

---

## Research and Applications

# Development of an algorithm to link electronic health record prescriptions with pharmacy dispense claims

Megan Hoopes,<sup>1</sup> Heather Angier,<sup>2</sup> Lewis A Raynor,<sup>1</sup> Andrew Suchocki,<sup>2</sup>  
John Muench,<sup>2</sup> Miguel Marino,<sup>2,3</sup> Pedro Rivera,<sup>1</sup> and Nathalie Huguet<sup>2</sup>

<sup>1</sup>OCHIN, Inc., Portland, Oregon, USA, <sup>2</sup>Department of Family Medicine, Oregon Health & Science University, Portland, Oregon, USA and <sup>3</sup>School of Public Health, Oregon Health & Science University – Portland State University, Portland, Oregon, USA

Corresponding Author: Megan Hoopes, MPH, OCHIN, Inc. 1881 SW Naito Pkwy, Portland, OR 97201, USA (hoopesm@ochin.org)

Received 26 February 2018; Revised 21 June 2018; Editorial Decision 23 June 2018; Accepted 27 June 2018

### ABSTRACT

**Objective:** Medication adherence is an important aspect of chronic disease management. Electronic health record (EHR) data are often not linked to dispensing data, limiting clinicians' understanding of which of their patients fill their medications, and how to tailor care appropriately. We aimed to develop an algorithm to link EHR prescribing to claims-based dispensing data and use the results to quantify how often patients with diabetes filled prescribed chronic disease medications.

**Materials and Methods:** We developed an algorithm linking EHR prescribing data (RxNorm terminology) to claims-based dispensing data (NDC terminology), within sample of adult (19-64) community health center (CHC) patients with diabetes from a network of CHCs across 12 states. We demonstrate an application of the method by calculating dispense rates for a set of commonly prescribed diabetes and cardio-protective medications. To further inform clinical care, we computed adjusted odds ratios of dispense by patient-, encounter-, and clinic-level characteristics.

**Results:** Seventy-six percent of cardio-protective medication prescriptions and 74% of diabetes medications were linked to a dispensing record. Age, income, ethnicity, insurance, assigned primary care provider, comorbidity, time on EHR, and clinic size were significantly associated with odds of dispensing.

**Discussion:** EHR prescriptions and pharmacy dispense data can be linked at the record level across different terminologies. Dispensing rates in this low-income population with diabetes were similar to other populations.

**Conclusion:** Record linkage resulted in the finding that CHC patients with diabetes largely had their chronic disease medications dispensed. Understanding factors associated with dispensing rates highlight barriers and opportunities for optimal disease management.

**Key words:** linkage algorithm, medication adherence, electronic health records, pharmacy claims, diabetes

---

### BACKGROUND

Implementation of electronic health records (EHRs) and e-prescribing has dramatically increased in recent years,<sup>1</sup> allowing improved data capture of prescribed medications. EHRs generally record medication prescribing but are often unequipped to capture whether those prescriptions are filled. External dispensing data (eg, claims,

pharmacy dispensing databases) can be retroactively linked to prescribing data; however, prescribed and dispensed medications are often stored using different terminology systems.<sup>2</sup> In addition, there is no gold standard for how to assess medication adherence for research,<sup>3</sup> and limited evidence about the validity of using prescribing data as a proxy for dispensed medications.<sup>4-6</sup>

Yet, the extent to which patients receive and take medications prescribed by their healthcare providers is an important component of chronic disease management. Failure to adhere to medication protocols leads to increased patient morbidity and mortality coupled with higher medical costs and increased utilization of medical resources.<sup>7</sup> Yet, studies have shown about 25% of prescriptions go unfilled.<sup>6,8,9</sup>

Individuals receiving care at community health centers (CHCs), our nation's healthcare "safety net," are largely low-income and either publicly insured (ie, Medicaid, Medicare) or uninsured. Non-elderly adult Medicaid recipients are sicker and have a greater burden of chronic disease than the general population;<sup>10</sup> consequently, patients seen in CHCs likely face many barriers to medication adherence. Identifying barriers to medication use for CHCs serving economically marginalized and sicker patients creates opportunities for interventions that can improve patient health and reduce healthcare costs for CHCs. A primary challenge to measuring this is the lack of integrated data on prescribing and dispensing.

## Objective

This study aimed to develop an algorithm linking EHR prescribing data from a large national network of CHCs to claims-based dispensing data at the record level. To demonstrate an application of the approach and to contribute to the literature on chronic disease medication adherence, we then computed dispensing rates for patients with diabetes by medication class and conducted an exploratory analysis of patient-, encounter-, and facility-level factors associated with medication dispensing. We considered a list of diabetes-related and cardiovascular disease (CVD) medications commonly covered by Medicaid. We focused on commonly prescribed drugs for diabetes management and CVD prevention because treatment recommendations include both, as patients with diabetes are at higher risk for CVD than those without diabetes.<sup>11-13</sup> Further, patients with diabetes often show poor adherence to prescribed medication regimens.<sup>14,15</sup>

## METHODS

### Data sources

The National Patient-Centered Clinical Research Network (PCORnet) created 13 clinical data research networks (CDRNs) based on a common data model (CDM) to facilitate community-based research that is generalizable to many populations.<sup>16</sup> The Accelerating Data Value Across a National Community Health Center Network (ADVANCE) CDRN is a multi-center collaborative led by the OCHIN (not an acronym) community health information network.<sup>17</sup> For the current study, we utilized the PRESCRIBING and DISPENSING tables, standardized in the PCORnet common data model version 3.1,<sup>18</sup> from the ADVANCE data warehouse. The PRESCRIBING table is populated with source data from member clinics' EHRs, and data in the DISPENSING table are obtained from an external pharmacy claims vendor (Surescripts), which is integrated into the clinics' EHRs. For each appointment scheduled on the next day, the OCHIN EHR automatically queries the Surescripts database, which returns a list of all medications that have been dispensed for that patient in the past 12 months from pharmacies in the Surescripts network. The provider can view that list and perform medication reconciliation with data entered in the EHR. The PRESCRIBING and DISPENSING tables contain a patient identification variable but are not explicitly linked (ie, the ADVANCE CDM does not

contain a variable that directly links a specific prescribing record and one or more dispensing records).

