# PLOS ONE

# Quantifying rural disparity in healthcare utilization in the United States: Analysis of a large midwestern healthcare system

Akua Nuako[1], Jingxia Liu[2,3], Giang Pham[1], Nina Smock[1], Aimee James[2,3], Timothy Baker[4], Laura Bierut[1], Graham Colditz[2,3], Li-Shiun Chen[1,2]*

1 Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, United States America, 2 Alvin J. Siteman Cancer Center at Barnes-Jewish Hospital, Washington University School of Medicine, St. Louis, MO, United States America, 3 Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI, United States America, 4 Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, MO, United States America

* li-shiun@wustl.edu

## Abstract

### Purpose

The objective of this study is to identify how predisposing characteristics, enabling factors, and health needs are jointly and individually associated with epidemiological patterns of outpatient healthcare utilization for patients who already interact and engage with a large healthcare system.

### Methods

We retrospectively analyzed electronic medical record data from 1,423,166 outpatient clinic visits from 474,674 patients in a large healthcare system from June 2018-March 2019. We evaluated patients who exclusively visited rural clinics versus patients who exclusively visited urban clinics using Chi-square tests and the generalized estimating equation Poisson regression methodology. The outcome was healthcare use defined by the number of outpatient visits to clinics within the healthcare system and independent variables included age, gender, race, ethnicity, smoking status, health status, and rural or urban clinic location. Supplementary analyses were conducted observing healthcare use patterns within rural and urban clinics separately and within primary care and specialty clinics separately.

### Findings

Patients in rural clinics vs. urban clinics had worse health status [$\chi^2 = 935.1$, df = 3, p<0.0001]. Additionally, patients in rural clinics had lower healthcare utilization than patients in urban clinics, adjusting for age, race, ethnicity, gender, smoking, and health status [2.49 vs. 3.18 visits, RR = 0.61, 95%CI = (0.55,0.68), p<0.0001]. Further, patients in rural clinics had lower utilization for both primary care and specialty care visits.

## Conclusions

Within the large healthcare system, patients in rural clinics had lower outpatient healthcare utilization compared to their urban counterparts despite having potentially elevated health needs reflected by a higher number of unique health diagnoses documented in their electronic health records after adjusting for multiple factors. This work can inform future studies exploring the roots and ramifications of rural-urban healthcare utilization differences and rural healthcare disparities.

## Introduction

Rural populations, who make up around ~19% of the United States' population [1, 2], have heightened mortality rates from several chronic diseases relative to urban populations [2]. The rural-urban mortality disparity has grown over recent decades despite overall declines in mortality from such conditions [3]. The causes of rural health outcome disparities are multiple and complex, but importantly include higher rates of poverty and behavioral risk factors, larger proportions of elderly individuals and individuals with disabilities, lower rates of insurance and education, and limited access to healthcare providers and healthcare encounters in rural regions [4–8]. Given rural communities' amplified barriers to acquiring healthcare, further quantification of differences in healthcare utilization across rural and urban populations using healthcare systems data can potentially provide important insight into rural health inequities among patients engaged in healthcare systems.

Healthcare utilization is a broad term that describes individual or population-level use of healthcare services delivered in outpatient facilities, hospitals, or homes [9, 10]. While individual patterns of healthcare utilization may vary based on numerous factors [11], there are existing comprehensive theoretical models such as Andersen's Behavioral Model of Health Services Use that predict healthcare utilization based on predisposing characteristics, enabling factors, and health needs [12]. Therefore, when studying utilization trends across population groups, it important to also account for predisposing characteristics such as sociodemographic and behavioral factors that have been found to be associated with use of healthcare utilization. For example, a higher proportion of rural residents have no healthcare visits in a year compared to urban residents [13]. Additionally, younger people, Black people, and males have fewer physician's office visits in a year than older people, White people, and females, respectively [14–17]. Furthermore, people who smoke have been shown to have higher overall outpatient healthcare utilization despite using preventive healthcare services less frequently than people who do not smoke [18–22].

There are a few reasons that highlight the need for quantifying rural disparity in healthcare utilization using real world healthcare system data. First, healthcare system data represents a complementary population to general population survey data. Much of the existing evidence on healthcare utilization has been based on general population surveys or Medicare fee-for-service claims data [23–25]. It is possible that general population surveys on healthcare utilization may be largely impacted by responses from people with few healthcare needs relative to people who interact with the healthcare system. In fact, nearly 16% of respondents to the National Health Interview Survey (NHIS), a prototypical general population survey examining health behaviors and outcomes, have no healthcare encounters in a year [26]. While this could partially be due to a relative lack of healthcare access amongst this group, NHIS surveys have

been found to have a bias for healthier patients relative to the general public [27]. The Medicare-based studies may also be non-representative of the overall adult population given their focus on a population in which adults age 65+ are overrepresented. Second, healthcare system data will allow us to not only quantify the differences between healthcare utilization patterns in patients in rural and urban clinics, but also to adjust for additional predictors like individual health status and smoking status that are potentially associated with utilization. This methodology can attenuate the effects of confounding factors that could impact national data given that the study populations could come from different healthcare systems or structures. It is important to examine whether factors observed in existing studies based on surveys and Medicare claims data hold true among a general adult population who accesses care within the same large healthcare system. To further current evidence on rural health disparity based on national survey or claims data, our study focuses on a more homogeneous population from a single large healthcare system among patients with existing access to healthcare, and uses real-world EHR data to derive utilization and health status, filling a gap on disparity among patients accessing the same healthcare system.

The objective of this study is to identify how predisposing characteristics, enabling factors, and health needs are jointly and individually associated with epidemiological patterns of outpatient healthcare utilization for patients who already interact and engage with a healthcare system. We hypothesize that, among this patient population, healthcare utilization among those accessing care will be lower in patients in rural versus urban clinics, even after adjusting for other significant predictors.

