Using electronic health record data to evaluate chlamydia screening in community health centers based on differences in patient primary language

Author: Michael Saladik, OHSU MD/MPH Candidate, Class of 2016
Committee: Miguel Marino, PhD; Ana Quiñones, PhD, MS; Willi Horner-Johnson, PhD

Table of Contents

Acknowledgements

I am indebted to the following people for their assistance and patience throughout this thesis process: my thesis committee including Miguel Marino, PhD (research mentor, biostatistician), Ana Quiñones, PhD, MS (thesis chair), and Willi Horner-Johnson, PhD (committee member); the OCHIN/OHSU family medicine CARDIAC team, in particular Jennifer DeVoe, Seffani Bailey, Miguel Marino, John Heintzman, Megan Hoopes, Jean O'Malley, and Brigit Hatch; my MD/MPH cohort Michelle (DeChant) Barton, Jenn (Jacobs) Siegfried, Xiang Gao, Liška Havel, Maddy Grandy, Lindsey Pearson; MD/MPH department faculty and staff John Stull, William Lambert, and Natalie Chin; Career advisors Tim Joslin, and Moira Ray, as well as and my partner Monica and my family.

Abbreviations

ANOVA – analysis of variance
API – Asian and Pacific Islander
CDC – Centers for Disease Control and Prevention
EHR – electronic health record
EMR – electronic medical record
FPL – federal poverty level
HEDIS – Healthcare Effectiveness Data and Information Set
NH – Non-Hispanic
OCHIN – [formerly] Oregon Community Health Information Network
OHA – Oregon Health Authority
OHP – Oregon Health Plan
OR – odds ratio
PID – pelvic inflammatory disease
SD – standard deviation
STD – sexually transmitted disease
STI – sexually transmitted infection
USPSTF – United States Preventive Services Task Force

List of Tables and Figures

Abstract

***Background*:** Research has demonstrated the existence of health disparities based on a race, ethnicity, socioeconomic status, insurance status, and other patient characteristics. These disparities in the provision of health care remain common and, with continued research, are becoming more apparent. While there have been investigations of disparities in provision of preventive services, no study to date has used electronic health record (EHR) data to evaluate for differences in chlamydia screening on the basis of patient primary language. Chlamydia is a sexually transmitted disease that is common in young sexually-active women, is readily detectable, and simple treatment can prevent serious consequences of chronic infection. Identification of barriers to testing faced by speakers of a non-English primary language could lead to targeted interventions including greater emphasis on matching patients with a language-concordant health care provider or ensuring the availability of high-quality interpreting services. Thus, the goal of this study was to use EHR data from OCHIN community health centers to describe differences in guideline-recommended testing for chlamydia based on patient primary language.

***Methods*:** This project used a data set of 41,269 patients and 790,501 encounter records to retrospectively evaluate for differential screening for chlamydia. Of those records, there were a total of 1,788 women who had encounters in OCHIN community health centers between 2006 and 2011 and for whom chlamydia screening would have been appropriately recommended according to United States Preventive Services Task Force (USPSTF) guidelines. Demographic comparisons were performed using ANOVA, chi-square, and Kruskal-Wallis tests. Simple logistic regression models were created evaluating the relationships between the primary independent variable of interest, patient primary language (English, Spanish, Other, and Unknown), potential covariates, and the binary outcome designating chlamydia testing as having been ordered for a patient or not. Multiple logistic regression modeling was used to determine which hypothesized confounders were statistically important in describing the relationship between patient language and chlamydia screening. Model fitting and diagnostics ensured the final multiple logistic regression model appropriately fit the data, and sensitivity analyses were performed to compare the primary model to other plausible comparative models.

Results: The primary model showed non-significantly higher odds of chlamydia test ordering among primary Spanish speakers (adjusted OR 1.1666, $p = 0.559$) and statistically significantly lower odds among those primarily speaking a language other than English or Spanish (adjusted OR 0.3982, $p < 0.001$). We also found consistently and statistically significantly higher odds of chlamydia testing among minority race/ethnicity groups. Sensitivity analyses supported these findings.

***Conclusions*:** This study showed a difference in chlamydia testing based on patient primary language, particularly lower odds in those who speak a primary language other than English or Spanish. These findings may be driven by a lower likelihood of language concordant encounters or lack of access to appropriate interpreting services in those languages and suggests the need for further investigation. The higher odds of testing among minority race/ethnicity groups again highlights a need for focused research to assess whether this represents a particular focus on testing in minority groups due to greater risk factors or whether it is attributable to diagnostic testing when symptoms are present.

Background

Chlamydia trachomatis infection is a common sexually transmitted disease (STD) that affected 6,691.5 per 100,000 women under the age of 25 in 2014. [1] This number likely underestimates the true burden of chlamydia because it so frequently goes undiagnosed since genital chlamydia infections are usually asymptomatic[2], suggesting reactionary testing is not sufficient to diagnose a significant proportion of chlamydial infections. Of greater concern, ascending infection (infection starting at common sites including the vagina or cervix and "ascending" to the upper genital tract including the uterus and fallopian tubes) if untreated can lead to pelvic inflammatory disease (PID) and chronic complications such as tubo-ovarian abscess, scar tissue formation inside and outside of the fallopian tubes, chronic pelvic pain, ectopic pregnancy, and infertility. [3,4] Perinatal transmission of urogenital chlamydia infection to a neonate can also lead to ophthalmia neonatorum or infant pneumonia. [5] Yet when uncomplicated urogenital chlamydia infection is diagnosed, it can be treated and easily cured with antibiotics. The two regimens primarily recommended by the Centers for Disease Control and Prevention (CDC) have been shown to have up to a 97-98% cure rate.[5]

Infection with *Chlamydia trachomatis* has been shown to be associated with a number of different risk factors and shows significantly different prevalence based on age, gender, race/ethnicity, and other common demographics. Young age (age 24 years or younger) is considered a risk factor as nearly two-thirds of incident cases occur in youth aged 14-24 years. Any sexual activity, but in particular sex without use of barrier contraception or with multiple partners, and lower socioeconomic status have been considered risk factors.[6] It has also been shown on numerous occasions that chlamydia disproportionately affects racial and ethnic minority groups as well as the men who have sex with men (MSM) population.[7,8]

The United States Preventive Services Task Force (USPSTF) has taken note of the burden of disease and complications associated with *Chlamydia trachomatis* infection, and has determined that chlamydia is an important disease, screening tests can accurately detect chlamydia, diagnosis and subsequent treatment reduces complications of disease, and there are small to no harms associated with screening. As a result, the USPSTF has consistently recommended screening in certain subsets of the population.[9] The recommendation has gone through multiple updates since the turn of the century. In 2001, the USPSTF recommended screening for chlamydia in women regardless of pregnancy status if they are at increased risk for infection. In 2007, an updated recommendation strongly recommended (grade A[a]) screening in all non-pregnant sexually active women age 24 or younger regardless of other risk factors and screening in non-pregnant women age 25 years or older in the presence of increased risk. The 2007 recommendations included a recommendation (grade B[b]) to screen all pregnant women age 24 years or younger and pregnant women age 25 years or older who are at increased risk.[10] Most recently, an update in 2014 changed the recommendation to once again combine pregnant and non-pregnant women, giving a grade B recommendation to screening all sexually active

[a] Grade A: The USPSTF recommends this service. There is high certainty that the net benefit is substantial.[10]
[b] Grade B: The USPSTF recommends this service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial.[10]

women age 24 years or younger and women age 25 years or older who are at increased risk.[9]

Despite these relatively consistent screening recommendations, screening for chlamydia still occurs at insufficient rates. Chlamydia screening is included in the Healthcare Effectiveness Data and Information Set (HEDIS), with data collected and reported each year to track rates of chlamydia screening. Using health plan information (including Medicaid), the CDC assessed 4,131,193 women in the US in 2014 for whom chlamydia screening is recommended based on age and sexual activity, and found that 49.9% of those women were tested for chlamydia. Similarly in Oregon, 44.3% out of 34,703 were screened. [11] Though these numbers represent an increase from rates around 21% in 2001, there are still many women at risk who are not being tested.