## Medication terminologies

### Prescribing data

ADVANCE prescribing data are standardized using RxNorm, which is an open-source program created by the National Library of Medicine to provide a single system for unambiguously identifying brand-name and generic drugs.<sup>19</sup> RxNorm provides a normalized name to each drug in addition to a concept unique identifier (RxCUI) that makes it possible to clearly identify a given drug; drugs that map to the same RxCUI are the same drug (identical in ingredients, strengths, and dose forms).<sup>19</sup> This terminology allows for medications to be exchanged across EHRs, making the usage of RxNorm a criterion for Meaningful Use Stage 2 EHR certification.<sup>20</sup> The PCORnet CDM specifications call for each prescription to be mapped to RxCUI at the highest possible specificity in order to keep one record per prescription.

### Dispensing data

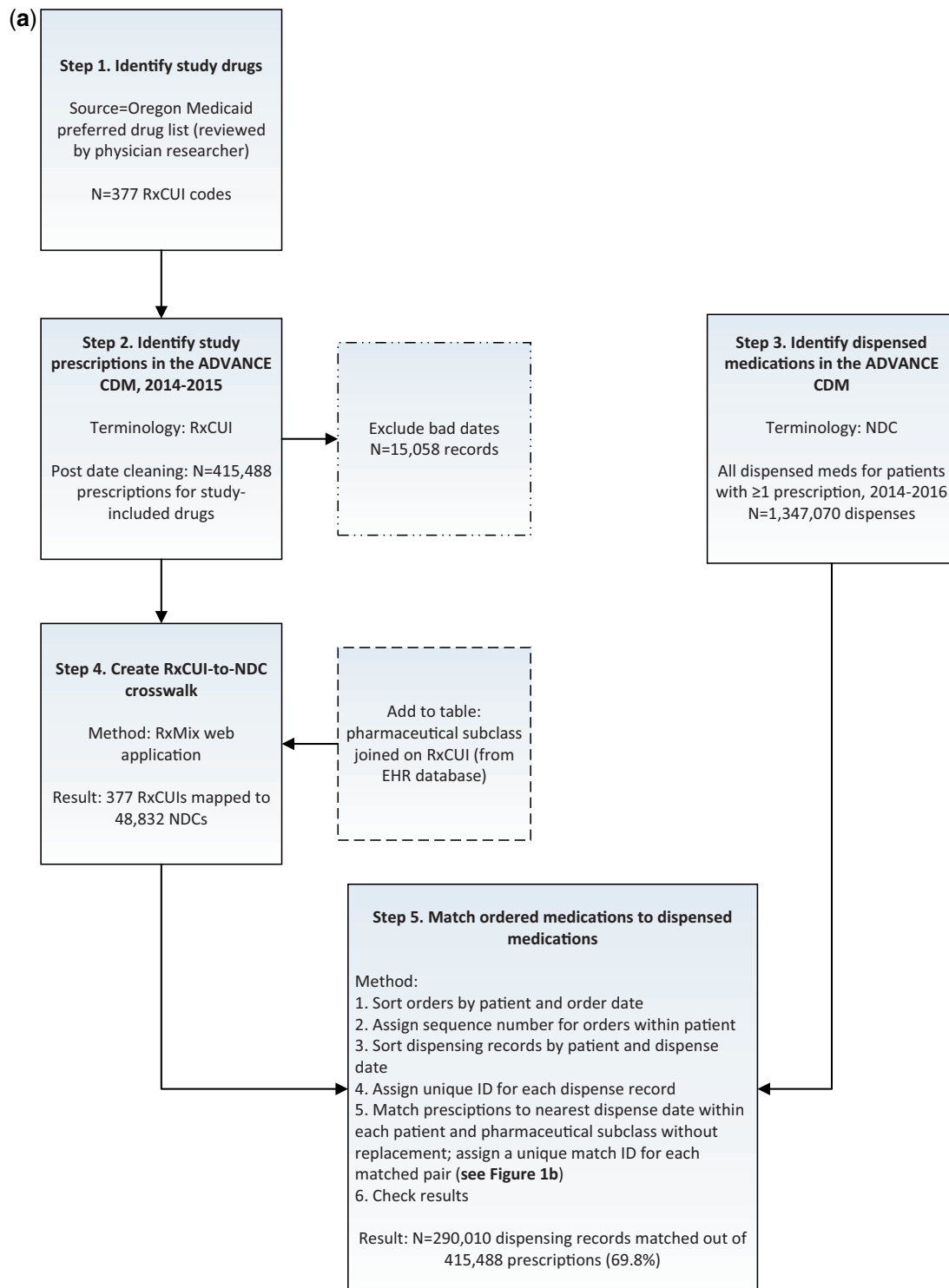
ADVANCE dispensing data are stored in National Drug Code (NDC) nomenclature, a universal product identifier published by the Food and Drug Administration.<sup>21</sup> Each NDC is a unique three-segment number that identifies the labeler (ie, the manufacturer or distributor), product (ie, strength, dosage form, and formulation), and package (ie, package sizes and types). Dispensing records are based on insurance claims; thus, dispensing information is available only for patients who are insured, publically or privately, at the time of dispense.