## Materials and methods

### Design and setting

This study used a cross-sectional, retrospective design to examine the level of outpatient healthcare utilization among patients serving the greater St. Louis, southern Illinois, and mid-Missouri regions stratified by age, gender, race, ethnicity, smoking status, health status and rural versus urban clinic location. Data used for these analyses included electronic health records (EHR) for outpatient visits from 2 June 2018 to 31 March 2019, a timeframe that encompassed the first ten months of the healthcare system's use of the EHR system observed in this study. Of the 766 included clinics, 693 were classified as urban and 73 were classified as rural.

### Population

Our healthcare system provided care for 474,674 adult patients aged 18 years and older with 1,423,166 documented in-person outpatient clinical evaluation and management encounters. For this study to evaluate patients from rural vs. urban clinics, we evaluated 124,688 rural outpatient visits (from 50,250 patients who visited only rural clinics) and 1,269,975 urban outpatient visits (from 424,424 patients who visited only urban clinics). For the purpose of comparing patients in rural vs. urban clinics, we excluded a small portion of 28,503 outpatient visits (from 4,456 patients who visited both urban and rural clinics).

### Variables

The number of outpatient visits during the study period was the outcome variable chosen as a proxy for level of outpatient healthcare utilization. The independent variables were selected based on several "predisposing characteristics" from an adopted Andersen model of healthcare utilization that were consistently documented and accessible within the EHR for the majority of patients [12]. These included several predisposing characteristics (age, gender, race,
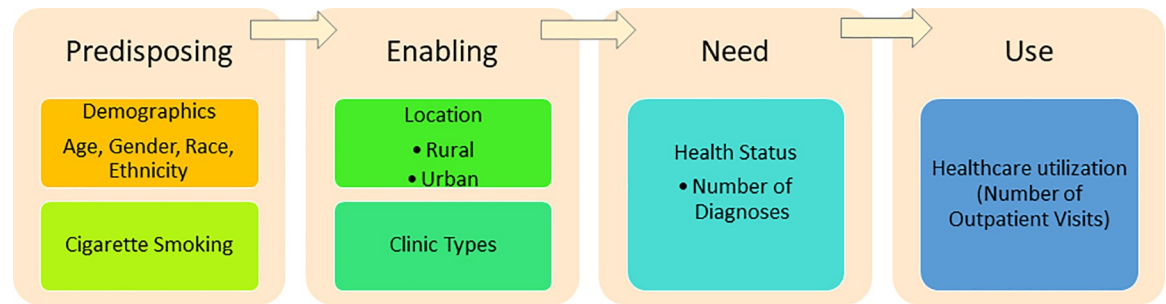
**Fig 1. Adapted Andersen Behavioral Model of Health Services Use.**

ethnicity, cigarette smoking), enabling factors (rural or urban clinic location, clinic types), and health need (health status) (See Fig 1).

**Age, gender, race, and ethnicity.** Information on age (coded as 18–49, 50–59, 60–69, ≥70), gender (male or female), race (White, Black, or Other), and ethnicity (Hispanic or Non-Hispanic) were obtained in the EHR.

**Cigarette smoking.** If an individual was ever documented as a person who smoked in a clinic encounter examined within the data timeframe, they were classified as a person who smoked. This identity was self-reported at the time of clinic encounter.

**Location.** We categorized rural versus urban location using clinic instead of patient home addresses in order to more directly examine the setting in which outpatient clinical encounters occurred. To categorize clinics by location, we input clinic address into Google Earth Pro and classified the clinic's county into one of six levels based on the CDC's 2013 National Center for Health Statistics Urban–Rural Classification Scheme for Counties. Following this guideline, metropolitan ("urban") clinics included the first four levels of the Metropolitan Statistical Areas (large central metro, large fringe metro, medium metro, and small metro), while non-metropolitan ("rural") clinics included those in micropolitan and noncore statistical areas [28].

**Health status.** Health status was determined by the number of unique diagnoses in the EHR data, and was converted into quartiles in analyses (≤2, 3–5, 6–8, and ≥9 diagnoses). The number of diagnoses is used as the measure for patient health status adapted from other studies [29–31]. We also used an alternative measure of comorbidity, the Elixhauser comorbidity score in additional analyses [32].

**Clinic type.** In secondary analyses, we evaluated rural disparity in primary vs. specialty clinics. Primary care clinics included Family Medicine, Internal Medicine, and Pediatrics. All other clinics such as Ancillary Services, Dermatology, Surgery Departments, Obstetrics & Gynecology, Ophthalmology, Pain Management, Physical Medicine and Rehabilitation, Psychiatry, Radiation Oncology and Radiology were considered specialty clinics, see **S1 Table** for the full list of clinics.

**Health professional shortage.** In secondary analyses, we defined level of health professional shortage in each clinic location by obtaining data from the Area Health Resource File (AHRF) from data.HRSA.gov [7] and dividing counties into tertiles (low, medium, high).

## Statistical analysis

The descriptive statistics were provided for the primary outcome and covariates. The univariate and multivariate generalized estimating equation (GEE) models with Poisson link function were employed to investigate the effect of the covariates on healthcare utilization, in which the correlation among patients within clinics was considered. Relative risks (RRs) and 95% CIs

were estimated and the standard errors were calculated using the GEE sandwich method when accounting for within-clinic correlation. The comparisons of the distributions of predisposing characteristics, enabling factors and health needs were conducted through $\chi^2$ tests. Supplementary analyses were conducted observing healthcare utilization patterns within rural and urban clinics separately and primary care and specialty clinics separately. To address the concern of multiple comparisons, we adjusted our significance threshold from 0.05 to 0.001 given the number of tests conducted. All analyses were conducted using R version 3.5.3 for Microsoft Windows and SAS software, Version 9.4m6 of the SAS system for Microsoft Windows.

### Ethics statement

The Institutional Review Board determined this study involved non-human subjects and waived the requirement for informed consent because this is part of conducting a quality improvement project. This study did not meet federal definitions under the jurisdiction of the Institutional Review Board and falls outside the purview of the Human Protection Office.

