes Res Clin Pract* 2019; **157:** 107843.

(19) World health survey Saudi Arabia: 2019 final report. Ministry of Health. Available: https://www.moh.gov.sa/en/Ministry/Statistics/Population-Health-Indicators/Documents/World-Health-Survey-Saudi-Arabia.pdf [Accessed 10 Oct 2022].

(20) Chen G, Khan N, Walker R, Quan H. Validating ICD coding algorithms for diabetes mellitus from administrative data. *Diabetes Res Clin Pract* 2010; **89:** 189-95.

(21) Lipscombe LL, Hwee J, Webster L, Shah BR, Booth GL, Tu K. Identifying diabetes cases from administrative data: a population-based validation study. *BMC Health Serv Res* 2018; **18:** 316,018-3148-0.

(22) Khokhar B, Jette N, Metcalfe A, et al. Systematic review of validated case definitions for diabetes in ICD-9-coded and ICD-10-coded data in adult populations. *BMJ Open* 2016; **6:** e009952,2015-009952.

(23) Moreno-Iribas C, Sayon-Orea C, Delfrade J, et al. Validity of type 2 diabetes diagnosis in a population-based electronic health record database. *BMC Med Inform Decis Mak* 2017; **17:** 34,017-0439-z.

(24) de Burgos-Lunar C, Salinero-Fort MA, Cárdenas-Valladolid J, et al. Validation of diabetes mellitus and hypertension diagnosis in computerized medical records in primary health care. *BMC Med Res Methodol* 2011; **11:** 146,2288-11-146.

(25) American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2020. *Diabetes Care* 2020; **43:** S14-31.

(26) Buderer NM. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emerg Med* 1996; **3:** 895-900.

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	2
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	5, 6
	4	Study objectives and hypotheses	5, 6
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	7
Participants	**6**	Eligibility criteria	7
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	7
	8	Where and when potentially eligible participants were identified (setting, location and dates)	7
	9	Whether participants formed a consecutive, random or convenience series	7
Test methods	**10a**	Index test, in sufficient detail to allow replication	7-10
	10b	Reference standard, in sufficient detail to allow replication	7-10
	11	Rationale for choosing the reference standard (if alternatives exist)	7-10
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	7-10
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	7-10
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	7-10
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	7-10
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	10
	15	How indeterminate index test or reference standard results were handled	7- 10
	16	How missing data on the index test and reference standard were handled	N/A
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	N/A
	18	Intended sample size and how it was determined	10
RESULTS			
Participants	**19**	Flow of participants, using a diagram	N/A
	20	Baseline demographic and clinical characteristics of participants	11
	21a	Distribution of severity of disease in those with the target condition	11
	21b	Distribution of alternative diagnoses in those without the target condition	11
	22	Time interval and any clinical interventions between index test and reference standard	11
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	11-13
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	11-13
	25	Any adverse events from performing the index test or the reference standard	-
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	13-15
	27	Implications for practice, including the intended use and clinical role of the index test	13-15
OTHER INFORMATION			
	28	Registration number and name of registry	-
	29	Where the full study protocol can be accessed	-
	30	Sources of funding and other support; role of funders	15

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.