With recent momentum toward expansion of health insurance and increasing health care access, there is increased need for research focused on cost-effective primary care and preventive medicine as a way to allay the burden on the health care system and control costs associated with providing care for greater numbers of people. In conjunction, as previously uninsured individuals gain insurance and the affordable access that insurance conveys, it is important to consider the remaining barriers to cost-effective care.

Health disparities research is one mode of describing differences in care and subsequently identifying both potential mechanisms leading to those differences and targets for intervention. Recent research has aimed to identify disparities in provision of preventive services, and has described differences based on many dimensions of social determinants of health. McMorrow et al [12] recently evaluated determinants of receipt of recommended preventive services, including in their analysis an assessment of insurance coverage, age, gender, race/ethnicity, citizenship, education level, employment status, family status, general health and mental health, as well as personal risk aversion. The study findings suggested receipt of preventive services varied widely depending on the service and income, with insurance status, education, age, and health status as other important factors. DeVoe et al [13] showed rates of preventive services differ by insurance status and having a "usual source of care", while Heintzman et al [14] confirmed insurance status affects rates of preventive services using clinical data collected from an electronic health record. Further, studies have shown differences in receipt of preventive services by race, ethnicity, language, health literacy, income, insurance status, education level, age, disability, health status, and having a usual source of care.[13,15-17]

Patient language (in particular in the US, non-English patient language) has long been considered a potential predictor of health disparities and this holds for disparities in preventive services. The association between patient language and various preventive services has been studied previously, often showing non-English speakers have lower odds of receiving a particular preventive service. Woloshin et al [18] attempted to isolate the effect of patient language by studying rates in a setting of universal access within the Canadian health care system and showed that women whose main language spoken was not English were significantly less likely to receive a breast exam or mammogram and appeared less likely to have received Pap testing (odds ratio was not significant). DeAlba et al [19] also showed that lack of

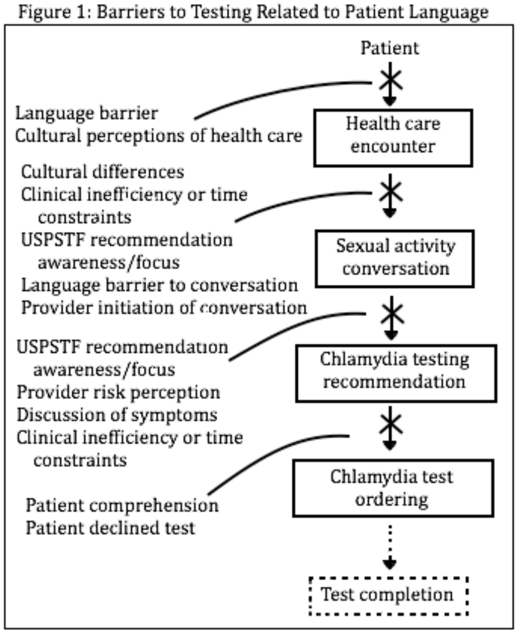
English proficiency was associated with a lower likelihood of Pap test recommendation. Johnson-Kozlow et al [20] suggested limited English proficiency was associated with lower likelihood of colorectal cancer screening.

Patient-physician language concordance (primary language agreement between patient and provider) has been evaluated as a potential association and mechanism to explain the connection between patient language and discrepancies in provision of preventive services. Two recent reports, however, failed to show patient-physician concordance was associated with differences in proportions of patients receiving mammography or influenza vaccination, and language concordance was actually associated with lower likelihood of colorectal cancer screening in studies among an Asian population[21] and a Spanish-speaking population[22] respectively.

Despite the null or inverse findings in prior studies, it remains plausible that in a US health care system, in which the vast majority of providers speak English as a primary language, a non-English primary language could be a barrier to seeking care or having a discussion about sexual practices. These barriers would be expected to reduce the odds of being appropriately screened for chlamydia. Since barriers to care and other social determinants of health have been shown to reduce the provision of preventive services, and non-English patient language specifically has been shown to reduce the provision of preventive services other than chlamydia screening in young sexually active women, we hypothesized that non-English patient language would be associated with a lower likelihood of recommended chlamydia screening.

It was therefore the goal of this investigation to search for differences in chlamydia screening in community health centers based on patient primary language. Conceptually, we believe this difference could come about in multiple ways, as diagrammed in Figure 1. Points at which a patient may fall off the pathway include the patient's non-English primary language presenting a barrier to visiting a community health center; a non-English primary language leading to language barriers with a provider who primarily speaks English, resulting in clinic inefficiencies and focus on acute patient concerns while the opportunities for routine preventive services are missed; or a non-English primary language leading to a language barrier that makes the sensitive conversation about sexual activity and chlamydia risk factors difficult and prevents proper risk assessment for the patient.

Figure 1: Barriers to Testing Related to Patient Language



Past studies looking at the association of patient language and preventive services have generally looked at survey data and insurance claims, but these modes of data collection have their weaknesses. Surveys are particularly subject to interviewer and recall bias, while claims data is likely to miss uninsured patients who frequently get their care at community health centers. The data for this study were collected in Oregon community health centers, capturing information from patients with a wide range of demographics and including uninsured patients. Also unique to this investigation is the use of electronic health record (EHR) data to focus on the association between patient language and recommended chlamydia screening. While insurance claims data allow quantification of the testing itself, the use of EHR data allows for an investigation of the appropriate recommendation and ordering of screening for patients and captures occasions in which the test may have been recommended and ordered but was not completed or was completed at a location that is not part of the OCHIN network.

Heintzman et al [23] recently investigated the agreement of Medicaid claims and EHRs as means of assessing rates or odds of USPSTF-recommended preventive services. In a comparison of 11 preventive services provided to greater than 13,000 individuals with Medicaid insurance in 43 Oregon community health centers in 2011, the group concluded EHRs represented an appropriate modality for evaluating adherence to preventive services recommendations.

The primary benefit of this evaluation of differential receipt of preventive services lies in the potential to identify non-English primary language as a risk factor for sub-optimal provision of preventive care. This study can highlight areas where further research is needed and suggest primary care interventions to focus on increasing rates of STD screening among specific groups in primary care clinics. Such interventions have the potential to be of particular importance given the association of non-English primary language and non-white race or ethnicity and higher rates of chlamydia. [24] Intervention would therefore hold two-fold benefit by targeting a population that may have both higher rates of disease and potentially lower rates of screening.

To evaluate for differential provision of preventive services in Oregon OCHIN community health centers, this study addressed the following research questions:

Research Question 1: Describe the population of 19-24 year-old females in the OCHIN database who have been designated as sexually active and are therefore appropriate for chlamydia screening. These descriptive statistics will include distributions within the total sample and within each patient primary language group of age, race and ethnicity, federal poverty level, insurance status, chronic disease burden, number of primary care encounters, and number of years in the study.

Hypothesis 1: Primary language subgroups within this population of young women will display a significantly different distribution of race and ethnicity, but no significant differences in distribution of age, income, insurance status, chronic disease, number of primary care encounters, or number of years in the study.

Research Question 2: Identify a primary model of interest and use univariable and multivariable logistic regression to estimate chlamydia screening in Oregon OCHIN community health centers based primarily on the effect of patient primary language while adjusting for confounders and effect modifiers.

Hypothesis 2: Sexually active women age 19–24 years who visited Oregon OCHIN community health centers between 2006 and 2011 were less likely to receive chlamydia screening according to USPSTF guidelines if their primary language was Spanish or Other compared to those with English as a primary language, after adjusting for confounders and considering effect modifiers.

Research Question 3: Perform sensitivity analyses using univariable and multivariable logistic regression modeling by varying the treatment of data in "Unknown" category levels, comparing the sexually active population and the population not designated as sexually active, modeling within the Hispanic race/ethnicity group, and stratifying by primary care encounters and years in the study.