## Study period and population

Our algorithm was developed within a population of adult patients with diabetes seen over a two-year study period (2014-2015). We included ADVANCE patients from OCHIN clinics (the subset for which dispensing data were available), aged 19-64 with diabetes<sup>22</sup> as of the start of the study period and  $\geq 1$  prescription for an included medication during the study period. Patients who had a pregnancy in the study period, no ambulatory visits, those uninsured throughout the entire study period, and patients with an unknown sex were excluded. Patients with unstable health insurance coverage (eg, had a combination of insured and uninsured visits) were retained in the sample. After exclusions, the study population included 24 130 eligible patients with DM and  $\geq 1$  study medication prescribed; study patients came from 273 CHCs in 50 health systems across 12 states.

## Explanatory variables

We obtained patient-, encounter-, and clinic-level variables from the ADVANCE CDM to assess a range of factors associated with dispensing. Patient-level variables included demographics [sex, age, race/ethnicity, household income as percent of federal poverty level (FPL), urban/rural residence], insurance coverage status and patterns (Medicaid, Medicare, private, multiple types of coverage, partially insured, newly insured in 2014-2015 after being uninsured in 2013), primary care provider (PCP) assignment, diagnoses, and comorbidity burden. The Charlson comorbidity index was used as a measure of clinical complexity.<sup>23</sup> Encounter-level factors included reason for visit, provider type, visit with assigned PCP, and whether medications prescribed at visit were e-prescribed and primary vs. refills. At the clinic level, we collected information about clinic type and



**Figure 1.** Process used to identify and match prescribed and dispensed medications. Notes: ADVANCE CDM=Common Data Model from the ADVANCE Clinical Data Research Network. RxCUI=RxNorm concept unique identifier. NDC=National Drug Code. RxMix=web interface from US National Library of Medicine, used to create mappings among different drug terminologies. Figure 1b. Representation of adjudication loop for matching medications. Description: Sort medication prescriptions and dispensing records by date within each patient. N=distinct patients with  $\geq 1$  prescription, M=maximum number of prescription records for any given patient, O=distinct patients with  $\geq 1$  dispensing record, and P=maximum number of dispensing records for any given patient.

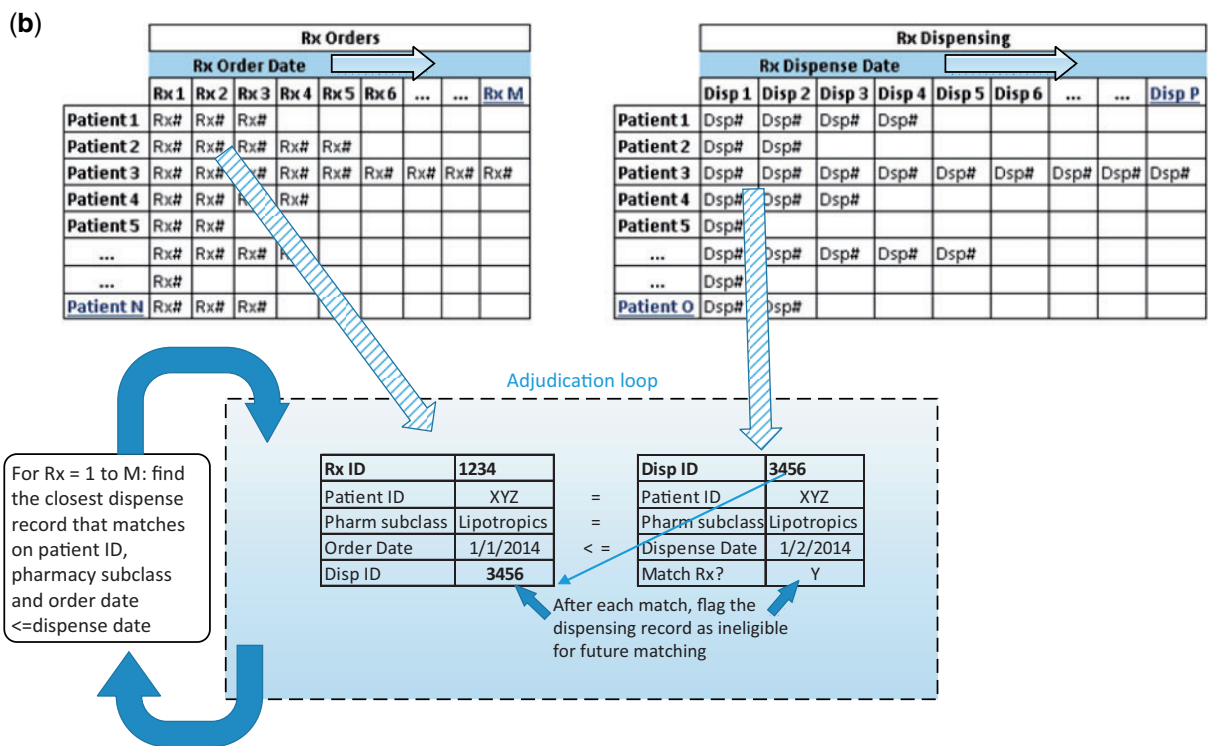


Figure 1. Continued.

specialty (primary care, other), length of time on the EHR, clinic size (number of active patients), and whether the clinic has attested for Meaningful Use.

### Medication matching

We developed a 5-step process to identify and match prescribed and dispensed medications (see Figure 1a and b).

