



Published in final edited form as:

Med Care. 2020 June ; 58(Suppl 6 1): S46–S52. doi:10.1097/MLR.0000000000001298.

Using Electronic Health Records in Longitudinal Studies:

Estimating Patient Attrition

Nathalie Huguet, PhD*, Jorge Kaufmann, ND, MS*, Jean O'Malley, MPH*, Heather Angier, PhD, MPH*, Megan Hoopes, MPH†, Jennifer E. DeVoe, MD, DPhil*, Miguel Marino, PhD*,‡

*Department of Family Medicine, Oregon Health and Science University

†OCHIN Inc.

‡Biostatistics Group, Oregon Health and Science University—Portland State University School of Public Health Portland, Portland, OR.

Abstract

Background: Electronic health records (EHRs) provide rich data on many domains not routinely available in other data, as such, they are a promising source to study changes in health outcomes using longitudinal study designs (eg, cohort studies, natural experiments, etc.). Yet, patient attrition rates in these data are unknown.

Objective: The objective of this study was to estimate overall and among adults with diabetes or hypertension: (1) patient attrition over a 3-year period at community health centers; and (2) the likelihood that patients with Medicaid permanently switched their source of primary care.

Research Design: A retrospective cohort study of 2012–2017 data from the Accelerating Data Value Across a National Community Health Center Network (ADVANCE) Clinical Data Research Network of community health centers were used to assess EHR data attrition. Oregon Medicaid enrollment and claims data were used to estimate the likelihood of changing the source of primary care.

Subjects: A total of 827,657 patients aged 19–64 with 1 ambulatory visit from 76 community health center systems across 20 states. In all, 232,891 Oregon Medicaid enrollees (aged 19–64) with a gap of 6 months following a claim for a visit billed to a primary care source.

Measures: Percentage of patients not returning within 3 years of their qualifying visit (attrition). The probability that a patient with Medicaid permanently changed their primary care source.

Results: Attrition over the 3 years averaged 33.5%; attrition rates were lower (< 25%) among patients with diabetes and/or hypertension. Among Medicaid enrollees, the percentage of provider change after a 6-month gap between visits was 12% for community health center patients compared with 39% for single-provider practice patients. Over 3 years, the likelihood of a patient

Correspondence to: Nathalie Huguet, PhD, 3181 Southwest Sam Jackson Park Road, Portland, OR 97239. huguetn@ohsu.edu.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.lww-medicalcare.com.

The authors declare no conflict of interest.

changing to a new provider increased with length of time since their last visit but remained lowest among community health center patients.

Conclusion: This study demonstrates the use of the EHR dataset is a reliable source of data to support longitudinal studies while highlighting variability in attrition by primary care source and chronic conditions.

Keywords

electronic health records; attrition; natural experiment; community health center; primary care; Medicaid

As the reach and connectivity of electronic health records (EHRs) have grown in the United States,^{1,2} large multi-state networks of clinics with linked EHRs have the potential to provide rich longitudinal data on many domains not routinely available in other data sources. For example, EHRs contain information on metabolic biomarkers (eg, blood pressure, glycosylated hemoglobin A1c, etc.), comprehensive diagnostic information, millions of encounter data points, and reliable information on insurance and uninsurance status from billing data at the time of the encounter.^{3,4} As such, EHRs are a promising data source to investigate changes in health outcomes using longitudinal study designs (eg, cohort studies, natural experiments, etc.). Although all longitudinal data sources can experience loss to follow-up, in many studies researchers have some control over the engagement and follow-up activities that potentially impact participant attrition. However, when utilizing EHR data for longitudinal research, data capture is driven by clinical and administrative encounters that characterize the patient-provider environment rather than research needs. There are several reasons why patient attrition might occur within EHR data, including: (1) patients infrequently seek care at any clinic (nonutilizers); (2) patients seek care at a clinic or setting outside of the EHR; data network (provider change), or (3) death (not captured systematically by most EHR systems). Understanding rates of attrition in EHR data is important to inform longitudinal study design considerations, power/sample size estimation, and selection of appropriate statistical methods to address attrition. In addition, when planning longitudinal study designs for subpopulations, it is likely that attrition rates vary from the overall population and therefore should be accounted for. For example, patients with chronic conditions may seek care more frequently (lowering attrition) or be more likely to seek care elsewhere (increasing attrition). Thus, it is important to conduct stratified analyses to assess whether attrition rates differ among patients with common chronic conditions compared with rates among the overall population. To our knowledge, there are no studies assessing attrition using EHR data, overall or by disease. Therefore, the objective of this study was to provide several estimates of patient attrition from the EHR. We used the Affordable Care Act (ACA)⁵ natural experiment as a case study.^{6,7} Briefly, the ACA increased the availability of health insurance by developing a health insurance marketplace and expanding Medicaid eligibility to adults with income \leq 138% of the federal poverty level in states that chose to expand. Since a large number of patients without insurance gained coverage following the ACA,^{8,9} it was expected that newly insured patients would seek care elsewhere, thus attrition rates may increase following the ACA. We leveraged EHR data

from community health centers because they serve a large proportion of patients without insurance or with Medicaid coverage.¹⁰