Hypothesis 3: Sensitivity analyses will show the same direction of effect and similar magnitude compared to the primary multivariable logistic regression model.

Methods

Data Sources: OCHIN Electronic Medical Records

In Oregon, much of the care provided in community health centers is provided within OCHIN (formerly the Oregon Community Health Information Network, now simply known as OCHIN after the network expanded regionally and nationally) community health centers and data are centralized in one vast health record system. This database is ideal for estimating differences among primary language groups given the large nature of the medical record, the diversity of patients covered across Oregon, and the broad range of primary care clinics whose data is available in the OCHIN database. The OCHIN network of community health centers encompasses 34 safety net clinics within Oregon alone, with 46.8% of patients below 100% of federal poverty level (FPL), 34.6% of patients on Medicaid, and 33.4% self-pay. [25] In 2011, the OCHIN EHR included 515,575 registered individuals and allows extensive de-identified databases to be compiled from patient information. This centralized location is the source of demographic, commercial insurance, clinical, and laboratory information.

Data Sources: Oregon Medicaid

Oregon Medicaid enrollment data was linked to the EHR in previously published work[14] and provided information on longitudinal Medicaid coverage.

Study Population

This study includes information from clinic encounters at OCHIN community health centers from 2006 to 2011. The primary population of interest was comprised of all female patients age 19-24 years, who were designated as sexually active based on the presence at any point in the study period of one or more electronic health record codes denoting sexual activity or a pregnancy test (excluding tests within seven days prior to retinoid prescription or diagnostic imaging). A secondary population was used for sensitivity analysis and included the female patients age 19-24 years who were not designated as sexually active and therefore would not have been recommended for chlamydia screening under the USPSTF recommendations in place during the study period. Exclusion criteria included patient age > 24 years throughout the entirety of the study period. Male patients in the OCHIN EHR were excluded from this study based on the USPSTF's conclusion that evidence is insufficient to assess the harms and benefits of screening for chlamydia and gonorrhea in men, a recommendation grade I.[c]

Variables

Primary Dependent Variable: Patient Primary Language

OCHIN EHRs include a listing of patient primary language. Patient primary language has been coded as a categorical variable with four categories: English, Spanish, Other, and Unknown. The flag does not include all languages in which a patient is fluent, rather flags one single primary language. It also does not necessarily represent the language in which encounters were conducted but instead the patient's identified preferred language. A series of design variables were created to represent patient primary language, using English as the reference group.

[c] Grade I: The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined. [9]

Outcome: Chlamydia Screening

Chlamydia screening was coded using a binary variable denoting whether a chlamydia test was ordered and therefore documented in the EHR for a patient at any point during the study period.

Covariates

Multiple patient characteristics were investigated as confounders or effect modifiers. Patient demographics included age at the start of the study period (discrete variable), race/ethnicity (categorical variable including Hispanic, Non-Hispanic White, Non-Hispanic Other, Non-Hispanic Black, Non-Hispanic Asian and Pacific Islander, and Unknown), health insurance status (categorical variable including continuously uninsured, partially insured with public insurance only, partially insured with any private insurance, continuously insured with public insurance only, continuously insured with any private insurance), income category (categorical variable including income ≤ 138% FPL, income > 138% FPL, or unknown income), number of chronic diseases (discrete variable including total diseases of coronary artery disease, hypertension, diabetes mellitus, dyslipidemia, and asthma), and presence of any chronic disease (binary variable including any disease or no disease). These potential covariates were evaluated to determine the appropriateness of the inclusion in the final association model. In addition, number of primary care encounters (discrete variable) and length of time over which an individual contributed information to the study, defined by the time between first and last encounter during the study period (continuous variable), were considered for use in stratification of the primary model.

Statistical Analyses

This study compared patient demographics among the total study population, primary English speakers, Spanish speakers, and speakers of “Other” or “Unknown” languages. Comparison of descriptive statistics across primary language groups was performed with oneway ANOVA tests for normally distributed continuous variables, chi-square test of proportions for categorical variables, or Kruskal-Wallis non-parametric tests for continuous variables with skewed distributions.

In addressing differences in chlamydia screening based on primary language, crude proportions within each primary language category were determined using as a denominator the number of patients identified as sexually active (for the second sensitivity analysis the denominator was comprised of the number of patients not identified as sexually active) and between the ages of 19 and 24 at any point in the study period with the numerator being those for whom chlamydia screening was ordered. In order to describe the likelihood of chlamydia screening, the binary yes/no variable describing whether a chlamydia-screening test was ordered for each patient was used as the dependent variable, the four-level categorical variable identifying each patient’s primary language was the independent variable of interest, and other variables were included as confounders, effect modifiers, or omitted from analysis as deemed appropriate.

Covariates for inclusion in the maximum likelihood multiple logistic regression models were determined through the use of a purposeful selection method. Modeling began with simple logistic regression models fit for the patient primary language variable as well as all other potential confounders and effect modifiers. Individual covariates showing overall F-test significance at an $\alpha = 0.25$ level were kept in the model for further investigation. After this initial variable selection step, a preliminary multivariable logistic regression model was fit. At this stage, covariates were compared to a $\alpha = 0.05$ significance level, and covariates not meeting this significance cutoff were considered for removal from the model (a categorical variable was considered significant based on a priori hypothesis or $p < 0.05$ for any level of the categorical variable). After removal of nonsignificant covariates, the reduced model was compared to the larger model using a likelihood ratio test to determine if the reduced model exhibited a more favorable fit. If any variables were removed based on nonsignificance at the $\alpha = 0.05$ level, they were re-introduced individually to assess for important cofounding, using a cut-off of a 20% change in the coefficients of variables retained in the reduced model. Finally, variables that were not included in the model based on the initial simple logistic regression significance level of $\alpha = 0.25$ were re-incorporated into the reduced model one at a time and kept in the model if the added variable showed statistical significance at an $\alpha = 0.05$ level. Interaction terms were created and investigated, keeping interaction terms showing statistical significance at an $\alpha = 0.05$ level. Model diagnostics/goodness of fit procedures were carried out, successively testing the Pearson residual statistic, Deviance residual statistic, Hosmer-Lemeshow goodness-of-fit statistic, as well as using graphic assessment diagnostics for logistic regression to identify influential points and poorly fitting covariate patterns.

The primary model of interest was a multivariable logistic regression model determined using the purposeful selection method described above. This model was performed among sexually active females age 19-24 years, included patient primary language as the variable of primary interest, and included an initial covariate pool of age at the start of the study period, insurance status, income, combined race/ethnicity, number of chronic diseases, and presence of any chronic disease as possible confounders. Descriptive statistics, including distributions for various methods of categorizing race and ethnicity, were evaluated prior to determining this primary model. A race/ethnicity variable including category levels Hispanic, Non-Hispanic (NH) White, NH Other, NH Black, NH Asian and Pacific Islander, and Unknown race/ethnicity fit the study population best, thus this variable was used for this primary model and subsequent analyses. To maintain optimal power, unknown designations of language, race/ethnicity, or income were kept in the analysis as a categorical level in these particular variables.

Sensitivity analyses were performed by comparing the magnitude and direction of the estimated odds ratios of patient primary language on chlamydia testing across multiple plausible models. The first sensitivity analysis treated unknown values as missing data in order to assess the potential risk of bias arising from individuals with “unknown” primary language. A second sensitivity analysis duplicated the primary model in the population not designated as sexually active to evaluate our decision to include only women designated as sexually active in the primary model. Chlamydia tests were ordered for individuals not listed as sexually active, suggesting some proportion of individuals were likely sexually active and either denied sexual activity but were tested based on symptoms and risk factors or were not marked as sexually active in the EHR despite acknowledgement of sexual activity. Modeling within the "Hispanic" race/ethnicity category was completed to highlight differences between primary English and Spanish speakers by using descriptive statistics and logistic regression modeling within this ethnicity group.