#### Step 1. Identify study drugs

We identified diabetes and CVD medications using the publicly available Oregon Medicaid preferred drug list.<sup>24</sup> This list identifies the system, class, and preferred drugs covered by Medicaid in Oregon, with associated RxNorm codes. We confirmed the appropriateness and generalizability of this list with CHC clinician-researchers.

#### Step 2. Identify study medication prescriptions in the ADVANCE CDM

We identified all diabetes and CVD medication prescriptions in the ADVANCE CDM using the list of RxNorm codes from Step 1. Prescriptions were excluded if they had: (a) a start date equal to the end date; (b) an end date prior to an order date; (c) an end date prior to a start date; (d) a start date more than two weeks before an order date; and/or (e) an end date more than one year after the end of our study period.

#### Step 3. Identify dispensed medications in the ADVANCE CDM

We identified all dispensing records (not limited to diabetes and CVD drugs) from the CDM for patients with one or more of our study medications prescribed and a dispense date between January 1, 2014, and December 31, 2016. We extended the date range for dispensed medications to one year after the study period to allow for

potential lags between prescriptions and dispenses, both because patients may not immediately fill prescriptions, and because dispensed claims will not be linked to a patient’s chart until his/her next clinic visit. The ADVANCE CDM DISPENSING table includes the drug’s NDC and dispense date but no further detail on drug name or class.

#### Step 4. Create a crosswalk between RxCUI (prescribing) and NDC (dispensing) terminologies

RxMix is web application that allows users to map between different terminologies of prescription drugs.<sup>2</sup> We used RxMix to identify all current and historical NDCs associated with each of the study RxCUIs (batch processing method through the user interface: input=RxCUI, function=getallhistoricalNDCs). The resulting crosswalk included 377 RxCUIs mapped to 48 832 NDCs. We then extracted the pharmaceutical subclass (see Table 2 for list) for each RxCUI, as this field was used in the algorithm to match individual prescriptions to dispense records.

#### Step 5. Algorithm to match prescribing to dispensing records

We sorted prescription records by patient and order date and assigned a sequence number of prescription records for each patient. Next, we sorted dispensing records by patient and dispensing date and created a unique identification (ID) number for matching purposes. We wrote a Structured Query Language (SQL) script to match prescriptions to the closest dispensing ID within patient ID and pharmaceutical subclass without replacement, ie, dispensing records were ineligible for future matches once matched to a prescription (see Figure 1b; the SQL script for this step is included in Supplementary Materials). Finally, we added the unique dispensing

ID to the prescribed medication table where we paired distinct prescriptions with dispenses.

### Analytic application of linked medication results

After matching prescribed to dispensed medications, we constructed an analytic dataset to demonstrate an application of the method and to inform clinical care. Analysis was limited to health centers with  $\geq 50$  study prescription records to aid in the precision of health center-level estimates. After applying this and other patient-level exclusion criteria described above, our final sample included 363 693 diabetes and CVD prescriptions matched to 273 981 dispensing records for 24 130 patients with diabetes across 273 CHCs nested within 50 health centers.

We described the study sample demographics and calculated medication dispensing rates and distributions across health centers (median, first and third quartiles) by drug class. We then conducted an exploratory analysis to estimate adjusted odds of dispense for a range of patient-, encounter-, and clinic-level explanatory variables. Adjusted odds ratios were computed using generalized estimating equation models with a logit link, applying a robust sandwich estimator to cluster standard errors for repeated measures within patients nested within health centers (SAS PROC GENMOD). All models were adjusted for sex, age group, and race/ethnicity. Statistical significance was two-sided and set at  $\alpha = 0.05$ . All analyses were conducted using SAS version 9.4 (SAS Institute, Inc.). The study was conducted with Institutional Review Board approval.

## RESULTS

We identified 430 546 diabetes and CVD prescriptions within our study population in the ADVANCE CDM, of which 3.5% of prescription records were excluded due to the date cleaning parameters described above. The final prescription dataset contained 415 488 records. We then identified 1 347 070 total dispensing records for these patients. The matching algorithm resulted in 290 010 of 415 488 prescriptions (69.8%) being matched to a dispensing record.

Match results were verified by comparing matched brand and generic medication names, dosage, and units as recorded in the EHR database. Over 98% of matched pairs were an exact match on brand name or generic name; the remaining 2% matched within pharmaceutical subclass only. Ninety-six percent of matches were for the same dose and units, with the remaining differing slightly on dose or units between the prescribed and dispensed records. Fifty-nine percent of matches were dispensed within seven days of the prescribing date; the median number of days dispensed after prescription was 1.0.

After study exclusions, the 24 130 patients with DM received 363 693 DM and CVD prescriptions in the two-year study period (Table 1). This represented 92% of potential study patients with DM; the remaining 8% of DM patients did not have a study medication prescribed. Study patients were predominantly female (55%), most were 40 to 55 years of age (47%), low-income ( $\leq 138\%$  of FPL, 74%), non-Hispanic white (49%), and residents of urban areas (79%; Table 1). The population was largely insured by Medicaid (40% continuously Medicaid covered), but notably 26% were partially uninsured. Most patients had  $\geq 10$  ambulatory visits (44%), and comorbidities were common (hypertension on problem list: 69%; lipid disorder on problem list: 70%; Charlson comorbidity index  $\geq 5$ : 45%).