In addition to estimating overall patient attrition rates, we sought to estimate the proportion of patients who changed their source of primary care overall and among those with diabetes or hypertension. We did so by utilizing Medicaid enrollment and claims data to compare patient movement from community health centers to other primary care settings. The objective of this secondary analysis was to determine if patient attrition was mainly due to changing providers, rather than patients not seeking care. Thus, using these 2 data sources (EHR and Medicaid claims), this study aimed to estimate: (1) attrition rates in primary care community health center practices using EHR data; and (2) the likelihood that patients who did not return to a community health center changed primary care providers. We hypothesized that attrition rates would be even lower for adults with diabetes or hypertension, as they are likely to favor continuity of care to manage their condition. We also hypothesized that patients from community health centers would be less likely to switch providers than patients in other care settings.

METHODS

This study was approved by the Institutional Review Board.

Accelerating Data Value Across a National Community Health Center Network Electronic Health Record Data

In this retrospective cohort study spanning 2012–2017, we utilized EHR data from the Accelerating Data Value Across a National Community Health Center Network (ADVANCE) Clinical Data Research Network of community health centers, a network within PCORnet.¹¹ Patients of ADVANCE that met our inclusion criteria (described below) were from 76 community health centers with 340 delivery sites across 20 states (Alaska, California, Florida, Hawaii, Indiana, Kansas, Massachusetts, Maryland, Minnesota, Missouri, Montana, North Carolina, New Mexico, Nevada, Ohio, Oregon, Rhode Island, Texas, Washington, Wisconsin). We utilized encounter data from facilities that did not close before 2018 (closure date was inferred from the last encounter preceding 4 mo of inactivity). Patients must have had ≥ 1 ambulatory visit to an ADVANCE community health center in any of 2012, 2013, and/or 2014. We chose patients aged 19–64, as this is the age range most affected by the ACA Medicaid expansion, which is the longitudinal natural experiment we used as a case example for this study.^{6,7} We also excluded patients who were ever pregnant during the study period, as pregnancy confers different care needs and higher visit rates, which could bias our attrition estimates. Patients were followed for 3 years after their qualifying ambulatory visit consistent with criteria for distinguishing an established patient from a new patient in Evaluation and Management procedure codes. In all, 827,657 patients from the ADVANCE EHR dataset were included in this study sample.

Oregon Medicaid Data

We obtained Oregon Medicaid enrollment data and outpatient administrative claims data (January 1, 2002, to December 31, 2017) from the Oregon Health Authority. We identified

billing primary care providers associated with office visits in claims data. We grouped these billing providers into primary care sources using their National Provider Identifier reported legal business names. Primary care sources were categorized into 4 billing provider categories based on National Provider Identifier taxonomy groups, supplemented by Oregon Health Authority provider types for greater specificity. The 4 categories were: (1) community health centers (Federally qualified health centers, rural health centers, or other community health centers); (2) primary care clinics (organizations with predominantly primary care performing providers); (3) multispecialty clinics (organizations billing for a mix of specialists and primary care providers); and (4) single-primary care practice (billing providers with National Provider Identifier taxonomy codes indicating a physician rather than a clinic or business).

Cohort Definitions

We followed 3 EHR patient cohorts over time based on the year a patient had a qualifying visit to a community health center (2012, 2013, or 2014). The purpose of these stratified analyses was to evaluate if attrition differed by year particularly for our case example, as 2012–2013 were before the ACA Medicaid expansion and 2014 was after. For analyses that utilized ADVANCE EHR data, there were 827,657 patients with 1 ambulatory visit. For each cohort, we observed EHR data for 3 years and described attrition rates. For example, for the 2012 cohort, we estimated the percent of patients who *never had a visit* in 2013, 2014, or 2015 (*attrition*). These measures were also estimated for the 2013 and 2014 cohorts. Given that diabetes and hypertension are 2 of the most common chronic conditions in this population, we subset each cohort-year into groups of patients with either diabetes or hypertension and estimated the attrition measure described below (patients with diabetes n = 2012: 51,619; 2013: 59,769; 2014: 67,420; patients with hypertension n = 2012: 136,050; 2013: 158,470; 2014: 174,633). Patients with diabetes and hypertension included both patients with a new diagnosis and those with existing diagnosis at the start of each cohort-year.

Analyses that utilized Oregon Medicaid data consisted of 3 cohorts patients with 1 primary care visit in 2012–2014 billed by a community health center, primary care clinic, multispecialty clinic, or single-primary care practice that was followed by a visit gap of at least 6 months during which the patient was continuously enrolled in Medicaid (n = 232,891). Visit gaps are defined below.