## Results

### Study sample

Our sample included a total of 474,674 patients. The combined patient sample included 59.5% females, 80.1% White patients, 15.7% Black patients, 98.5% Non-Hispanic patients, and 14.7% patients who smoke, and 10.6% patients in exclusively rural clinics (**Table 1**).

### Patient characteristics in rural vs. urban clinics

The characteristics of the rural and urban samples are shown in **S2 Table**. Compared to the urban group, notable differences in the rural group include a higher proportion of White patients (94.4% vs 78.5%, $\chi^2$ = 7608, df = 2, p<0.0001), a higher proportion of patients who smoke (20.7% vs 13.9%, $\chi^2$ = 1504, df = 1, p<0.0001), and a higher proportion of patients with worse health status defined by more diagnoses (overall $\chi^2$ = 935.1, df = 3, p<0.0001).

**Healthcare utilization across patients characteristics.** The number of outpatient clinic visits across age, gender, race, ethnicity, smoking status, and health status groups are shown in **Table 2**, showing variable utilization across age, gender, race, smoking, and health status groups. Healthcare utilization across these patient characteristics in is also shown in **Table 3**, showing a general pattern of lower utilization in patients seen in rural vs. urban clinics.

First, most demographic, smoking, and health status factors were associated with utilization in univariate analyses. Specific results of univariate analyses from a GEE model that tests the association of health need and several predisposing characteristics and enabling factors with healthcare utilization are shown in **Table 4** left panel. Older age was associated with increasing visits [age $\geq$70 vs. age 18–49, RR = 1.37, 95%CI = (1.33, 1.41); p<0.0001]. Female vs. male gender was also associated with higher utilization [RR = 1.04, 95%CI = (1.03, 1.06), p<0.0001]. Black vs. White patients had higher utilization [RR = 1.06, 95%CI = (1.05, 1.08), p<0.0001] and patients from other racial groups had lower utilization [RR = 0.89, 95%CI = (0.87, 0.92), p<0.0001]. Smoking was associated with higher utilization [RR = 1.04, 95%CI = (1.02, 1.05), p<0.0001]. Worse health status was associated with higher utilization [Q4 vs. Q1, RR = 4.51, 95%CI = (4.25, 4.78); p<0.0001]. Additionally, we noted a trend of lower utilization in patients in rural vs. urban clinics that has not reached statistical significance in univariate analyses. Second, we found that rural location and health status were significant predictors for utilization in multivariate analyses when adjusting for demographic factors and all predictors (Table 4 right panel). The multivariate analyses demonstrated a lower utilization among patients in rural vs.

**Table 1. Sample characteristics[a].**

| | (N = 474,674) |
|---|---|
| | **n (%)** |
| **Age (years)** | |
| 18–49 | 171,566 (36.1%) |
| 50–59 | 91,079 (19.2%) |
| 60–69 | 105,622 (22.3%) |
| ≥70 | 106,407 (22.4%) |
| **Gender[b]** | |
| Male | 192,197 (40.5%) |
| Female | 282,283 (59.5%) |
| **Race[c]** | |
| White | 371,208 (80.1%) |
| Black | 72,575 (15.7%) |
| Other | 19,273 (4.2%) |
| **Ethnicity[d]** | |
| Hispanic | 6,493 (1.5%) |
| Non-Hispanic | 440,672 (98.5%) |
| **Smoking Status[e]** | |
| Non-smoker | 360,661 (85.3%) |
| Smoker | 62,120 (14.7%) |
| **Clinic location[f]** | |
| Urban | 424,424 (89.4%) |
| Rural | 50,250 (10.6%) |
| **Health Status[g]** | |
| Q1 (≤2 diagnoses) | 143, 702 (30.3%) |
| Q2 (3–5 diagnoses) | 102,795 (21.7%) |
| Q3 (6–8 diagnoses) | 126,909 (26.7%) |
| Q4 (≥9 diagnoses) | 101,268 (21.3%) |

[a] The sample was from outpatients, serving the greater St. Louis, southern Illinois, and mid-Missouri regions from June 2018- March 2019.

[b] Documented gender was missing for 194 patients.

[c] Documented race was missing for 11,618 patients.

[d] Documented ethnicity was missing for 27,509 patients.

[e] Smoking status was missing for a total of 51,893 patients.

[f] These patients are unique and exclusively visited urban or rural clinics.

[g] Health Status is defined as the number of ICD 10 diagnosis codes by quartile.

urban clinics [2.49 vs. 3.18 visits, RR = 0.61, 95%CI = (0.55, 0.68), p<0.0001] (Tables 3 and 4). Health status remained a significant predictor for utilization [Q4 vs. Q1, RR = 4.52, 95%CI = (4.26, 4.79); p<0.0001]. Age remained a significant predictor for utilization. However, the associations of gender, race, and smoking with level of healthcare utilization observed in uni-variate analyses were no longer significant after controlling for the other variables in the multi-variate model. In additional analyses, we reached similar results when adjusting for additional covariates such as the Elixhauser comorbidity score [32] and health professional shortage as shown in **S3 and S4 Tables**.

**Fig 1** visually demonstrates our adapted Andersen Behavioral Model of Health Services Use with the predisposing characteristics, enabling factors, and health needs that were considered in our study (Fig 1).

**Table 2. Healthcare utilization across patient characteristics[a].**

|  | Average number of visits |
|---|---|
|  | mean (sd) (N = 474,674) |
| Age |  |
|   18–49 | 2.57 (2.59) |
|   50–59 | 3.04 (3.15) |
|   60–69 | 3.41 (3.71) |
|   $\geq$70 | 3.71 (3.94) |
| Gender |  |
|   Male | 3.08 (3.46) |
|   Female | 3.12 (3.24) |
| Race |  |
|   White | 3.08 (3.33) |
|   Black | 3.46 (3.51) |
|   Other | 2.70 (2.82) |
| Ethnicity |  |
|   Hispanic | 2.96 (3.02) |
|   Non-Hispanic | 3.17 (3.39) |
| Smoking Status[b] |  |
|   Non-smoker | 3.09 (3.34) |
|   Smoker | 3.21 (3.28) |
| Location[c] |  |
|   Urban | 3.18 (3.38) |
|   Rural | 2.49 (2.84) |
| Health Status[d] |  |
|   Q1($\leq$2 diagnoses) | 1.51 (1.48) |
|   Q2 (3–5 diagnoses) | 2.08 (1.85) |
|   Q3(6–8 diagnoses) | 2.44 (2.85) |
|   Q4 ($\geq$9 diagnoses) | 6.15 (4.58) |

[a] Healthcare Utilization is defined as number of visits to any outpatient clinics in 766 clinics serving the greater St. Louis, southern Illinois, and mid-Missouri regions from June 2018- March 2019.