Finally, in this longitudinal data set, an additional sensitivity analysis was performed by stratification of the primary model by number of primary care encounters and by length of time over which an individual contributed information to the study. These variables approximate the degree of health care exposure of individuals and the amount of time spent in the study, and inclusion of these variables in analysis allows our study to evaluate whether women who had more frequent clinic encounters or were captured in the data over a longer period of time were more likely to have a chlamydia test ordered. These variables were stratified rather than adjusted in logistic regression because these are temporal variables counting time or events both before and after the possible outcome, and thus would not be appropriate to consider as a true confounder. Stratification also allows for further consideration of the number of primary care encounters and the time contributing to the study as effect modifiers. Strata were created by forming tertiles, with cut-points at the 33.33 percentile and the 66.67 percentile among number of primary care encounters and years in the study period (calculated as the difference in time from each individual's first and last encounter during the study period).

Power and Sample Size Calculation

With a nearly 5:1 ratio of sexually active English to Spanish speakers, calculation suggests that with a power of 0.8 and $\alpha = 0.05$, the minimum sample size to detect a 10% difference in chlamydia screening proportions would be $n > 1{,}311$ subjects, while the ideal sample size to detect a 5% difference would be $n > 5{,}141$. With a greater than 10:1 ratio of sexually active English to Other language speakers, calculation suggests that with a power of 0.8 and $\alpha = 0.05$, the ideal sample size to detect a 10% difference in chlamydia screening rates would be $n > 2{,}406$ subjects, while the ideal sample size to detect a 5% difference would be $n > 9{,}428$.

All statistical analysis was performed using STATA 13 statistical software.

Results

Population Characteristics

The total study population included 1,788 female patients between the ages of 19 and 24 years, with 1,264 primary English speakers, 272 primary Spanish speakers, 121 who primarily spoke a language designated as “Other”, and 131 whose primary language was unknown. All demographic categories evaluated differed significantly across language groups, as shown below in Table 1.

Table 1

Population Characteristics: Total Population and Breakdown by Patient Primary Language Among 19–24 Year Old Women in Oregon OCHIN Community Health Centers in 2006–2011

Characteristics	Total	English	Spanish	Other	Unknown	*p*-value
N	1788	1264	272	121	131	
Age at start of study						
Mean (sd)	21.2 (1.6)	21.2 (1.6)	21.7 (1.7)	20.9 (1.7)	21.1 (1.6)	<0.001[a]
Race/ethnicity, Number (%)						
NH White	943 (52.7)	846 (66.9)	0 (0.0)	33 (27.3)	64 (48.9)	<0.001[b]
Hispanic	386 (21.6)	104 (8.2)	272 (100)	2 (1.7)	8 (6.1)	
NH Other	41 (2.3)	36 (2.9)	0 (0.0)	2 (1.7)	3 (2.3)	
NH Black	227 (12.7)	185 (14.6)	0 (0.0)	26 (21.5)	16 (12.2)	
NH API	91 (5.1)	34 (2.7)	0 (0.0)	51 (42.2)	6 (4.6)	
Unknown	100 (5.6)	59 (4.7)	0 (0.0)	7 (5.8)	34 (26.0)	
Income, Number (%)						
≤ 138% FPL	1367 (76.5)	985(77.9)	208(76.5)	102(84.3)	72 (55.0)	<0.001[b]
> 138% FPL	312 (17.5)	204(16.1)	64(23.5)	19(15.7)	25 (19.1)	
Unknown	109 (6.1)	75(5.9)	0(0.0)	(0.0)	34 (26.0)	
Insurance, Number (%)						
Cont uninsured	327 (18.3)	105(8.3)	176(64.7)	15(12.4)	31 (23.7)	<0.001[b]
Partially Public	815 (45.6)	607(48.0)	65(23.9)	89(73.6)	54 (41.2)	
Partially Private	358 (20.0)	294(23.3)	14(5.2)	9(7.4)	41 (31.3)	
Cont Public	240 (13.4)	220(17.4)	13(4.8)	6(5.0)	1 (0.8)	
Cont Private	48 (2.7)	38(3.0)	4(1.5)	2(1.7)	4 (3.1)	
Presence of Chronic Disease, Number (%)						
None	1478 (82.7)	999 (79.0)	250 (91.9)	116 (95.9)	113 (86.3)	<0.001[b]
Any	310 (17.3)	265 (21.0)	22 (8.1)	5 (4.1)	18 (13.7)	
Encounters						
Mean (sd)	9.0 (9.0)	10.0 (9.8)	7.5 (6.7)	7.0 (6.0)	4.2 (3.8)	<0.001[c]
Time in study, years (%)						
Mean (sd)	2.0 (1.4)	2.1 (1.4)	2.2 (1.5)	1.5 (1.3)	1.0 (1.2)	<0.001[c]

[a]Oneway ANOVA test
[b]Chi-square test
[c]Kruskal-Wallis test

Means across primary language groups in the age at the start of study ranged between 20.9±1.7 years in the Other language group and 21.7±1.7 years in the Spanish language group, and were evaluated with a one-way ANOVA test based on the normal distribution of this discrete age variable. Differences in frequencies across primary language in the race/ethnicity, income, insurance status, and chronic disease categorical variables were tested using a non-distributional chi-square test. Finally, differences of means across primary language groups among number of primary care encounters and years contributing study information were tested using a non-parametric Kruskal-Wallis test based on their right skewed distributions. Number of primary care encounters ranged from a mean of 4.2±3.8 encounters in the group with Unknown primary language compared to a mean of 10.0±9.8 primary care encounters in the English language group. Number of years in the study ranged from a mean of 1.0±1.2 years in the group with Unknown primary language compared to a mean of 2.2±1.5 in the Spanish language group.

Univariable Logistic Regression

Primarily, univariable logistic regression was used to show that in an unadjusted model, patient primary language was associated with differences in odds of chlamydia testing. In the primary population of interest, Spanish primary language was associated with 70.1% higher odds (95% CI: 26.5%–128.7%) of chlamydia testing compared to English primary language, while "Other" language was associated with 30.6% lower odds (95% CI: 52.4% lower–1.0% higher), and "Unknown" language was associated with 12.2% lower odds (95% CI: 39.2% lower to 26.9% higher) of chlamydia testing. The association of all other potential covariates with chlamydia testing was modeled using univariable logistic or linear regression as appropriate, and the resulting unadjusted odds ratios are shown below in Table 2.

Table 2
Univariable Logistic Regression: Unadjusted Odds Ratios Comparing of Any Chlamydia Testing During the Study Period (Primary Outcome), and Patient Primary Language (Independent Variable of Interest) and All Potential Covariates (*n* = 1788)

Variable	OR	95% CI (*p*)
Language		
English (ref)	---------	-------------------------------
Spanish	1.7006*	1.2647–2.2867 (<0.001)
Other	0.6936	0.4764–1.0099 (0.056)
Unknown	0.8782	0.6076–1.2691 (0.489)
Start Age	0.9835	0.9273–1.0431 (0.579)
Race/Ethnicity		
Non-Hispanic White (ref)	---------	-------------------------------
Hispanic	1.9547*	1.5071–2.5353 (<0.001)
Non-Hispanic Other	0.8238	0.4399–1.5426 (0.545)
Non-Hispanic Black	2.0730*	1.4975–2.8695 (<0.001)
Non-Hispanic API	1.9861*	1.2242–3.2222 (0.005)
Unknown	0.9431	0.6218–1.6000 (0.783)
Income		
> 138% FPL (ref)	---------	-------------------------------
≤ 138% FPL	1.1503	0.8860–1.4935 (0.293)
Unknown	0.7370	0.4966–1.0939 (0.130)
Insurance Coverage Type		
Continuously Uninsured (ref)	---------	-------------------------------
Partially-Insured, Public Only	0.8621	0.6334–1.1374 (0.294)
Partially-Insured, Any Private	0.5313*	0.3879–0.7278 (<0.001)
Continuously-Insured, Public Only	0.6985*	0.4916–0.9925 (0.045)
Continuously-Insured, Any Private	1.0545	0.5420–2.0517 (0.876)
Number of Chronic Diseases	1.1296	0.9076–1.4060 (0.275)
Presence of Any Chronic Disease		
No Chronic Disease (ref)	---------	-------------------------------
Any Chronic Disease	1.1843	0.9131–1.5361 (0.202)
# of Primary Care Encounters	1.0540*	1.0391–1.0690 (<0.001)
# of Years in Study	1.3090*	1.2198–1.4046 (<0.001)

*Statistically significant at $\alpha < 0.05$ level

Multivariable Logistic Regression – Primary Model

In the adjusted primary multivariable logistic regression model, Spanish language remained positively associated with chlamydia testing, though this association was not statistically significant.