Chronic Condition Definitions

Patients with diabetes were identified using a modified version of the Surveillance, PREvention, and ManagEment of Diabetes Mellitus (SUPREME-DM) definition¹² in the EHR data. The SUPREME-DM requires 2 of any of the following events within 2 years: diabetes-related medications prescribed (eg, insulin, sulfonylurea, etc.), hemoglobin A1c or glucose laboratory test result, and/or a diabetes diagnosis on the outpatient problem list. The original definition included in-patient diagnosis and dispensed medication, which are not available in our dataset. Patients with hypertension were identified using a definition from Selby et al,¹³ which requires any 2 of the following within a 2-year period: hypertension diagnosis, hypertension or blood pressure prescription, or consecutive high blood pressure

measures (≥ 140 mm Hg systolic and/or 90 mm Hg diastolic). As with the year cohorts, these disease-specific cohorts were not mutually exclusive. Our Medicaid data did not include laboratory results or vital signs, so patients with diabetes and hypertension in our Medicaid yearly cohorts were defined as patients with at least 2 diagnosis codes for the condition on or before the date of the visit starting the visit gap (defined below).

Primary Care Visit Gaps in Medicaid Data

The time between any visit billed to a primary care source and a subsequent visit billed to the same primary care source was defined as a visit gap. Visits with gaps of ≥ 6 months were linked to the patient enrollment data to determine whether the patient was continuously enrolled in Medicaid during the visit gap. If a patient was no longer enrolled during the gap, the length of the gap was censored at the last enrollment date. We chose a minimum period of 6 months because the majority of consecutive visits to a primary care provider were spaced ≥ 6 months apart and patients with chronic conditions require frequent visits. Patients could have multiple visits and visit gaps during the study period. The objective of this approach was to estimate the proportion of patients who had had no contact with a primary care provider for a specified length of time (non-utilizer) from those who had actually changed their primary care provider in that time interval (attrition).

Primary Care Provider Change

To estimate the likelihood of a patient with a visit gap sought care elsewhere, visit gaps from patients from Medicaid data were linked to patient dates of primary care visits from other primary care sources. This linkage was done to determine the date at which the patient changed their source of primary care (defined as the date of the first primary care visit billed by a different primary care source that was not followed by a return to the primary care source that billed the visit starting the visit gap). Patients who changed providers early in a gap (ie, in the first 6 mo after their last visit) were included in the proportion who changed in subsequent time intervals (12, 18, 24, 30, and 36 mo) if patient enrollment continued.

Statistical Analysis

We describe demographic characteristics of the entire study population and our diabetes or hypertension subpopulations for both the ADVANCE EHR and Medicaid data. We report counts and percentages for all groups with regard to attrition.

To determine the probability of a patient permanently changing primary care provider source as a function of time since the patient's last visit, we used Medicaid data. We subdivided the gaps into periods of 6 months and assessed whether or not the patient switched to another provider before the end of each 6-month period in the gap. Once a patient changed primary care providers, this change was carried forward into any remaining periods in the gap (eg, a patient who changed providers 15 mo after the visit that started the gap would be recorded as not having changed at 6 mo and 12 mo but would be recorded as having changed at 18 mo and at all subsequent 6-mo intervals). The probability of provider change was modeled separately for each patient cohort and time interval (6, 12, 18, 24, 30, and 36 mo). As patients could have multiple gaps of 6 months or more between visits resulting in multiple observations per patient within each analysis, we used generalized estimating equation

(GEE) logistic regression to account for within-patient correlation using a robust sandwich variance estimator. GEE models included primary care provider type, cohort year, and the interaction term between those 2 terms. Odds ratios from GEE logistic models were transformed to estimate predicted probabilities of provider change. Analyses were conducted in Stata version 14.2 and SAS enterprise v.7.15.

RESULTS

Demographic Characteristics

Among the 827,657 patients in the study sample, the mean age was 39 (SD: 13 y), 58% were female, 41% non-Hispanic white, and 29% Hispanic/Latino. The vast majority (80%) of patients spoke English as their primary language. Over 70% had a household income 138% of the federal poverty level (the ACA Medicaid expansion eligibility limit) at their last visit, and the majority resided in an urban area (Table 1). Almost half of all patients (48%) had public health insurance at some point during their observation period, while nearly 30% were always uninsured. Among the Medicaid enrollees, characteristics were similar to patients in the community health center EHR data, except for the racial/ethnicity distribution (Medicaid enrollees were more likely to be non-Hispanic whites). Demographic characteristics for patients with diabetes and hypertension were similar to that of the overall population and are reported in Appendix Table 1 (Supplemental Digital Content 1, <http://links.lww.com/MLR/B968>).

Attrition Within Community Health Center Electronic Health Record Data

Overall, attrition over a 3-year period averaged 33.5%, implying that 66.5% of the patients had 1 subsequent visit in the next 3 years (Table 2). The attrition rates increased slightly between 2012 (31.8%) and 2014 cohorts (34.3%). More than half (57%) of the overall population returned in the year following their cohort start. For patients with diabetes or hypertension, the attrition rate averaged <25% (with an average 78% retention rate over 3 y). The attrition rate was slightly higher among those with diabetes compared with patients with hypertension; about 73% of patients with diabetes or hypertension had a visit in the subsequent year following their cohort start (Table 2).