[b] Individuals were classified as smokers in this study if they were ever documented as a smoker in a clinic encounter recorded within the data timeframe. This identity was self-reported at the time of clinic encounter.

[c] These patients are unique and exclusively visited urban or rural clinics.

[d] Health Status is defined as the number of ICD 10 diagnosis codes by quartile.

https://doi.org/10.1371/journal.pone.0263718.t002

## Medical specialties of clinics by location

The distribution of primary vs. specialty clinics also differ in rural vs. urban settings as shown in **S1 Table**. "Primary care" clinic types included Family Medicine, Internal Medicine, and Pediatrics clinics while "specialty" clinics included those from all other specialties. There were notably fewer clinics in rural areas [73 rural clinics (9.5%); 693 urban clinics (90.5%)]. Primary care clinics made up a significantly higher proportion of clinics in rural areas compared to urban areas (71% of rural clinics; 44% of urban clinics) and specialty clinics were more prevalent in urban areas.

## Healthcare utilization across patients characteristics: Primary vs. specialty care

Given the predominance of primary care clinics and the relative lack of specialty clinics in rural areas, we evaluated if rural disparity in utilization was observed in in primary care vs.

**Table 3. Healthcare utilization across patient characteristics in rural vs. urban clinics[a].**

| | Overall | Rural | Urban |
|---|---|---|---|
| | *Average number of visits* | *Average number of visits* | *Average number of visits* |
| | *mean (sd)* | *mean (sd)* | *mean (sd)* |
| **Age** | | | |
| 18–49 | 2.57 (2.59) | 2.12 (1.74) | 2.64 (2.68) |
| 50–59 | 3.04 (3.15) | 2.37 (2.38) | 3.11 (3.22) |
| 60–69 | 3.41 (3.71) | 2.64 (3.19) | 3.50 (3.75) |
| ≥70 | 3.71 (3.94) | 3.10 (4.00) | 3.79 (3.92) |
| **Gender** | | | |
| Male | 3.08 (3.46) | 2.46 (3.14) | 3.16 (3.49) |
| Female | 3.12 (3.24) | 2.52 (2.62) | 3.20 (3.30) |
| **Race** | | | |
| White | 3.08 (3.33) | 2.52 (2.87) | 3.17 (3.39) |
| Black | 3.46 (3.51) | 2.35 (2.60) | 3.48 (3.52) |
| Other | 2.70 (2.82) | 2.04 (2.14) | 2.75 (2.86) |
| **Ethnicity** | | | |
| Hispanic | 2.96 (3.02) | 2.46 (2.45) | 3.00 (3.06) |
| Non-Hispanic | 3.17 (3.39) | 2.54 (2.90) | 3.25 (3.43) |
| **Smoking Status[b]** | | | |
| Non-smoker | 3.09 (3.34) | 2.48 (2.93) | 3.16 (3.37) |
| Smoker | 3.21 (3.28) | 2.57 (2.48) | 3.33 (3.39) |
| **Health status[c]** | | | |
| Q1(≤2 diagnoses) | 1.51 (1.48) | 1.46 (2.39) | 1.52 (1.34) |
| Q2 (3–5 diagnoses) | 2.08 (1.85) | 1.71 (1.47) | 2.12 (1.88) |
| Q3(6–8 diagnoses) | 2.44 (2.85) | 2.15 (1.80) | 2.95 (2.51) |
| Q4 (≥9 diagnoses) | 6.15 (4.58) | 4.69 (3.89) | 6.34 (4.62) |

[a] Healthcare Utilization is defined as number of visits to any outpatient clinics in 766 clinics serving the greater
St. Louis, southern Illinois, and mid-Missouri regions from June 2018- March 2019.
[b] Individuals were classified as smokers in this study if they were ever documented as a smoker in a clinic encounter
recorded within the data timeframe. This identity was self-reported at the time of clinic encounter.
[c] Health Status is defined as the number of ICD 10 diagnosis codes by quartile.

specialty clinic types (**S5** and **S6** **Tables**). We observed a similar trend of rural disparity in utilization across primary and specialty care clinics.

## Discussion

Using data from a large healthcare system, we have evaluated patient characteristics and healthcare utilization in rural vs. urban clinics to better understand the extent and pattern of health disparity. We present evidence to quantify a significantly lower utilization of outpatient care among patients in rural clinics, even after adjusting for health status and several predisposing characteristics and enabling factors.

First, we confirmed prior findings that predisposing demographic factors are associated with utilization. Older patients had higher utilization, even considering other predictors such as health status. This could be due in part to the high prevalence of Medicare coverage above age 65 minimizing barriers to healthcare access caused by lack of insurance. In addition, we also identified the impact of race, gender, and smoking on utilization; with Black race, female gender, and cigarette smoking associated with higher utilization. However, these associations