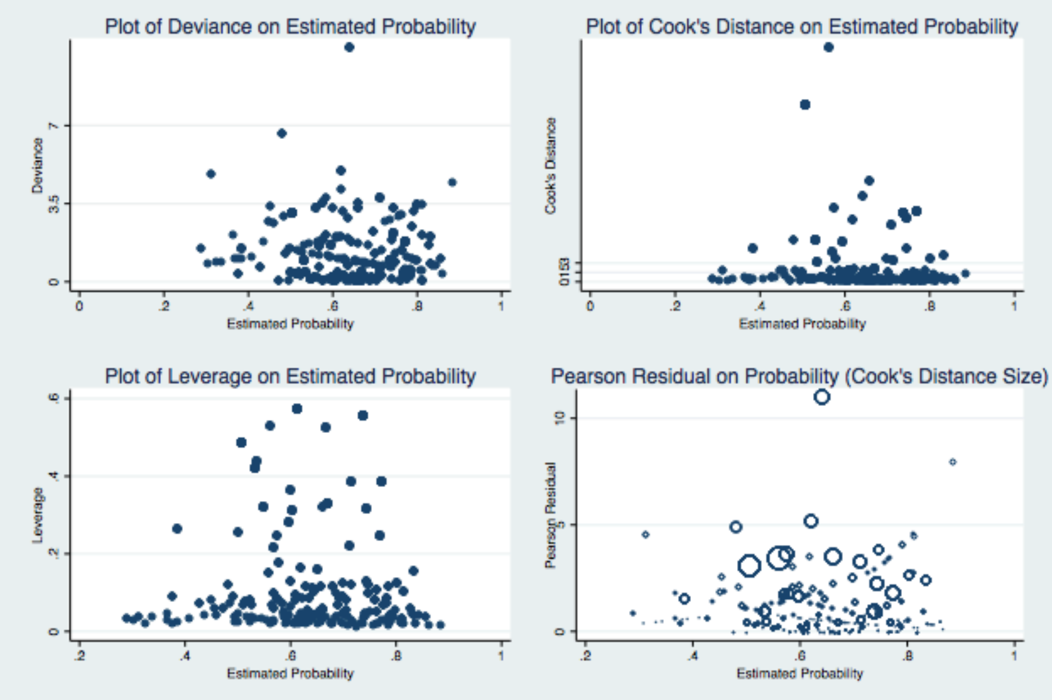
Table 3
Primary Model: Association Between Chlamydia Testing and Patient Primary Language Among Sexually Active Females Using Univariable and Multivariable Logistic Regression, Unknown Data Treated as a Categorical Level (*n* = 1788)

Variable	% with Ordered Tests	Unadjusted OR	Adjusted OR+	Adjusted OR 95% CI (*p*)
Language				
English (ref)	63.37	---------	---------	-----------------------------
Spanish	74.63	1.7006*	1.1666	0.6959–1.9556 (0.559)
Other	54.55	0.6936	0.3982*	0.2526–0.6277 (<0.001)
Unknown	60.31	0.8782	0.8691	0.5814–1.2992 (0.494)
Race/Ethnicity				
NH White (ref)			---------	-----------------------------
Hispanic			1.5940*	1.0409–2.4412 (0.032)
NH Other			0.8138	0.4313–1.5357 (0.525)
NH Black			2.1587*	1.5347–3.0366 (<0.001)
NH API			3.2158*	1.8188–5.6859 (<0.001)
Unknown			0.9585	0.5814–1.2992 (0.494)
Income				
> 138% FPL (ref)			---------	-----------------------------
≤ 138% FPL			1.1891	0.9030–1.5650 (0.217)
Unknown			1.1285	0.7226–1.7626 (0.595)
Insurance Coverage Type				
Cont.-Uninsured (ref)			---------	-----------------------------
Part.-Insured, Public Only			1.0421	0.7547–1.4391 (0.802)
Part.-Insured, Any Private			0.6776*	0.4702–0.9766 (0.037)
Cont.-Insured, Public Only			0.7549	0.5065–1.1251 (0.167)
Cont.-Insured, Any Private			1.2900	0.6437–2.5851 (0.473)
Presence of Any Chronic Disease				
No Chronic Disease (ref)			---------	-----------------------------
Any Chronic Disease			1.2907	0.9846–1.6921 (0.065)

*Statistically significant at α < 0.05 level
+Adjusted for all variables listed

A primary Spanish speaker had 16.66% higher odds (OR 1.1666, 95% CI: 0.6959–1.9556) of being tested during the study period than a primary English speaker. Those for whom primary language was designated as "Other" had 60.18% lower odds (OR 0.3982, 95% CI: 0.2526–0.6277) of being tested for chlamydia, while those with Unknown language had 13.09% (OR 0.8691, 95% CI: 0.5814–1.2992) lower odds of being tested. Among other factors, Hispanic, non-Hispanic black, and non-Hispanic API race/ethnicity were statistically significantly and positively associated with chlamydia testing while partial coverage with private insurance was significantly negatively associated with testing. No interaction terms were statistically significant. Model diagnostics suggested this multivariable logistic regression model displayed good overall fit based on the Hosmer-Lemeshow statistic of 9.14 with 7 degrees of freedom (p = 0.2427), the graphic representations below in Figure 2, and a lack of any identified influential points or poorly fitting covariate patterns.

Figure 2: Model Diagnostic Graphic Assessments



Multivariable Logistic Regression – Sensitivity Analysis

The first model used for sensitivity analysis included a purposeful variable selection process independent of the primary model as well as treatment of the "Unknown" categories as missing data rather than a categorical level. Income was neither statistically significant nor an important confounder, but all other primary model variables were statistically important. In Table 4, the odd ratios comparing chlamydia testing in speakers of Spanish and Other languages to English speakers retained the same direction and similar magnitude as the primary model. The odds ratios for all other variables in this sensitivity analysis also maintained the same direction and similar magnitude compared to their respective reference groups.

Table 4

Association Between Chlamydia Testing and Patient Primary Language Among Sexually Active Females Using Univariable and Multivariable Logistic Regression, Unknown Data Treated as Missing Data (n = 1591)

Variable	% with Ordered Tests	Unadjusted OR	Adjusted OR+	Adjusted OR 95% CI (p)
Language				
English (ref)	63.37	---------	---------	-----------------------------
Spanish	74.63	1.7006*	1.1663	0.6825–1.9930 (0.574)
Other	54.55	0.6936	0.4289*	0.2661–0.6913 (0.001)
Race/Ethnicity				
NH White (ref)			---------	-----------------------------
Hispanic			1.2834*	1.1139–2.7236 (0.015)
NH Other			0.8693	0.4491–1.6829 (0.678)
NH Black			2.4009*	1.6788–3.4336 (<0.001)
NH API			2.9865*	1.6599–5.3731 (<0.001)
Insurance Coverage Type				
Cont.-Uninsured (ref)			---------	-----------------------------
Part.-Insured, Public Only			1.0958	0.7746–1.5503 (0.605)
Part.-Insured, Any Private			0.7979	0.5402–1.1786 (0.257)
Cont.-Insured, Public Only			0.8467	0.5555–1.2905 (0.439)
Cont.-Insured, Any Private			1.2362	0.6035–2.5322 (0.562)
Presence of Any Chronic Disease				
No Chronic Disease (ref)			---------	-----------------------------
Any Chronic Disease			1.2834	0.9649–1.7071 (0.086)

*Statistically significant at $\alpha < 0.05$ level
+Adjusted for all variables listed

The second sensitivity analysis retained the variables determined to be statistically significant or important confounders in the primary model, but modeled within the group that was not designated as sexually active in the EHR, rather than the sexually active group. As seen in Table 5, the odd ratios comparing chlamydia testing in speakers of Spanish and other languages to English speakers retained the same direction but with magnitude greater than from the primary model, while speakers of "Other" languages continued to display significantly lower odds of testing compared to English speakers, and speakers of Unknown language had nonsignificantly lower odds.