Probability of Changing Source of Primary Care Using Medicaid Data

Overall, 64.6% of patients experienced a 6-month visit gap at some point in the 3-year time period. As seen in Figure 1, community health center patients were significantly less likely to change their source of primary care than patients in other settings. In 2012, 12% of community health center patients who had not been seen in their community health center for 6 months had permanently changed to a different primary care billing provider compared with 23% for noncommunity health center primary care clinics, 17% for multispecialty clinics, and 39% for single-primary care practice. Patients with a 36-month visit gap had a high probability of having changed providers, with the highest rates observed in single-primary care practice (85%) and the lowest in community health centers (62%). Overall, the probability of changing providers was stable across the yearly cohorts (2012, 2013, and/or 2014) for all care settings (Appendix Table 2, Supplemental Digital Content 2, <http://links.lww.com/MLR/B969>).

A smaller percentage of patients with diabetes (41.8%) or hypertension (41.0%) experienced a 6-month visit gap at some point in the 3-year period. Figure 2 displays the results for changing the source of primary care among patients with diabetes or hypertension for the 2014 cohort (the results for the remaining cohorts are in Appendix Table 3, Supplemental Digital Content 3, <http://links.lww.com/MLR/B970>). Those seen in community health centers remained the least likely to change primary care providers. As noted above, patients with diabetes or hypertension were less likely to have a visit gap meeting the inclusion criteria. Patients with diabetes and hypertension who did have a visit gap of 6 months were more likely to have changed providers than patients overall. Among patients who did not have a visit 36 months after their last visit, 66% of community health center patients (diabetes or hypertension) had permanently changed compared with > 70% for noncommunity health center primary care clinics or for multispecialty clinics, and 80% for single-primary care practice.

DISCUSSION

The increasing use of EHRs in clinics and hospitals offer access to rich health and health care data that can be utilized for longitudinal studies. However, as EHRs were not originally developed for research, the extent to which these data are reliable for longitudinal studies was unknown. Here, we estimated attrition rates and the probability of changing primary care providers using the ACA natural experiment as a case study. Specifically, we assessed attrition rates using EHR data and the probability of patients permanently changing sources of care with Medicaid data. We found an attrition rate over a 3-year period in the community health centers' EHR dataset of around 33%, which is similar to longitudinal studies using prospectively collected data.¹⁴⁻¹⁶ Overall, attrition was similar for each yearly cohort despite the introduction of the ACA, suggesting that attrition rates are likely to remain stable throughout a defined analysis period in the community health center setting. Contrary to expectation, our results did not support the assumption that patients who gained coverage following the ACA, sought care outside of community health centers as the attrition rates and the probability of changing providers remained constant over the ACA, implementation. In fact, the study findings suggest that low-income patients receiving care in community health centers are more likely to remain in these health centers than patients who sought care in other primary care sources initially. In addition, it is not likely that community health center patients without an observed follow-up visit re-established care elsewhere given results from Medicaid data analyses showing that community health center patients are less likely to permanently change their source of care compared with patients in other settings. In other words, it cannot be assumed that patients missing from subsequent years in a community health center EHR dataset changed providers; it is probable that these underserved patients are not seeking/receiving health care at all. This is probably because noncommunity health center settings are less likely to accept patients with Medicaid coverage or those who are uninsured than community health centers.

This study demonstrated that the health status of individual patients or cohorts should be considered in the study planning and analytic phase of EHR-based longitudinal studies as attrition varied by a patient's chronic health needs. This study confirmed our hypothesis that attrition rates were low in community health centers and that patients with diabetes or

hypertension would have lower rates of attrition than the overall population, which is not surprising given that patients with chronic conditions require more frequent contact with health care professionals to manage and treat their conditions.^{17,18} As such, they are more likely to have regular ambulatory visits and maintain a usual source of primary care. Patients with diabetes or hypertension who had visit gaps, however, were more likely to have changed providers than patients without chronic conditions, suggesting that their visit gaps are less likely to be due to not accessing care for an extended period and more likely to be due to changing providers.

Overall, these findings support the use of EHR data in the longitudinal design such as those described in this case study. These findings also highlight that attrition in EHR-based studies is nontrivial and researchers should plan to address attrition bias in their analytic protocols using common statistical methods including, but not limited to descriptively comparing those who left to those who remained in the system, inverse probability weighting methods,¹⁹ full-information maximum likelihood,²⁰ and/or multiple imputations.²¹ Future research is needed to compare which of these methods would be most effective in reducing attrition bias in EHR data.