**Table 4. Association of patients characteristics and healthcare utilization: Univariate and multivariate analyses.**

| | Univariate Analysis | | | Multivariate Analysis | | |
|---|---|---|---|---|---|---|
| | RR[b] | 95% CI | p[*] | RR | 95% CI | p |
| Age | | | | | | |
| 18–49 | | Reference | | | Reference | |
| 50–59 | 1.13 | 1.12, 1.15 | **<0.0001** | 0.97 | 0.96, 0.98 | **<0.0001** |
| 60–69 | 1.26 | 1.24, 1.28 | | 0.99 | 0.98, 1.00 | |
| ≥70 | 1.37 | 1.33, 1.41 | | 1.02 | 1.00, 1.03 | |
| Gender | | | | | | |
| Male | | Reference | | | Reference | |
| Female | 1.04 | 1.03, 1.06 | **<0.0001** | 0.99 | 0.98, 1.00 | 0.307 |
| Race | | | | | | |
| White | | Reference | | | Reference | |
| Black | 1.06 | 1.05, 1.08 | **<0.0001** | 1.00 | 0.99, 1.01 | 0.207 |
| Other | 0.89 | 0.87, 0.92 | | 0.99 | 0.97, 1.00 | |
| Ethnicity | | | | | | |
| Hispanic | 0.96 | 0.94, 0.99 | 0.031 | 1.01 | 1.00, 1.03 | 0.18 |
| Non-Hispanic | | Reference | | | Reference | |
| Smoking Status[c] | | | | | | |
| Non-smoker | | Reference | | | Reference | |
| Smoker | 1.04 | 1.02, 1.05 | **<0.0001** | 1.00 | 0.99, 1.01 | 0.76 |
| Health status[d] | | | | | | |
| Q1(≤2 diagnoses) | | Reference | | | Reference | |
| Q2 (3–5 diagnoses) | 1.68 | 1.61, 1.76 | **<0.0001** | 1.69 | 1.62, 1.76 | **<0.0001** |
| Q3(6–8 diagnoses) | 2.40 | 2.27, 2.53 | | 2.40 | 2.27, 2.54 | |
| Q4 (≥9 diagnoses) | 4.51 | 4.25, 4.78 | | 4.52 | 4.26, 4.79 | |
| Location[e] | | | | | | |
| Urban | | Reference | | | Reference | |
| Rural | 0.93 | 0.70, 1.23 | 0.594 | 0.61 | 0.55, 0.68 | **<0.0001** |

[a] Healthcare Utilization is defined as number of visits to any outpatient clinics in 766 clinics serving the greater St. Louis, southern Illinois, and mid-Missouri regions from June 2018- March 2019.

[b] RR: Relative risk.

[c] Individuals were classified as smokers in this study if they were ever documented as a smoker in a clinic encounter recorded within the data timeframe. This identity was self-reported at the time of clinic encounter.

[d] Health Status is defined as the number of ICD 10 diagnosis codes by quartile.

[e] These patients are unique and exclusively visited urban or rural clinics.

[*]Boldface indicates statistical significance (p<0.001).

were likely mediated through individual's health status as they became insignificant when adjusting for other factors. We have conjectures as to why these populations might have higher health needs based on health status, including that Black populations face higher burdens of several diseases than do White populations [33], obstetric and/or gynecological care recommendations for female patients may increase their prescribed health needs relative to men, and smoking is associated with a number of health consequences that lead to increased health needs.

Second, we found lower healthcare utilization among rural clinic patients compared to their urban counterparts, with 22% fewer interactions with the healthcare system on average than urban clinic patients over the study period. Adjusting for other variables (age, gender,

race, ethnicity, smoking status, and health status), patients in rural vs. urban clinics were associated with significantly less outpatient visits. This rural disparity becomes more apparent when multiple predictors for utilization are also considered. For example, our findings shows that rural vs. urban clinic patients were more likely to be smokers (20.7% vs 13.9%) with more health problems (number of diagnoses> = 6) (54.2% vs 47.4%), but their utilization were actually lower, suggesting a combination of higher disease burden and lower healthcare use. These findings suggest that disparity may be masked if the uneven distribution of health risk factors of patients in rural and urban clinics is not considered. Even with a higher proportion of primary care clinics in rural areas in our study, patients in primary care clinics in rural areas had lower utilization than patients in primary care clinics in urban areas after adjusting for other factors. In specialty clinics, we observed a similar pattern of lower utilization among patients in rural clinics. This supports that a thorough evaluation of healthcare utilization is complex and requires consideration of multiple factors [34]. The phenomenon of lower healthcare utilization in rural areas could reflect geographical [24] or sociocultural barriers to accessing care in rural areas, issues with insurance coverage, time constraints, a relative lack of provider availability, or a lack of transportation [35]. With fewer healthcare encounters, rural patients may have less access to health information that could curtail health behaviors such as smoking, drinking, and lack of exercise, which are seen at higher incidences in rural areas [34]. There is also the possibility that rural clinics may operate more efficiently than urban clinics, meaning less visits noted, but similar health outcomes achieved. However, our findings show there is significant disparity in health status in patients in rural clinics when examining the association of patient characteristics visiting primary care clinics. In addition to the possible contributors to lower rural healthcare utilization discussed above, there could be differences in rural and urban clinic scheduling practices given lower provider availability in rural areas that contribute to less frequent encounters for patients in rural clinics.

Our findings argue for the critical importance to address barriers to healthcare with innovative sustainable solutions to limit rural disparity. One potential solution is to leverage the use of advancements in technology [36, 37] such as, telehealth and telecommunication [38, 39], health information technology, health management, decision support tools, patient education using outreach [40–42], and point of care models [5, 43].

To our knowledge, our study is unique in quantifying the healthcare utilization patterns of a diverse general adult patient population over a substantial study period using direct electronic medical record information. Previous data on factors associated with healthcare utilization have largely been driven by general population surveys, provider-based surveys, or Medicare claims information. Our data suggests that examining healthcare system level data yields similar conclusions for patients who access care. Even among patients with access to the same healthcare system, rural vs. urban clinic patients have worse health status and lower utilization, suggesting multiple factors (e.g., transportation) other than the general access to local healthcare may contribute to the overall health disparity.