Table 5

Association Between Chlamydia Testing and Primary Language Among Females Not Designated as Sexually Active Using Univariable and Multivariable Logistic Regression, Unknown Data Treated as a Categorical Level (n = 1297)

Variable	% Tested	Unadjusted OR	Adjusted OR+	Adjusted OR 95% CI (p)
Language				
English (ref)	20.66	---------	---------	------------------------------
Spanish	33.33	1.9200*	1.8653	0.8130–4.2794 (0.141)
Other	12.94	0.5707*	0.5498*	0.3161–0.9562 (0.034)
Unknown	13.04	0.5759	0.7388	0.4058–1.3450 (0.322)

*Statistically significant at $\alpha < 0.05$ level

+Adjusted for race/ethnicity, income, insurance coverage type, and presence of any chronic disease

The data was subsequently modeled among those with Hispanic race/ethnicity, primarily to assess differences between English and Spanish speakers of Hispanic race/ethnicity. As seen in Table 6, after adjusting for income, insurance status, and chronic disease, test ordering among Spanish speakers showed a nonsignificant 3.80% (OR 0.9620, 95% CI: 0.5282–1.7520) lower odds compared to English speakers. Though nonsignificant, the odds of chlamydia testing were 72.11% (OR 0.2789, 95% CI: 0.0162–4.7997) lower among speakers of Other languages compared to English speakers, while those with Unknown language had 46.99% lower odds (OR 0.5301, 95% CI: 0.1143–2.4586) of being tested.

Table 6
Association Between Chlamydia Testing and Primary Language Among Sexually Active Hispanic Females Using Univariable and Multivariable Logistic Regression, Unknown Data Treated as a Categorical Level (*n* = 386)

Variable	% with Ordered Tests	Unadjusted OR	Adjusted OR+	Adjusted OR 95% CI (*p*)
Language				
English (ref)	72.12	---------	---------	------------------------------
Spanish	74.63	1.1376	0.9620	0.5282–1.7520 (0.899)
Other	50.00	0.3867	0.2789	0.0162–4.7997 (0.379)
Unknown	50.00	0.3867	0.5301	0.1143–2.4586 (0.417)

§This analysis included only 2 Other language speakers and 8 Unknown language speakers. The adjusted OR comparing odds of test ordering in Spanish compared to English speakers was stable when these 10 individuals were removed from analysis (change of 2.2%), thus the full Hispanic group (*n* = 386) was included.
+Adjusted for race/ethnicity, income, insurance coverage type, and presence of any chronic disease

In order to evaluate the role of increasing contact with community health centers on the odds of chlamydia testing, the primary model was estimated and stratified by number of primary care encounters over the study period and number of years in the study. Tertiles were formed based on the 33.33 and 66.67 percentiles. These percentiles among primary care encounters were calculated to be 4 and 10 encounters respectively and fit clinically relevant cut-points used in prior research. [14] The calculated percentiles for the years in study variable were 1.00 and 2.93 suggesting clinically relevant cut-points of 1 and 3 years. The results of these stratified analyses are shown below in Table 7 and Table 8.

While the direction and magnitude of the association between Spanish or Unknown primary language and chlamydia testing changed across both the primary care encounter and time in study strata, the stratified models consistently showed lower odds of chlamydia testing in patients speaking languages Other than English or Spanish.

Table 7

Association Between Chlamydia Testing and Primary Language Among Sexually Active Females Using Univariable and Multivariable Logistic Regression, Primary Model Stratified by Primary Care Encounters (*n* = 1788)

Variable	% with Ordered Tests	Unadjusted OR	Adjusted OR+	Adjusted OR 95% CI (*p*)
≤ 4 encounters, *n* = 695				
English (ref)	49.66	---------	---------	-----------------------------
Spanish	66.67	2.0278*	1.2844	0.6188–2.6657 (0.502)
Other	43.14	0.7692	0.4525*	0.2257–0.9075 (0.026)
Unknown	55.06	1.2420	1.0901	0.6584–1.8049 (0.737)
4.001–10 encounters, *n* = 550				
English (ref)	69.15	---------	---------	-----------------------------
Spanish	78.57	1.6357	0.8185	0.2882–2.3241 (0.707)
Other	53.33	0.5098*	0.2740*	0.1184–0.6345 (0.003)
Unknown	71.88	1.1400	1.1807	0.4946–2.8185 (0.708)
> 10 encounters, *n* = 543				
English (ref)	71.82	---------	---------	-----------------------------
Spanish	83.82	2.0334*	0.9996	0.3157–3.1652 (0.999)
Other	80.00	1.5696	0.7761	0.2510–2.3995 (0.660)
Unknown	70.00	0.9156	0.6919	0.1564–3.0608 (0.627)

*Statistically significant at α < 0.05 level

+Adjusted for race/ethnicity, income, insurance coverage type, and presence of any chronic disease

Table 8

Association Between Chlamydia Testing and Primary Language Among Sexually Active Females Using Univariable and Multivariable Logistic Regression, Primary Model Stratified by Time in Study (years) (*n* = 1788)

Variable	% with Ordered Tests	Unadjusted OR	Adjusted OR+	Adjusted OR 95% CI (*p*)
≤ 1 year, *n* = 596				
English (ref)	50.40	---------	---------	-----------------------------
Spanish	62.03	1.6077	1.0235	0.4239–2.4713 (0.959)
Other	45.45	0.8202	0.4447*	0.2112–0.9363 (0.033)
Unknown	56.63	1.2850	1.1573	0.6852–1.9545 (0.585)
1.001–2.999 years, *n* = 621				
English (ref)	66.59	---------	---------	-----------------------------
Spanish	82.02	2.2889*	1.8712	0.7393–4.7359 (0.186)
Other	56.52	0.6522	0.3244*	0.1495–0.7039 (0.004)
Unknown	64.86	0.9262	0.9242	0.4277–1.9972 (0.841)
≥ 3 years, *n* = 571				
English (ref)	71.33	---------	---------	-----------------------------
Spanish	77.88	1.4155	0.7143	0.2662–1.9165 (0.504)
Other	75.00	1.2058	0.7466	0.2393–2.3297 (0.615)
Unknown	72.73	1.0718	0.9523	0.2314–3.9187 (0.946)

*Statistically significant at α < 0.05 level

+Adjusted for race/ethnicity, income, insurance coverage type, and presence of any chronic disease

Discussion

We hypothesized that our primary model would show the odds of chlamydia test ordering for sexually active women in OCHIN community health centers age 19–24 years would be lower if their primary language was Spanish or Other when compared to women with English as their primary language. Interestingly, our hypothesis does not hold for Spanish speakers but appears to hold for non-English/non-Spanish ("Other" language) speakers. Consistent with prior studies comparing provision of preventive services based on patient language, we hypothesized that patients speaking a primary language other than English would have lower odds of chlamydia testing. However, primary Spanish speakers were most often shown to have higher odds or no significant difference in odds of testing when compared to primary speakers of English. In unadjusted analyses, Spanish speakers had significantly higher odds of a test being ordered. In the primary multivariable logistic regression model, the association was in the same direction but was no longer significant after adjustment for race/ethnicity, income, insurance status, and presence of any chronic disease. By contrast, our results support our hypothesis among primary speakers of languages designated as "Other" in the OCHIN electronic health record. This group had lower odds of testing compared to primary English speakers consistently and most often statistically significantly across the primary model and all sensitivity analyses. It is important to note that the demographic breakdown of the "Other" language group appears different based on demographics, comprised of a mix of NH white, NH black, NH API race/ethnicity groups. This group includes the greatest proportions with incomes

≤ 138% FPL and partial public insurance, but the least documented chronic disease, fewer encounters and less time in the study than Spanish or English speakers. Still, differences in odds remain after adjustment or stratification for these confounders.