**Table 1.** Demographic characteristics of community health center patients with diabetes (N = 24 130)

Cardiovascular drugs prescribed (N)	203 504
Diabetes drugs prescribed (N)	160 189
Sex, N (%)	
Female	13 209 (54.7)
Male	10 921 (45.3)
Age group, N (%)	
20-39	3494 (14.5)
40-55	11 307 (46.9)
56-64	9329 (38.7)
Race/ethnicity, N (%)	
Hispanic	6192 (25.7)
Non-Hispanic white	11 707 (48.5)
Non-Hispanic black	4765 (19.7)
Non-Hispanic other	1066 (4.4)
Missing/unknown	400 (1.7)
% of federal poverty level (FPL), N (%)	
$\leq 138\%$	17 829 (73.9)
$> 138\%$	3668 (15.2)
Missing/unknown	2633 (10.9)
Urbanicity, N (%)	
Urban	18 943 (78.5)
Rural	2728 (11.3)
Missing/unknown	2459 (10.2)
State of residence, N (%)	
Medicaid expansion state	20 325 (84.2)
Medicaid non-expansion state	3805 (15.8)
Insurance status, N (%)	
Medicaid only	9560 (39.6)
Medicare only	3650 (15.1)
Private only	2877 (11.9)
Multiple types	1731 (7.2)
Partially uninsured	6312 (26.2)
Newly insured in 2014, N (%)	5487 (22.7)
N visits in study, N (%)	
1-2	2241 (9.3)
3-5	4889 (20.3)
6-9	6407 (26.6)
$\geq 10$	10 593 (43.9)
N visits with PCP, N (%)	
0-2	4499 (18.6)
3-5	7044 (29.2)
6-9	6995 (29.0)
$\geq 10$	5592 (23.2)
% of visits with PCP, N (%)	
$\leq 25\%$	2038 (8.4)
26-50%	4165 (17.3)
51-75%	6736 (27.9)
75-100%	11 191 (46.4)
Charlson comorbidity index, N (%)	
0-2	7255 (30.1)
3-4	6074 (25.2)
5-7	6675 (27.7)
$\geq 8$	4126 (17.1)
Hypertension on problem list, N (%)	16 595 (68.8)
Lipid disorder on problem list, N (%)	16 820 (69.7)

Notes: FPL=federal poverty level.

PCP=primary care provider.

Charlson comorbidity index calculated from encounter and problem list diagnoses as of December 31, 2013; "enhanced" index used by permission (Mary Charlson, 2017).

**Table 2.** Dispensing percentages and distributions by health center, by medication class

	N prescribed	N dispensed	% dispensed	Distribution of proportion dispensed across health centers, median (Q1, Q3)
Cardiovascular drugs, total	203 504	154 742	76.0%	78.3 (66.7, 85.2)
ACEs, ARBs, and DRIs	60 158	45 565	75.7%	78.3 (64.7, 85.7)
Statins	50 178	38 455	76.6%	77.5 (65.1, 85.6)
Diuretics	27 274	21 259	77.9%	80.3 (70.0, 85.0)
Beta blockers	23 687	18 754	79.2%	79.4 (66.5, 86.8)
Platelet inhibitors	18 369	12 418	67.6%	63.5 (50.0, 76.3)
Calcium channel blockers: dihydropyridine, oral	13 951	10 805	77.4%	78.8 (66.8, 88.7)
Other dyslipidemia drugs	4957	3636	73.4%	76.9 (57.9, 82.4)
Antianginals	2913	2271	78.0%	77.0 (53.6, 86.4)
Calcium channel blockers: non-dihydropyridine, oral	2017	1579	78.3%	85.2 (71.4, 93.3)
Diabetes drugs, total	160 189	119 239	74.4%	77.4 (64.3, 84.8)
Miscellaneous antidiabetic agents (metformin)	63 581	47 640	74.9%	74.6 (66.8, 83.7)
Insulins	59 035	44 028	74.6%	73.0 (61.4, 85.5)
Sulfonylureas	28 727	21 513	74.9%	77.4 (67.8, 85.8)
DPP-4 inhibitors	5544	3653	65.9%	65.9 (50.5, 76.1)
Thiazolidinediones	2146	1726	80.4%	85.1 (77.8, 92.8)
SGLT2 inhibitors	653	363	55.6%	66.7 (25.0, 84.0)
GLP-1 Receptor agonists	503	316	62.8%	66.3 (50.0, 90.3)

## Notes:

ACE=angiotensin-converting enzyme inhibitors.

ARB=angiotensin II receptor blockers.

DRI=direct renin inhibitor.

DPP-4 = dipeptidyl peptidase-4.

SGLT2 = sodium-glucose cotransporter-2.

GLP-1 = glucagon-like peptide-1.

Overall, CVD drugs prescribed for patients with diabetes were dispensed 76% of the time (Table 2). Dispensing rates ranged from 68% for platelet inhibitors to 79% for beta blockers. Diabetes drugs were dispensed at slightly lower rates than CVD medications: overall 74% of DM prescriptions were dispensed, ranging from 56% for SGLT2 inhibitors to 80% for thiazolidinediones. While distributions across health centers show some variation in dispensing rates, the vast majority of clinics had over 50% of their CVD and diabetes medications dispensed (Table 2). We computed dispensing rates within several subpopulations as a sensitivity analysis to explore whether certain patient factors were driving these results. Dispense rates excluding patients ever uninsured, Hispanic, and those with only one study visit were largely consistent with those found in our full sample (Supplementary Materials, Table 1).