Despite its novelty, there are limitations to this study. First, due to the restrictions on what factors are consistently reported or easily accessible in electronic health records, our analysis did not examine additional important factors such as insurance status, socioeconomic status, access to transportation, or family structure that could reasonably be associated with healthcare utilization. These results therefore need to be considered with caution. We suggest caution in interpreting these results that this study of patients engaged in a large healthcare system may suffer from an inherent selection bias that most patients accessing healthcare are likely to have insurance support. The impact of insurance on health status and utilization needs to be further studied. Second, using the number of diagnoses as a proxy for health status may not fully account for the severity of the individual diagnoses, which has important implications

when determining one's health status. Some of the quartiles also have a wide range of number of diagnoses, which might suggest a wide range of health statuses within a quartile. In secondary analysis, we reached similar results using either the number of diagnoses or Elixhauser comorbidity score as a proxy for patient overall health [29–31]. Third, this study categorized patients based on their clinic locations rather than personal addresses. To reduce the potential crossover (e.g, rural residing patients visiting an urban clinic), we evaluated patients who exclusively visit rural or urban clinics and excluded patients who visited both rural and urban clinics (only ~1% of the sample, suggesting that clinics largely served either a majority of patients living in rural areas or a majority of patients living in urban areas). The low crossover rate also reduces the concern of skewing results via the exclusion of rural patients needing specialty care requiring visits to urban clinics, which might reflect increased health needs and entail increased healthcare encounters among such patients. Lastly, although this study examines healthcare utilization patterns across large adult patient populations, we acknowledge these results may be limited in generalizability to other systems outside the greater St. Louis, southern Illinois, and mid-Missouri regions. However, given the sociodemographic, behavioral, and geographic diversity of the sample included in our study, we feel our findings can be informative on patterns of healthcare utilization to a wide range of healthcare systems.

In summary, our study adds to existing healthcare utilization research by quantifying the extent to which patients in rural clinics utilize healthcare differently than patients in urban clinics in the context of their predisposing characteristics, enabling factors, and health needs within a large healthcare system. In our study, patients in rural clinics had worse health status and lower healthcare utilization than patients in urban clinics, even within the same large healthcare system. Understanding the root causes of rural disparity is an important topic requiring further qualitative and quantitative research. The current work has the potential to guide future research on the association of rural healthcare system utilization with rural healthcare outcomes disparities and ways to address barriers to healthcare access once patients enter the healthcare system.

## Conclusions

Our research examining a large population of adults within a diverse healthcare system demonstrates lower outpatient healthcare utilization amongst rural populations despite worse health status amongst this group. This conclusion supports findings from previous healthcare utilization data which has largely been driven by general population surveys or Medicare claims information which might have a bias for healthier and older populations, respectively. Further studies exploring how unique healthcare utilization patterns among rural groups are related to health outcomes in other large scale healthcare systems can help generalize findings and identify possible solutions to eliminating rural healthcare disparities for patients who already access healthcare.

## Supporting information

**S1 Table. Distributions of clinic specialties differ in rural vs. urban clinics.**
(DOCX)

**S2 Table. Sample characteristics in rural vs. urban clinics.**
(DOCX)

**S3 Table. Healthcare utilization[a] across patient characteristics including health professional shortage.**
(DOCX)

**S4 Table. Healthcare utilization[a] across patient characteristics including Elixhauser.**
(DOCX)

**S5 Table. Association of patients characteristics and primary care utilization: Multivariate analyses[a,b].**
(DOCX)

**S6 Table. Association of patients characteristics and specialty care utilization: Multivariate analyses [a,b].**
(DOCX)

# Author Contributions

**Conceptualization:** Akua Nuako, Nina Smock, Aimee James, Timothy Baker, Laura Bierut, Graham Colditz, Li-Shiun Chen.

**Data curation:** Jingxia Liu, Giang Pham, Nina Smock, Li-Shiun Chen.

**Formal analysis:** Jingxia Liu, Giang Pham, Li-Shiun Chen.

**Investigation:** Akua Nuako, Nina Smock, Li-Shiun Chen.

**Methodology:** Jingxia Liu, Giang Pham, Nina Smock, Li-Shiun Chen.

**Project administration:** Nina Smock.

**Software:** Jingxia Liu, Giang Pham, Nina Smock.

**Supervision:** Li-Shiun Chen.

**Validation:** Jingxia Liu, Giang Pham, Nina Smock, Li-Shiun Chen.

**Visualization:** Akua Nuako, Giang Pham, Nina Smock.

**Writing – original draft:** Akua Nuako, Giang Pham, Nina Smock, Li-Shiun Chen.

**Writing – review & editing:** Akua Nuako, Jingxia Liu, Giang Pham, Nina Smock, Aimee James, Timothy Baker, Laura Bierut, Graham Colditz, Li-Shiun Chen.

# References

1. American Community Survey: 2011–2015. The United States Census Bureau. Retrieved November 28, 2020, from https://www.census.gov/newsroom/press-releases/2016/cb16-210.html.

2. Garcia MC. Potentially Excess Deaths from the Five Leading Causes of Death in Metropolitan and Non-metropolitan Counties—United States, 2010–2017. MMWR Surveill Summ. 2019; 68. https://doi.org/10.15585/mmwr.ss6810a1 PMID: 31697657

3. Singh GK, Siahpush M. Widening Rural–Urban Disparities in All-Cause Mortality and Mortality from Major Causes of Death in the USA, 1969–2009: 21.

4. Knudson A, Meit M, Tanenbaum E, Brady J, Gilbert T, Klug MG, et al. Exploring Rural and Urban Mortality Differences. August 2015: 1.

5. Ramsey AT, Baker TB, Pham G, Stoneking F, Smock N, Colditz GA, et al. Low Burden Strategies Are Needed to Reduce Smoking in Rural Healthcare Settings: A Lesson from Cancer Clinics. International journal of environmental research and public health. 2020; 17(5). Epub 2020/03/12. https://doi.org/10.3390/ijerph17051728 PMID: 32155775; PubMed Central PMCID: PMC7084618.

6. United States Census Bureau. Rates of Uninsured Fall in Rural Counties, Remain Higher Than Urban Counties. https://www.census.gov/library/stories/2019/04/health-insurance-rural-america.html.

7. Bureau of Health Workforce, Health Resources and Services Administration (HRSA), U.S. Department of Health & Human Services. https://data.hrsa.gov/tools/shortage-area/hpsa-find Accessed December 7, 2021.