There are multiple possible explanations for the results of this study. First, the statistically null findings comparing chlamydia testing in English and Spanish speakers may be a true null relationship, while speakers of "Other" languages have statistically significantly reduced odds of chlamydia testing. Sensitivity analysis modeling the association between patient language and chlamydia testing within only the Hispanic ethnicity group suggests that the odds of chlamydia testing among Spanish speakers are in fact not different from the odds among English speakers. This model run within one single race/ethnicity group with two predominant primary languages (English and Spanish) likely comes closest to isolating the effect of language differences in chlamydia testing between these two language groups. There likely remain unmeasured differences, in particular among Hispanic race/ethnicity groups. Patient primary language may parallel degree of acculturation, especially within such a narrow age range. First or second generation Hispanic immigrants may hold different values and health-seeking behavior on health care, testing, and STDs. However, the adjusted odds ratio of 0.962 suggests there is no difference in odds of chlamydia testing between primarily Spanish and primarily English speaking people of Hispanic race/ethnicity. While this analysis was underpowered, with a sample size of only 386, 3.8% difference in odds would not likely be significant in a larger sample (a sample size of 879 would be required to show a 10% difference), nor would it be clinically relevant if statistically significant.

One plausible mechanism driving these relationships could be language concordance. Given the prevalence of Spanish language speakers and the availability of quality Spanish language medical interpreters in Oregon, it is plausible to hypothesize that Spanish speakers have greater access to language concordant health care providers, in-person medical interpreting, and ubiquitous phone interpreting, while primary speakers of Other languages would be less likely to have a language concordant health care provider, ready access to in-person interpreting, may not have phone interpreting in a particular language or dialect, may have untrained family members serve as interpreters, or may simply conduct an appointment in English despite suboptimal English proficiency.

Another mechanism that could explain part of the difference in odds between primary English and "Other" language speakers was suggested in the stratified models. While the odds of chlamydia testing among speakers of "Other" languages was statistically significantly lower than the odds of testing among English speakers in the lower two tertiles of primary care encounters and time in the study, the odds were only non-significantly lower in the upper tertile. This finding is likely the result of the small number of Other language speakers in these upper tertiles and the corresponding wide confidence intervals. However, the small number of Other language speakers in these upper tertiles also suggests a lesser degree of health care exposure among this group. Thus, while it appears that speaking a non-English, non-Spanish primary language is associated with lower odds of chlamydia testing compared to primary English speakers, this group also had fewer individuals in the highest tertiles of primary care encounters and number of years contributing study information. The language barriers associated with speaking these "Other" languages may alone explain the lower odds of testing, but the lower health care exposure alone could also be at least a partial driver of the lower odds of testing.

Second, the primary model could be underpowered to detect a significant difference between the Spanish and English language groups and the true association in fact may be greater odds of chlamydia testing among the primary Spanish speaking population as the unadjusted model and trends in the primary model and initial sensitivity analyses suggest. Power and sample size calculations suggested that this study was powered to detect a 10% proportional difference, but detection of a 5% difference would require a sample size in excess of 5,141. There are plausible mechanisms that could explain higher odds of testing in this group: (1) It is possible there were concerted efforts to increase rates of recommended chlamydia screening among clinics that treat a large proportion of Spanish-speaking patients; (2) Testing in this data set may have been ordered based on symptoms suggestive of chlamydia infection or other STDs. Previous research in the Pacific Northwest has shown higher odds of chlamydia test positivity among Black, American Indian and Alaska Native (AI/AN), Asian and Pacific Islander, and Hispanic women when compared to white women. [8] CDC statistics in 2014 showed higher prevalence of chlamydia again among Black, AI/AN, Pacific Islander, and Hispanic women compared to white women. [1,2] It is possible that Spanish-speaking patients reported such symptoms at rates greater than those who speak primarily

English due to greater burden of disease among those of Hispanic race/ethnicity and therefore also likely among primary Spanish speakers in this population, while language barriers may have limited report of these symptoms in the "Other" language group, leading to lower odds of testing. Conversely, English speakers (potentially based on a lack of language barriers) may have had a more comprehensive discussion of sexual activity and risk factors and declined appropriate screening recommendations; (3) Screening recommendations may have been made not on the blanket USPSTF recommendations to test sexually active women 24 years old and younger, but rather based on racial/ethnic demographics and evidence of greater chlamydia burden among Black, Hispanic, and other minority race/ethnicity groups described above[1,2,8], or based on provider perceptions of lifestyle and STD risk in minority race/ethnicity groups. Findings in this study lend greater support to the interpretation that testing is based on reported symptoms or on true or perceived risk among minority race/ethnicity groups, given the significantly and persistently elevated odds of chlamydia testing among Hispanic, non-Hispanic Black, and non-Hispanic Asian and Pacific Islander race/ethnicity groups.

Finally, residual or unknown confounding must be considered as a possible explanation of the negative association between the "Other" primary language and chlamydia testing. It is possible that the presence of an unknown or unmeasured confounder could explain this relationship, however this seems unlikely given the wide range of confounders considered and overlap of confounders with prior studies. Chlamydia screening can occur during routine prenatal care thus pregnancy could be an important unmeasured confounder. Our data structure, however, did not allow for assessment of the temporal relationship between pregnancy and the encounter during which chlamydia screening was documented, thus it was not included in the models. Pregnancy, however, was not associated with language or race/ethnicity in this dataset, thus it is unlikely to be a statistically important confounder and alter the results in any meaningful way. Age was hypothesized to be a potentially important confounder, but did not prove to be statistically important in the primary model variable selection process (most likely because of the narrow age range in this study), thus was not retained as a covariate in the models.

This retrospective, observational study is susceptible to bias. Fortunately, the outcome is definitive in the sense that there is little potential for error when measuring whether a chlamydia screen was ordered in the EHR or not. Of greatest concern in this study is under-reporting of patient sexual activity. Though the proportion of individuals not designated as sexually active with ordered chlamydia tests was much lower than the proportion among sexually-active individuals, the non-zero proportion suggests there is some number of sexually-active individuals who were not appropriately designated as such. This under-reporting of sexual activity was most likely a nondifferential bias given the sensitivity analysis model among those not designated as sexually active and the similar direction and magnitude of findings compared to the primary model. Misclassification or recall bias was unlikely in the measurement of the covariates as the covariates either do not require recall (primary language, race/ethnicity) or are measurable in the EHR without subjective input from patient or provider.

The findings in this study have numerous public health and clinical implications. Lower odds of chlamydia testing among those who speak a primary language other than English or Spanish suggest a persistent barrier to an important recommended preventive service that can detect otherwise asymptomatic disease that has serious long-term consequences. The lower odds of testing among those in the "Other" language group, but null findings in the Spanish language group suggests that there may be a lack of access to language-concordant encounters or effective and efficient translation that allows comprehensive discussions with young non-English, non-Spanish speaking women about sexual activity and appropriate risk stratification. Higher odds of testing seen among multiple minority race/ethnicity groups could be viewed one of two ways: either this represents appropriate testing among groups previously identified to have greater burden of disease or inappropriate under-screening among sexually active non-Hispanic white women who are considered at enough risk for USPSTF blanket screening recommendations. The overall proportions in this sample suggest 45–85% of those for whom testing is recommended are having these recommended tests ordered in the EHR. However, an unknown subset is being tested in response to symptoms concerning for an STD, thus the proportion tested among asymptomatic patients for whom screening is still recommended could be significantly lower. All these findings suggest the need for increased focus on chlamydia screening, particularly among young sexually-active females who speak a language other than English or Spanish and young sexually-active non-Hispanic white women.