Figure 2 shows patient-, encounter-, and facility-level predictors of dispensing for patients with diabetes after adjusting for sex, age, and race/ethnicity. Both CVD and diabetes drug classes had equal odds of being dispensed (OR = 1.01, 95% CI = 0.99-1.02). Patient-level factors associated with greater odds of dispensing were sex (OR = 1.14, 95% CI = 1.09-1.20, female vs. male), older ages, and income  $\leq$ 138 FPL (OR = 1.42, 95% CI = 1.32-1.53, vs. >138 FPL). Ethnicity (OR = 0.49, 95% CI: 0.46-0.53, Hispanic) and non-white race were associated with lower odds of dispensing relative to non-Hispanic whites. All insurance coverage categories had significantly lower odds of being dispensed relative to those fully covered by Medicaid, and patients newly insured in 2014 (no coverage in 2013) had lower odds of being dispensed than the continuously insured (OR = 0.68, 95% CI = 0.64-0.72). Having a PCP assigned, greater proportion of visits with a PCP, residence in a Medicaid expansion state, and higher comorbidity scores were all associated with increased odds of being dispensed.

At the encounter level, prescriptions that were refills (OR = 1.03, 95% CI = 1.01-1.04), e-prescribed (OR = 1.18, 95% CI = 1.14-1.22), and prescribed by the patient's PCP (OR = 1.11, 95% CI = 1.08-1.14) had higher odds of being dispensed. Encounter provider types other than medical doctors were marginally associated with lower odds of adherence, but reason for visit and patients' visit counts at the same clinic were not. At the clinic level, attesting to Meaningful Use (OR = 1.36, 95% CI = 1.20-1.53), larger clinic size, and longer time on the EHR were significant predictors of dispensing. Table 2 in Supplementary Materials lists dispensing proportions at each level of all predictor variables and adjusted odds of dispense with confidence intervals.

## DISCUSSION

We created an algorithm that successfully linked EHR-based prescriptions to claims-based dispensing data for a population of CHC patients with diabetes, and found dispensing rates for diabetes and CVD drugs ranging from 56% to 80% depending on drug class. Despite the vulnerability of our patient population, these rates align with previous studies in other populations showing all prescription fill rates were about 75%.<sup>6,8,9</sup> In addition, meta-analyses have shown medication adherence of approximately 70% for individuals with DM;<sup>25</sup> antihypertensive medication fill rates were previously measured at 66% within a population of Medicare patients.<sup>26</sup>

Much attention has been given to the challenges of mapping among different medication terminology systems.<sup>27-29</sup> This work can be time and resource intensive and difficult to automate, due in part to one-to-many relationships between terminologies, such as RxNorm and NDC, differences in granularity among terminologies,



**Figure 2.** Odds of medication dispensed, by patient-, encounter-, and clinic-level factors. Footnote: Odds ratios and 95% CIs from GEE models with a logit link; robust standard error estimator applied to account for repeated measures within patients and patients nested within health centers. All models adjusted for sex, age group, and race/ethnicity. FPL predictor model restricted to facilities with sufficient variation for model conversion.

and ever-evolving systems that require mappings to be continually updated.<sup>27</sup> We took a rather generalized approach by mapping medications prescribed to those dispensed within pharmaceutical subclass. This method allows matches to be made even when an appropriate alternative medication was dispensed (eg, differing trade names for prescribed and dispensed medications with the same active ingredient), as long as they fell within the same pharmaceutical subclass and matched within the RxNorm-NDC crosswalk con-

structed through the RxMix tool. For the current work, we stopped short of including additional elements such as brand or generic name, dosage, strength, or number of days' supply, but such data could be integrated into the matching algorithm for further refinement where additional specificity is required.

Similar algorithms have been developed to match EHR prescriptions to dispensing data in integrated healthcare systems or within national pharmacy chains.<sup>6,30</sup> The reproducibility of such efforts in

the broader US healthcare environment is challenged by diverse geographies, fragmented healthcare systems, multiple payers, and patient populations with unstable insurance coverage and access to care. By developing this algorithm within the framework of the PCORnet CDM, we present a generalizable method with applications across the diverse PCORnet distributed research network, which currently represents more than 100 million patients.<sup>31</sup>

Emerging extensions to this work include the movement toward a national quality measure for primary medication nonadherence<sup>32</sup> and potential applications to medication reconciliation workflows. For example, a recent paper presents a web-based software capable of an e-medication reconciliation application to automate medication reconciliation between community and hospital drug lists in Canada.<sup>33</sup> Such efforts have the potential to dramatically improve patient safety and disease management through improved focus on medication adherence and the reduction of barriers; over time, the need for more automated and accurate medication-matching methods across different sources will likely increase.

In the current study, medications with the lowest dispense rates were DPP4 inhibitors, SGLT2 inhibitors, and GLP-1 receptor agonists, which are new and relatively costly DM medications. These medications often require prior authorization from a provider attesting that less costly medications have been used but failed to control glucose levels;<sup>24</sup> therefore, it was unsurprising to find many of these were prescribed but not dispensed. Our methods likely represent conservative estimates for classes that contain low-cost over-the-counter drugs; for example, the dispense rate for platelet inhibitors increases from 68% to 75% when excluding aspirin (Supplementary Materials, Table 1).

We also found that patient characteristics (eg, age, ethnicity, insurance type), provider type, and clinic factors (eg, time on EHR, clinic size) were associated with prescribed medications being dispensed. These findings reinforce prior studies of the multifactorial influences on medication non-adherence. Health literacy, lack of communication, and fragmented healthcare systems have been highlighted as reasons for non-adherence.<sup>26,34</sup> The current finding that the strongest associations were at the patient level adds to this literature. In particular, having Medicaid insurance was strongly associated with having prescriptions dispensed. Given that most Medicaid plans require a \$0 or very minimal co-pay, our data suggest that lack of prescription cost sharing is an important barrier-reducing factor in this population.