This study has some limitations. Our EHR data were derived from community health centers and thus the attrition rate may be different in other settings as suggested by the Medicaid data. Although our EHR-only analyses spanned 20 US states, our Medicaid data analyses are from 1 state and results may not be generalizable to other states as Medicaid eligibility and the provider mix available to patients with Medicaid varies by state. In addition, mortality data were not available so we do not know the proportion of the attrition rate due to patient death. Future research is needed to confirm EHR attrition estimates in other primary care settings and with taking mortality into account. Extensions to this work could also examine attrition patterns across different age groups, disease cohorts, and insurance type.

CONCLUSIONS

This study demonstrates the use of the EHR dataset is a reliable source of data to support longitudinal studies while highlighting variability in attrition by primary care source and chronic conditions. Our EHR data analyses suggest that community health center patient attrition rates average 33% and about 67% of patients had 1 returning visit within a 3-year period. Although these attrition rates vary by clinic location and patient chronic health needs, they appear stable over time. Findings from our analyses utilizing Medicaid data suggest that only a minority of community health center patients without an observed follow-up visit changed primary care providers and that community health center attrition rates may be largely due to patients not seeking regular care, especially among patients without documented chronic health conditions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This research was conducted with Accelerating Data Value Across a National Community Health Center Network (ADVANCE) Clinical Research Network (CRN), a member of PCORnet, an initiative originally funded by the Patient-Centered Outcomes Research Institute (PCORI) and now funded by the People-Centered Research Foundation (PCRF). The ADVANCE network is led by OCHIN in partnership with Health Choice Network, Fenway Health, Oregon Health and Science University, and the Robert Graham Center/HealthLandscape. ADVANCE is funded through PCRF contract number 1237.

Supported by Cooperative Agreement Number U18DP006116 jointly funded by the US Centers for Disease Control and Prevention and the National Institute of Diabetes and Digestive and Kidney Disease, and Patient-Centered Outcomes Research Institute. This work was also supported by the Agency for Healthcare Research and Quality, grant number R01HS025962 and by the National Cancer Institute grant number R01CA204267 and by the National Health, Lung, and Blood Institute grant number R01HL136575.

The views presented in this article are solely the responsibility of the authors and do not necessarily represent the views of the funding agencies.

REFERENCES

1. Office of the National Coordinator for Health Information Technology. Office-based Physician Electronic Health Record Adoption, Health IT Quick-Stat #50; 2019 Available at: <https://dashboard.healthit.gov/quickstats/pages/physician-ehr-adoption-trends.php>. Accessed September 13, 2019.
2. Evans RS. Electronic health records: then, now, and in the future. *Yearb Med Inform*. 2016;25(suppl 1):S48–S61.
3. Hatch B, Tillotson C, Angier H, et al. Using the electronic health record for assessment of health insurance in community health centers. *J Am Med Inform Assoc*. 2016;23:984–990. [PubMed: 26911812]
4. Marino M, Angier H, Valenzuela S, et al. Medicaid coverage accuracy in electronic health records. *Prev Med Rep*. 2018;11:297–304. [PubMed: 30116701]
5. The Henry J Kaiser Family Foundation. Summary of the Affordable Care Act. Menlo Park, CA: Henry J. Kaiser Family Foundation; 2013.
6. 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;12:14. [PubMed: 28183354]
7. Duru OK, Mangione CM, Rodriguez HP, et al. Introductory Overview of the Natural Experiments for Translation in Diabetes 2.0 (NEXT-D2) Network: Examining the Impact of US Health Policies and Practices to Prevent Diabetes and Its Complications. *Curr Diab Rep*. 2018;18:8. [PubMed: 29399715]
8. Angier H, Hoopes M, Gold R, et al. An early look at rates of uninsured safety net clinic visits after the Affordable Care Act. *Ann Fam Med*. 2015;13:10–16. [PubMed: 25583886]
9. Hoopes MJ, Angier H, Gold R, et al. Utilization of Community Health Centers in Medicaid Expansion and Nonexpansion States, 2013–2014. *J Ambul Care Manage*. 2016;39:290–298. [PubMed: 26765808]
10. National Association of Community Health Centers. Community health center chartbook; 2019 Available at: www.nachc.org/wp-content/uploads/2019/01/Community-Health-Center-Chartbook-FINAL-1.28.19.pdf. Accessed September 17, 2019.
11. Angier H, Gold R, Gallia C, et al. Variation in outcomes of quality measurement by data source. *Pediatrics*. 2014;133:e1676–e1682. [PubMed: 24864178]
12. Nichols GA, Desai J, Elston Lafata J, et al. Construction of a multisite DataLink using electronic health records for the identification, surveillance, prevention, and management of diabetes mellitus: the SUPREME-DM project. *Prev Chronic Dis*. 2012;9:E110. [PubMed: 22677160]
13. Selby JV, Lee J, Swain BE, et al. Trends in time to confirmation and recognition of new-onset hypertension, 2002–2006. *Hypertension*. 2010;56:605–611. [PubMed: 20733092]