8.  Zhao G, Okoro CA, Hsia J, Garvin WS, Town M. Prevalence of Disability and Disability Types by Urban-Rural County Classification-U.S., 2016. American journal of preventive medicine. 2019; 57 (6):749–56. Epub 2019/11/23. https://doi.org/10.1016/j.amepre.2019.07.022 PMID: 31753256.

9.  Bernstein AB, Hing E, Moss AJ, Allen KF, Siller AB, Tiggle RB. Health Care in America: Trends in Utilization.  Hyattsville, Maryland:  National Center for Health Statistics; 2004: 2.

10. Chen X, Orom H, Hay JL, Waters EA, Schofield E, Li Y, et al. Differences in Rural and Urban Health Information Access and Use. The Journal of rural health: official journal of the American Rural Health Association and the National Rural Health Care Association. 2019; 35(3):405–17. Epub 2018/11/18. https://doi.org/10.1111/jrh.12335 PMID: 30444935; PubMed Central PMCID: PMC6522336.

11. Committee on Health Care Utilization and Adults with Disabilities, Board on Health Care Services, Health and Medicine Division, National Academies of Sciences, Engineering, and Medicine. Summary: Factors that Affect Health-Care Utilization. In: Health-Care Utilization as a Proxy in Disability Determination.  Washington, D.C.:  National Academies Press; 2018:24969. https://doi.org/10.17226/24969 PMID: 29782136

12. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? Journal of health and social behavior. 1995; 36(1):1–10. Epub 1995/03/01. PMID: 7738325.

13. National Center for Health Statistics (US), Centers for Disease Control and Prevention. Health, United States, 2018. Hyattsville, MD: National Center for Health Statistics; 2019: Trend Tables, Table 30. https://www.ncbi.nlm.nih.gov/books/NBK551099/table/ch3.tab30/. Accessed January 24, 2021.

14. Rui P, Okeyode T. National Ambulatory Medical Care Survey: 2016 National Summary Tables. National Center for Health Statistics; 2019: 5,7. https://www.cdc.gov/nchs/data/ahcd/namcs_summary/2016_namcs_web_tables.pdf. Accessed January 24, 2021.

15. Lassman D, Hartman M, Washington B, Andrews K, Catlin A. US health spending trends by age and gender: selected years 2002–10. Health affairs (Project Hope). 2014; 33(5):815–22. Epub 2014/05/07. https://doi.org/10.1377/hlthaff.2013.1224 PMID: 24799579.

16. Bertakis KD, Azari R, Helms LJ, Callahan EJ, Robbins JA. Gender differences in the utilization of health care services. The Journal of family practice. 2000; 49(2):147–52. Epub 2000/03/16. PMID: 10718692.

17. Manuel JI. Racial/Ethnic and Gender Disparities in Health Care Use and Access. Health services research. 2018; 53(3):1407–29. Epub 2017/05/10. https://doi.org/10.1111/1475-6773.12705 PMID: 28480588; PubMed Central PMCID: PMC5980371.

18. Vander Weg MW, Howren MB, Cai X. Use of routine clinical preventive services among daily smokers, non-daily smokers, former smokers, and never-smokers. Nicotine & tobacco research: official journal of the Society for Research on Nicotine and Tobacco. 2012; 14(2):123–30. Epub 2011/12/20. https://doi.org/10.1093/ntr/ntr141 PMID: 22180595.

19. Kahende JW, Adhikari B, Maurice E, Rock V, Malarcher A. Disparities in health care utilization by smoking status—NHANES 1999–2004. International journal of environmental research and public health. 2009; 6(3):1095–106. Epub 2009/05/15. https://doi.org/10.3390/ijerph6031095 PMID: 19440435; PubMed Central PMCID: PMC2672402.

20. Hutcheson TD, Greiner KA, Ellerbeck EF, Jeffries SK, Mussulman LM, Casey GN. Understanding smoking cessation in rural communities. The Journal of rural health: official journal of the American Rural Health Association and the National Rural Health Care Association. 2008; 24(2):116–24. Epub 2008/04/10. https://doi.org/10.1111/j.1748-0361.2008.00147.x PMID: 18397444.

21. Roberts ME, Doogan NJ, Kurti AN, Redner R, Gaalema DE, Stanton CA, et al. Rural tobacco use across the United States: How rural and urban areas differ, broken down by census regions and divisions. Health & place. 2016; 39:153–9. Epub 2016/04/25. https://doi.org/10.1016/j.healthplace.2016.04.001 PMID: 27107746; PubMed Central PMCID: PMC4874850.

22. Xu X, Bishop EE, Kennedy SM, Simpson SA, Pechacek TF. Annual healthcare spending attributable to cigarette smoking: an update. American journal of preventive medicine. 2015; 48(3):326–33. Epub 2014/12/17. https://doi.org/10.1016/j.amepre.2014.10.012 PMID: 25498551; PubMed Central PMCID: PMC4603661.

23. Casey MM, Thiede Call K, Klingner JM. Are rural residents less likely to obtain recommended preventive healthcare services? American journal of preventive medicine. 2001; 21(3):182–8. Epub 2001/09/25. https://doi.org/10.1016/s0749-3797(01)00349-x PMID: 11567838.

24. Arcury TA, Gesler WM, Preisser JS, Sherman J, Spencer J, Perin J. The effects of geography and spatial behavior on health care utilization among the residents of a rural region. Health services research. 2005; 40(1):135–55. Epub 2005/01/25. https://doi.org/10.1111/j.1475-6773.2005.00346.x PMID: 15663706; PubMed Central PMCID: PMC1361130.

25. Kosar CM, Loomer L, Ferdows NB, Trivedi AN, Panagiotou OA, Rahman M. Assessment of Rural-Urban Differences in Postacute Care Utilization and Outcomes Among Older US Adults. JAMA network open. 2020; 3(1):e1918738. Epub 2020/01/09. https://doi.org/10.1001/jamanetworkopen.2019.18738

PMID: 31913495; PubMed Central PMCID: PMC6991315 the American Health Care Association outside the submitted work. No other disclosures were reported.