While we successfully developed an algorithm to link prescribing to dispensing data, we do not know if patients subsequently took the dispensed medications or used them as directed. More research is needed to measure and describe medication adherence and persistence (ie, the act of continuing a medication for the prescribed duration over time) in the CHC population, as well as to understand reasons for discontinuation. Further, identifying barriers to adherence and reasons for discontinuation will allow clinicians and researchers to create meaningful interventions that address systemic barriers to adherence. Informatics-based approaches combining multiple data sources should continue to be developed,<sup>30,35</sup> which may in turn facilitate patient-provider communication about appropriate medication use and ultimately improve patient adherence and disease control.

This study has several limitations. Approximately 2% of matched prescriptions did not agree on brand and/or generic names and were matched within pharmaceutical subclass only; this small proportion may represent false positives (ie, the dispensing record did not actually originate from its matched prescribing record). An algorithm based on brand and generic medication names instead of

pharmaceutical subclass may prove more sensitive. For the 30% of EHR prescriptions that did not match to a dispensing record, we cannot say for certain whether these went unfilled or failed to match a dispense record; however, the consistency of dispensing rates with prior published work, combined with our validity checks on matched records, lends confidence to our matching algorithm. Due to the retrospective nature of this work, we included historical (ie, inactive) NDC codes in our crosswalk. This is but one issue to be aware of when working with NDCs: because codes can be reused five years after being inactive, the same NDC can represent more than one drug.<sup>27</sup> Further, because NDCs are constantly updated, the NDC is not necessarily a reliable indicator of exactly what was dispensed, but rather what was billed. We were limited to assessing medications dispensed for patients insured at the time of dispense. Consequently, we had to exclude nearly 25% of our study sample who were uninsured throughout the study period. In addition, some degree of missingness in the claims-based dispensing data is likely.<sup>36</sup> For example, if patients are given manufacturer samples in the clinic or purchase medications under \$4 generic plans without using insurance, these dispenses would not be recorded.<sup>37</sup> Previous research suggests this is likely not drastically impacting observed dispense rates,<sup>36,38</sup> but for these reasons, our reported dispensing rates may be underestimated.

## CONCLUSION

This study developed an algorithm to link EHR-based prescriptions to claims-based dispensing data for a large network of CHCs, matching 70% of prescribed medications using data from the PCORnet common data model. Our analysis demonstrating an application of this algorithm found dispense rates in the range of many published studies assessing other populations. Our findings suggest that prescribing data may be used as a reasonable proxy for medication dispensing. While this study specifically examined chronic disease medication dispensing among patients with diabetes, the algorithm developed can be applied to other medication classes and patient populations. These findings lay the groundwork for future studies of medication adherence and chronic disease management using EHR data.

## FUNDING

This work was supported by the U.S. Centers for Disease Control and Prevention and the National Institute of Diabetes and Digestive and Kidney Disease through a joint cooperative agreement (Cooperative Agreement Number U18DP006116); the National Heart, Lung, and Blood Institute (grant number R01HL136575); and the Patient-Centered Outcomes Research Institute.

*Conflict of interest statement.* The authors have no competing interests to declare.

## REGISTRATION

This project is registered with ClinicalTrials.gov (NCT02685384). Registered February 18, 2016.

## CONTRIBUTORS

M. H. helped conceive of the study, extracted and linked the data, conducted the analysis, interpreted the data, and wrote the manuscript. H. A. helped interpret the data and contributed to writing the



manuscript. L.A. R. helped conceive of the study, interpreted the data, and contributed to writing the manuscript. A. S. provided clinical input on inclusion criteria, interpreted the data, and provided substantive edits. J. M. provided clinical input on inclusion criteria, interpreted the data, and provided substantive edits. M. M. interpreted the data and provided substantive edits. P. R. developed the medication-matching algorithm, contributed to the writing, and provided substantive edits. N. H. helped conceive of the study, interpreted the data, and provided substantive edits.

## SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

## REFERENCES

- Gabriel MH, Swain M. *E-Prescribing Trends in the United States*. Washington, DC: The Office of the National Coordinator for Health Information Technology; 2014.
- US Department of Health and Human Services, US National Library of Medicine. RxMix. Bethesda, MD; 2017.
- Lehmann A, Aslani P, Ahmed R, et al. Assessing medication adherence: options to consider. *Int J Clin Pharm* 2014; 36 (1): 55–69.
- Mabotuwana T, Warren J, Harrison J, Kenealy T. What can primary care prescribing data tell us about individual adherence to long-term medication? Comparison to pharmacy dispensing data. *Pharmacoepidemiol Drug Saf* 2009; 18 (10): 956–64.
- Lindgren A, Stroh E, Jakobsson K. Ever dispense of prescribed allergy medication in children growing up close to traffic: a registry-based birth cohort. *BMC Public Health* 2015; 15 (1): 1023.
- Fischer MA, Choudhry NK, Brill G, et al. Trouble getting started: predictors of primary medication nonadherence. *Am J Med* 2011; 124 (11): 1081.e9–22.
- Polonsky WH, Henry RR. Poor medication adherence in type 2 diabetes: recognizing the scope of the problem and its key contributors. *Patient Prefer Adherence* 2016; 10: 1299–307.
- Comer D, Couto J, Aguiar R, Wu P, Elliott D. Using aggregated pharmacy claims to identify primary nonadherence. *Am J Manag Care* 2015; 21 (12): e655–60.
- Zhang H, Plutzky J, Shubina M, Turchin A. Risk factors for lack of statin therapy in patients with diabetes and coronary artery disease. *J Clin Lipidol* 2016; 10 (6): 1406–13.
- Khanna R, Pace PF, Mahabaleshwarar R, Basak RS, Datar M, Banahan BF. Medication adherence among recipients with chronic diseases enrolled in a state Medicaid program. *Popul Health Manag* 2012; 15 (5): 253–60.
- Qaseem A, Barry MJ, Humphrey LL, Forciea M. Oral pharmacologic treatment of type 2 diabetes mellitus: a clinical practice guideline update from the American College of Physicians. *Ann Intern Med* 2017; 166 (4): 279–90.
- Stamler J, Vaccaro O, Neaton JD, Wentworth D. Diabetes, other risk factors, and 12-yr cardiovascular mortality for men screened in the Multiple Risk Factor Intervention Trial. *Diabetes Care* 1993; 16 (2): 434–44.
- Gold R, Bunce A, Cowburn S, et al. Cardiovascular care guideline implementation in community health centers in Oregon: a mixed-methods analysis of real-world barriers and challenges. *BMC Health Serv Res* 2017; 17 (1): 253.
- Khunti K, Seidu S, Kunutsor S, Davies M. Association between adherence to pharmacotherapy and outcomes in type 2 diabetes: a meta-analysis. *Diabetes Care* 2017; 40 (11): 1588–96.
- Cramer JA. A systematic review of adherence with medications for diabetes. *Diabetes Care* 2004; 27 (5): 1218–24.
- Corley DA, Feigelson HS, Lieu TA, McGlynn EA. Building data infrastructure to evaluate and improve quality: PCORnet. *J Oncol Pract* 2015; 11 (3): 204–6.
- DeVoe JE, Gold R, Cottrell E, et al. The ADVANCE network: accelerating data value across a national community health center network. *J Am Med Inform Assoc* 2014; 21 (4): 591–5.
- The National Patient-Centered Clinical Research Network (PCORnet). PCORnet Common Data Model; 2017.
- Wei M, Robin M, Vikraman G, Stuart N, Simon L. RxNorm: prescription for electronic drug information exchange. *IT Prof* 2005; 7: 17.
- US Department of Health and Human Services, Office of the National Coordinator for Health Information Technology. Federal Register, 45 CFR Part 170. Health Information Technology: Standards, Implementation Specifics, and Certification Criteria for Electronic Health Record Technology, 2014 Edition; Revisions to the Permanent Certification Program for Health Information Technology; 2012.
- US Department of Health and Human Services, US Food & Drug Administration. Silver Spring, MD: National Drug Code Directory; 2017.
- Huguet N, Angier H, Marino M, et al. Protocol for the analysis of a natural experiment on the impact of the Affordable Care Act on diabetes care in community health centers. *Implement Sci* 2017; 12 (1): 14.
- Charlson ME, Charlson RE, Peterson JC, Marinopoulos SS, Briggs WM, Hollenberg JP. The Charlson comorbidity index is adapted to predict costs of chronic disease in primary care patients. *J Clin Epidemiol* 2008; 61 (12): 1234–40.
- Oregon Health Authority. Oregon Fee-for-Service Enforceable Physical Health Preferred Drug List. Table 121-0030-1. Salem, OR; 2016.
- Tunceli K, Zhao C, Davies MJ, et al. Factors associated with adherence to oral antihyperglycemic monotherapy in patients with type 2 diabetes. *Patient Prefer Adherence* 2015; 9: 191–7.
- Brown MT, Bussell JK. Medication adherence: WHO cares? *Mayo Clin Proc* 2011; 86 (4): 304–14.
- Saitwal H, Qing D, Jones S, Bernstam EV, Chute CG, Johnson TR. Cross-terminology mapping challenges: a demonstration using medication terminological systems. *J Biomed Inform* 2012; 45 (4): 613–25.
- Pathak J, Chute CG. Analyzing categorical information in two publicly available drug terminologies: RxNorm and NDF-RT. *J Am Med Inform Assoc* 2010; 17 (4): 432–9.
- Nadkarni PM. Drug safety surveillance using de-identified EMR and claims data: issues and challenges. *J Am Med Inform Assoc* 2010; 17 (6): 671–74.
- Parker MM, Moffet HH, Adams A, Karter AJ. An algorithm to identify medication nonpersistence using electronic pharmacy databases. *J Am Med Inform Assoc* 2015; 22 (5): 957–61.
- PCORnet: The National Patient-Centered Clinical Research Network. PCORnet Data; 2018. <http://pcornet.org/pcor-net-data/>. Accessed April 5, 2018.
- Adams AJ, Stolpe SF. Defining and measuring primary medication nonadherence: development of a quality measure. *J Manag Care Spec Pharm* 2016; 22 (5): 516–23.
- Tamblyn R, Winslade N, Lee TC. Improving patient safety and efficiency of medication reconciliation through the development and adoption of a computer-assisted tool with automated electronic integration of population-based community drug data: the RightRx project. *J Am Med Inform Assoc* 2017; 25 (5): 482–95.
- De Geest S, Sabate E. Adherence to long-term therapies: evidence for action. *Eur J Cardiovasc Nurs* 2003; 2 (4): 323.
- Dixon BE, Jabour AM, Phillips EOK, Marrero DG. An informatics approach to medication adherence assessment and improvement using clinical, billing, and patient-entered data. *J Am Med Inform Assoc* 2014; 21 (3): 517–21.
- Lauffenburger JC, Balasubramanian A, Farley JF, et al. Completeness of prescription information in US commercial claims databases. *Pharmacoepidemiol Drug Saf* 2013; 22 (8): 899–906.
- Choudhry NK, Shrank WH. Four-dollar generics—increased accessibility, impaired quality assurance. *N Engl J Med* 2010; 363 (20): 1885–7.
- Zhang Y, Gellad WF, Zhou L, Lin YJ, Lave JR. Access to and use of \$4 generic programs in Medicare. *J Gen Intern Med* 2012; 27 (10): 1251–7.