14. Gustavson K, von Soest T, Karevold E, et al. Attrition and generalizability in longitudinal studies: findings from a 15-year population-based study and a Monte Carlo simulation study. *BMC Public Health*. 2012;12:918. [PubMed: 23107281]
15. Bjerkeset O, Nordahl HM, Larsson S, et al. A 4-year follow-up study of syndromal and sub-syndromal anxiety and depression symptoms in the general population: the HUNT study. *Soc Psychiatry Psychiatr Epidemiol*. 2008;43:192–199. [PubMed: 18064394]
16. Gucciardi E A systematic review of attrition from diabetes education services: strategies to improve attrition and retention research. *Can J Diabetes*. 2008;32:53–65.
17. Morrison F, Shubina M, Turchin A. Encounter frequency and serum glucose level, blood pressure, and cholesterol level control in patients with diabetes mellitus. *Arch Intern Med*. 2011;171:1542–1550. [PubMed: 21949161]
18. Guthmann R, Davis N, Brown M, et al. Visit frequency and hypertension. *J Clin Hypertens (Greenwich)*. 2005;7:327–332. [PubMed: 16088295]
19. Weuve J, Tchetgen Tchetgen EJ, Glymour MM, et al. Accounting for bias due to selective attrition: the example of smoking and cognitive decline. *Epidemiology*. 2012;23:119–128. [PubMed: 21989136]
20. Raykov T Analysis of longitudinal studies with missing data using covariance structure modeling with full-information maximum likelihood. *Struct Equ Modeling*. 2005;12:493–505.
21. Biering K, Hjollund NH, Frydenberg M. Using multiple imputation to deal with missing data and attrition in longitudinal studies with repeated measures of patient-reported outcomes. *Clin Epidemiol*. 2015;7:91–106. [PubMed: 25653557]

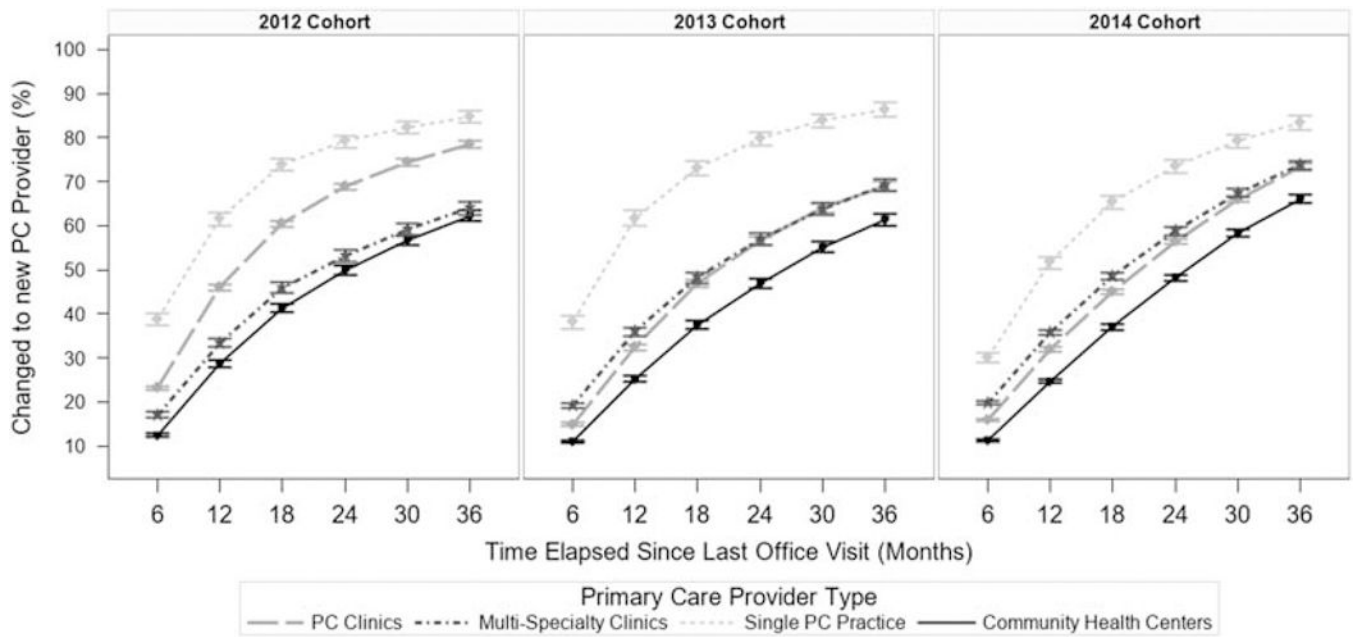


FIGURE 1. Percent of primary care (PC) patients who changed PC provider by time elapsed since their visit to their last provider, stratified by year cohort. The percent of patients who changed provider was modeled using generalized estimating equation (GEE) logistic regression. As patients could have multiple gaps at a specified length (multiple gaps of 6 mo or more between visits resulting in multiple observations per patient within each analysis), we used GEE models to account for within-patient correlation using a robust sandwich variance estimator. GEE models included PC provider type, cohort year, and the interaction term between those 2 terms. Odds ratios from GEE logistic models were transformed to estimate predicted probabilities of provider change has occurred. Models were stratified by the year in which the PC visit starting the time interval occurred.