26. Villarroel MA, Blackwell DL, Jen A. Tables of Summary Health Statistics for US Adults: 2018 National Health Interview Survey. National Center for Health Statistics; 2019: Table A-18c.

27. Keyes KM, Rutherford C, Popham F, Martins SS, Gray L. How Healthy Are Survey Respondents Compared with the General Population?: Using Survey-linked Death Records to Compare Mortality Outcomes. Epidemiology (Cambridge, Mass). 2018; 29(2):299–307. Epub 2018/02/02. https://doi.org/10.1097/EDE.0000000000000775 PMID: 29389712; PubMed Central PMCID: PMC5794231.

28. Ingram DD, Franco SJ. 2013 NCHS Urban-Rural Classification Scheme for Counties. National Center for Health Statistics; 2014: 2.

29. Colantonio A, Gerber G, Bayley M, Deber R, Yin J, Kim H. Differential profiles for patients with traumatic and non-traumatic brain injury. Journal of rehabilitation medicine. 2011; 43(4):311–5. Epub 2011/02/25. https://doi.org/10.2340/16501977-0783 PMID: 21347507.

30. McEwen SE, Elmi S, Waldman M, Bishev M. Inpatient oncology rehabilitation in Toronto: a descriptive 18-month retrospective record review. Supportive care in cancer: official journal of the Multinational Association of Supportive Care in Cancer. 2012; 20(7):1541–7. Epub 2011/08/06. https://doi.org/10.1007/s00520-011-1243-4 PMID: 21818640.

31. Khuu W, Chan V, Colantonio A. A systematic review protocol for measuring comorbidity in inpatient rehabilitation for non-traumatic brain injury. Systematic reviews. 2015; 4(1):14. Epub 2015/01/31. https://doi.org/10.1186/2046-4053-4-14 PMID: 25634135; PubMed Central PMCID: PMC4328947.

32. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Medical care. 1998; 36(1):8–27. Epub 1998/02/07. https://doi.org/10.1097/00005650-199801000-00004 PMID: 9431328

33. U.S. Department of Health and Human Services, Office of Minority Health. Profile: Black/African Americans [Internet]. minorityhealth.hhs.gov. [cited 2021 Jan 25]. Available from: https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=3&lvlid=61.

34. Suka M, Odajima T, Okamoto M, Sumitani M, Igarashi A, Ishikawa H, et al. Relationship between health literacy, health information access, health behavior, and health status in Japanese people. Patient education and counseling. 2015; 98(5):660–8. Epub 2015/03/06. https://doi.org/10.1016/j.pec.2015.02.013 PMID: 25739344.

35. Arcury TA, Preisser JS, Gesler WM, Powers JM. Access to transportation and health care utilization in a rural region. The Journal of rural health: official journal of the American Rural Health Association and the National Rural Health Care Association. 2005; 21(1):31–8. Epub 2005/01/26. https://doi.org/10.1111/j.1748-0361.2005.tb00059.x PMID: 15667007.

36. Madison M, Rudman WJ, Hart-Hester S, Caputo N, French M, Jones WA. Leveraging health information technology and health information management in rural america: a health summit commentary. The Journal of rural health: official journal of the American Rural Health Association and the National Rural Health Care Association. 2012; 28(1):4–7. Epub 2012/01/13. https://doi.org/10.1111/j.1748-0361.2011.00383.x PMID: 22236309.

37. McIlhenny CV, Guzic BL, Knee DR, Wendekier CM, Demuth BR, Roberts JB. Using technology to deliver healthcare education to rural patients. Rural and remote health. 2011; 11(4):1798. Epub 2011/10/15. PMID: 21995854.

38. Talbot JA, Burgess AR, Thayer D, Parenteau L, Paluso N, Coburn AF. Patterns of Telehealth Use Among Rural Medicaid Beneficiaries. The Journal of rural health: official journal of the American Rural Health Association and the National Rural Health Care Association. 2019; 35(3):298–307. Epub 2018/10/06. https://doi.org/10.1111/jrh.12324 PMID: 30288808.

39. Benavides-Vaello S, Strode A, Sheeran BC. Using technology in the delivery of mental health and substance abuse treatment in rural communities: a review. The journal of behavioral health services & research. 2013; 40(1):111–20. Epub 2012/10/25. https://doi.org/10.1007/s11414-012-9299-6 PMID: 23093443.

40. Duffy SA, Ronis DL, Ewing LA, Waltje AH, Hall SV, Thomas PL, et al. Implementation of the Tobacco Tactics intervention versus usual care in Trinity Health community hospitals. Implementation science: IS. 2016; 11(1):147. Epub 2016/11/07. https://doi.org/10.1186/s13012-016-0511-6 PMID: 27814722; PubMed Central PMCID: PMC5097410.

41. Duffy SA, Ronis DL, Karvonen-Gutierrez CA, Ewing LA, Hall SV, Yang JJ, et al. Effectiveness of the Tobacco Tactics Program in the Trinity Health System. American journal of preventive medicine. 2016; 51(4):551–65. Epub 2016/09/21. https://doi.org/10.1016/j.amepre.2016.03.012 PMID: 27647056; PubMed Central PMCID: PMC5031413.

42. Cruvinel E, Richter KP, Stoney C, Duffy S, Fellows J, Harrington KF, et al. CHARTing a Path to Pragmatic Tobacco Treatment Research. American journal of preventive medicine. 2016; 51(4):630–6.

Epub 2016/09/21. https://doi.org/10.1016/j.amepre.2016.05.025 PMID: 27647063; PubMed Central PMCID: PMC5919279.

43. Craig EJ, Ramsey AT, Baker TB, James AS, Luke DA, Malone S, et al. Point of care tobacco treatment sustains during COVID-19, a global pandemic. Cancer epidemiology. 2021:102005. Epub 2021/08/28. https://doi.org/10.1016/j.canep.2021.102005 PMID: 34446379.