One strength of this study is the use of EHR data from a robust centralized electronic health record that includes a wide range of community health centers in Oregon. This incorporates clinical and demographic information for a population that is actually more diverse than Oregon as a whole, a benefit in a study aiming to address disparities based primarily on patient language, but also considering race/ethnicity and federal poverty level. The use of the OCHIN EHR database also afforded the ability to consider a wide range of potential confounders in this complex relationship between language and chlamydia testing and to incorporate multiple years of data.

There were also limitations faced in this investigation. The chlamydia testing documented in this study is not necessarily representative of chlamydia tests ordered for "screening" purposes. Based on the data pulled from the electronic health record, it is not known whether individual tests were triggered by age and sexual activity alone (screening) versus patient report of symptoms (diagnostic testing). However, considering chlamydia testing is indicated for all sexually active women age 24 years and younger, testing based on symptoms is also valuable information and would preclude the need for further screening in that individual barring new or persistent risk factors over time. Tests in this EHR also indicate ordered tests but it is possible not all were collected and resulted, thus this study assesses chlamydia testing from the perspective of the health care provider recommending and ordering the test, but does not indicate the proportion of patients who had chlamydia tests resulted as positive or negative. However, this has the potential to capture information about recommendations even if tests were completed outside the OCHIN network.

OCHIN also does not encompass all possible testing sites, thus this dataset does not include chlamydia tests that may have been recorded in other health systems in the same area, and therefore does not include Planned Parenthood and other organizations that routinely provide STD testing, often for patients with similar socioeconomic status to OCHIN community health centers. There are limitations in the availability of data as well. As described previously, the data structure did not allow for an assessment of pregnancy, definitive language concordance, language of encounter, or presence and mode of interpreting services.

Finally, despite the large OCHIN database, the sample size was limited to the small proportion of the population for whom chlamydia screening is recommended (sexually active women age 24 years and younger). This limited the ability to carry out adequately powered sub-analyses.

Future research to continue this investigation would involve a study focusing on language concordance between patient and provider and include translation scenarios. This alternative approach could provide further clarification of the mechanism by which differences in odds of chlamydia testing arise, allowing for testing of the hypothesis that it is in fact different primary languages spoken by patient and provider that leads to the hypothesized differences in rates of recommended chlamydia testing. These future studies could also investigate differences in the methods of overcoming patient-provider language mismatch including visits that (1) are untranslated without patient and provider language concordance, (2) use phone interpretation, (3) use video interpretation, or (4) use in-person interpreting. Ideally, future research would also match patients in the OCHIN system with health records of other clinical entities that routinely perform STD screening. This approach could better capture all chlamydia testing and assess for differences in chlamydia testing based on a secondary source of care.

Conclusions

In conclusion, this study showed significantly lower odds of chlamydia testing among non-English, non-Spanish speaking women age 19–24 years. Analyses also showed higher odds of testing among minority race/ethnicity groups. Further research is needed to evaluate language concordance or modes of interpreting as the driver of these differences. Overall proportions of individuals tested vary widely and proportions screened may be even lower than those we observed. While greater emphasis on testing in all groups is indicated, a focus on improving efforts on chlamydia testing for those speaking non-English, non-Spanish languages may be particularly warranted.

References:

1. CDC – chlamydia statistics. http://www.cdc.gov/std/chlamydia/stats.htm. Accessed 4/28/2016, 2016.

2. Prevalence of chlamydia trachomatis genital infection among persons aged 14–39 years — united states, 2007–2012. http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6338a3.htm. Accessed 4/28/2016, 2016.

3. Detailed STD facts - pelvic inflammatory disease - CDC fact sheet. http://www.cdc.gov/std/pid/stdfact-pid-detailed.htm. Accessed 4/28/2016, 2016.

4. Pelvic inflammatory disease - CDC fact sheet. http://www.cdc.gov/std/pid/stdfact-pid.htm. Accessed 4/28/2016, 2016.

5. Chlamydial infections - 2015 STD treatment guidelines. http://www.cdc.gov/std/tg2015/chlamydia.htm. Accessed 4/28/2016, 2016.

6. Mishori R, McClaskey EL, WinklerPrins VJ. Chlamydia trachomatis infections: Screening, diagnosis, and management. *Am Fam Physician*. 2012;86(12):1127-1132.

7. Detailed STD facts - chlamydia. http://www.cdc.gov/std/chlamydia/stdfact-chlamydia-detailed.htm. Accessed 4/28/2016, 2016.

8. Fine D, Thomas KK, Nakatsukasa-Ono W, Marrazzo J. Chlamydia positivity in women screened in family planning clinics: Racial/ethnic differences and trends in the northwest U.S., 1997-2006. *Public Health Rep*. 2012;127(1):38-51.

9. Final recommendation statement: Chlamydia and gonorrhea: Screening - US preventive services task force. http://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/chlamydia-and-gonorrhea-screening. Updated December 2014. Accessed 4/28/2016, 2016.

10. Screening for chlamydial infection: U.S. preventive services task force recommendation statement. *Ann Intern Med*. 2007;147(2):128.

11. Chlamydia screening percentages. http://www.cdc.gov/std/chlamydia/chlamydia-screening-2014.htm. Accessed 4/28/2016, 2016.

12. McMorrow S, Kenney GM, Goin D. Determinants of receipt of recommended preventive services: Implications for the affordable care act. *Am J Public Health*. 2014;104(12):2392-2399.

13. DeVoe JE, Fryer GE, Phillips R, Green L. Receipt of preventive care among adults: Insurance status and usual source of care. *Am J Public Health*. 2003;93(5):786-791.

14. Heintzman J, Marino M, Hoopes M, et al. Using electronic health record data to evaluate preventive service utilization among uninsured safety net patients. *Prev Med*. 2014;67:306-310.

15. Brown DW, Shepard D, Giles WH, Greenlund KJ, Croft JB. Racial differences in the use of aspirin: An important tool for preventing heart disease and stroke. *Ethn Dis*. 2005;15(4):620-626.

16. Mainous A, Diaz VA, Koopman RJ, Everett CJ. Quality of care for hispanic adults with diabetes. *FAMILY MEDICINE-KANSAS CITY-*. 2007;39(5):351.

17. Horner-Johnson W, Dobbertin K, Andresen EM, Iezzoni LI. Breast and cervical cancer screening disparities associated with disability severity. *Womens Health Issues*. 2014;24(1):e147-e153.

18. Woloshin S, Schwartz LM, Katz SJ, Welch HG. Is language a barrier to the use of preventive services? *Journal of General Internal Medicine*. 1997;12(8):472 477.

19. De Alba I, Sweningson JM. English proficiency and physicians' recommendation of pap smears among hispanics. *Cancer Detect Prev*. 2006;30(3):292-296.

20. Johnson-Kozlow M, Roussos S, Rovniak L, Hovell M. Colorectal cancer test use among californians of mexican origin: Influence of language barriers. *Ethn Dis*. 2009;19(3):315-322.

21. Jih J, Vittinghoff E, Fernandez A. Patient-physician language concordance and use of preventive care services among limited english proficient latinos and asians. *Public Health Rep*. 2015;130(2).

22. Eamranond PP, Davis RB, Phillips RS, Wee CC. Patient-physician language concordance and primary care screening among spanish-speaking patients. *Med Care*. 2011;49(7):668-672.

23. Heintzman J, Bailey SR, Hoopes MJ, et al. Agreement of medicaid claims and electronic health records for assessing preventive care quality among adults. *J Am Med Inform Assoc*. 2014;21(4):720-724.

24. Eagar RM, Beach RK, Davidson AJ, Judson FN. Epidemiologic and clinical factors of chlamydia trachomatis in black, hispanic and white female adolescents. *West J Med*. 1985;143(1):37-41.

25. 2013 annual report 140424 - 2013-annual-report-140424.pdf. https://ochin.org/media/2013-Annual-Report-140424.pdf. Accessed 4/28/2016, 2016.