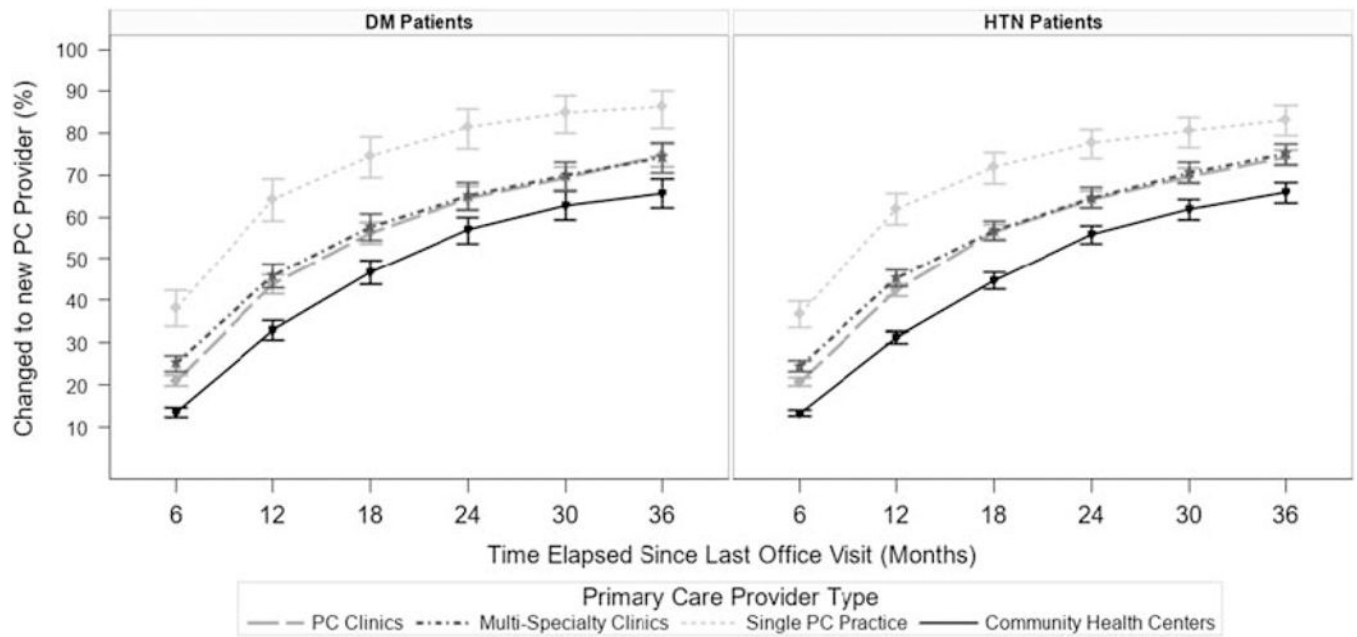


FIGURE 2.

Percent of PC patients in the 2014 cohort changed their PC provider by time elapsed since their visit to their last provider, stratified by disease diagnosis (diabetes or hypertension). The percent of patients who changed provider was modeled using GEE logistic regression. As patients could have multiple gaps at a specified length (multiple gaps of 6 mo or more between visits resulting in multiple observations per patient within each analysis), we used GEE models to account for within-patient correlation using a robust sandwich variance estimator. GEE models included PC provider type, cohort year, and the interaction term between those 2 terms. Odds ratios from GEE logistic models were transformed to estimate predicted probabilities of provider change has occurred. Models were stratified by the year in which the PC visit starting the time interval occurred. Patients with both diabetes and hypertension were included in both cohorts. Results for time intervals starting in 2014 are presented; results for 2012 and 2013 are included in Appendix 3 (Supplemental Digital Content 3, <http://links.lww.com/MLR/B970>). DM indicates diabetes mellitus; GEE, generalized estimating equation; HTN, hypertension; PC, primary care.

TABLE 1.

Study Characteristics by Data Source

Characteristics	EHR Data *	Oregon Medicaid Data †
PC provider organizations (N)		
Community health centers	76	90
PC clinics	NA	400
Multispecialty clinics	NA	56
Single-PC practice	NA	543
US States ‡	20	1
Total patients (N)	827,657	232,891
Cohorts § (N)		
2012	405,351	71,666
2013	451,970	68,943
2014	492,549	169,026
Female [n (%)]	479,964 (58.0)	126,369 (54.3)
Age, as of January 1, 2012 [mean (SD)]	39.4 (13.1)	39.3 (12.8)
Race-ethnicity [N (%)]		
Non-Hispanic white	336,776 (40.7)	162,164 (69.6)
Non-Hispanic black	144,982 (17.5)	8185 (3.5)
Non-Hispanic Other	29,896 (3.6)	15,093 (6.5)
Hispanic	241,645 (29.2)	24,649 (10.6)
No information	74,358 (9.0)	22,800 (9.8)
Primary language [n (%)]		
English	632,174 (76.4)	176,709 (75.9)
Spanish	161,943 (19.6)	6547 (2.8)
Other	22,797 (2.8)	4673 (2.0)
No information	10,743 (1.3)	44,962 (19.3)
Insurance group [n (%)]		
Some private	134,322 (16.2)	NA
Some public	397,317 (48.0)	232,891 (100)
Some private and public	63,357 (7.7)	NA
Uninsured always	232,661 (28.1)	NA
Federal poverty level ¶ [n (%)]		
138%	586,039 (70.8)	232,891 (100)
> 138%	138,042 (16.7)	NA
No information	103,576 (12.5)	NA
Urban/rural group ¶¶ [n (%)]		
Urban area	758,822 (91.7)	214,576 (92.1)
Rural area	65,252 (7.9)	17,958 (7.8)
No information	3583 (0.4)	357 (0.2)

All clinics in these samples delivered PC services.

* Characteristics of nonpregnant, adults (age 19–64), having had at least 1 ambulatory visit in 2012, 2013, or 2014, in any of 76 community health centers across 20 US states, 2012–2017.

† Patients having at least 1 claim for an office visit with a PC provider followed by a visit gap of at least 6 months while continuously enrolled. Cohort does not include patients with PC visits without 6 months visit gap.

‡ Alaska, California, Florida, Hawaii, Indiana, Kansas, Massachusetts, Maryland, Minnesota, Missouri, Montana, North Carolina, New Mexico, Nevada, Ohio, Oregon, Rhode Island, Texas, Washington, Wisconsin.

§ Cohorts are not mutually exclusive and will not sum to 100% of total patients.

∥ Last recorded federal poverty level.

¶ EHR RUCA classification code based on 2010 census tract information and patient's most recent zip code. Medicaid RUCA classification based on the patient's most recent zip code.

EHR indicates electronic health record; NA, not available; PC, primary care; RUCA, rural-urban commuting area.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 2.

Attrition and Retention in Electronic Health Record Data From 76 Community Health Centers With 340 Delivery Sites Across 20 States, Among Nonpregnant Adults Age 19–64 in the Accelerating Data Value Across a National Community Health Center Network Clinical Research Network, 2012–2017

EHR Cohorts	Attrition* [in (%)]		Retention† [in (%)]		
	No Visits in the 3-year Follow-up	Return at 1 in the 3-year Follow-up	Return in First Follow-up Year	Return in Second Follow-up Year	Return in Third Follow-up Year
2012 Cohort					
Follow-up years	2013–2015	2013–2015	2013	2014	2015
All (N = 405,351)	128,796 (31.8)	276,555 (68.2)	238,287 (58.8)	191,877 (47.3)	160,144 (39.5)
DM (N = 51,619)	9210 (17.8)	42,409 (82.2)	40,260 (78.0)	32,912 (63.8)	27,250 (52.8)
HTN (N = 136,050)	28,513 (21.0)	107,537 (79.0)	99,619 (73.2)	80,959 (59.5)	66,780 (49.1)
2013 cohort					
Follow-up years	2014–2016	2014–2016	2014	2015	2016
All (N = 451,970)	155,519 (34.4)	296,451 (65.6)	256,175 (56.7)	202,025 (44.7)	172,322 (38.1)
DM (N = 59,769)	12,251 (20.5)	47,518 (79.5)	44,969 (75.2)	35,975 (60.2)	30,550 (51.1)
HTN (N = 158,470)	37,901 (23.9)	120,569 (76.1)	111,727 (70.5)	88,752 (56.0)	75,194 (47.4)
2014 cohort					
Follow-up years	2015–2017	2015–2017	2015	2016	2017
All (N = 492,549)	169,103 (34.3)	323,446 (65.7)	277,039 (56.2)	223,533 (45.4)	192,408 (39.1)
DM (N = 67,420)	14,819 (22.0)	52,601 (78.0)	49,545 (73.5)	40,691 (60.4)	35,068 (52.0)
HTN (N = 174,633)	42,876 (24.6)	131,757 (75.4)	120,913 (69.2)	98,855 (56.6)	84,895 (48.6)

Patients with at least 1 ambulatory visit in 2012, 2013, or 2014 were included in each years' cohort. Results reported for all patients, and for those meeting diabetes mellitus (DM) or hypertension (HTN) criteria, for the 3 years following cohort-year, 2012–2017.

* Attrition is the percent of patients who *never had a visit* in the 3 years that follow their cohort-year.

† Retention is year-specific estimates of the percent of cohort patients who had at least 1 ambulatory visit in either of the follow-up years. For example, in the 2012 cohort, all patients are in the denominator for estimation of retention in 2013, 2014, and 2015. If a patient in the 2012 cohort had a visit in 2013 and 2015, the patient is in the numerator for each of those years but not in 2014 since they had no visit that year.