County-level predictors of retention in care status among people living with HIV in South Carolina from 2010 to 2016: A data-driven approach

Chengbo Zeng[a,b,c], **Jiajia Zhang**[a,c,d], **Xiaowen Sun**[a,c,d], **Zhenlong Li**[c,e], **Sharon Weissman**[c,f], **Bankole Olatosi**[a,c,g], **Xiaoming Li**[a,b,c]

[a]South Carolina SmartState Center for Healthcare Quality, University of South Carolina

[b]Department of Health Promotion, Education, and Behavior, Arnold School of Public Health, University of South Carolina

[c]University of South Carolina Big Data Health Science Center

[d]Department of Epidemiology and Biostatistics, Arnold School of Public Health

[e]Geoinformation and Big Data Research Lab, Department of Geography, College of Arts and Sciences

[f]School of Medicine, University of South Carolina, Columbia, South Carolina, USA.

[g]Department of Health Services, Policy, and Management, Arnold School of Public Health, University of South Carolina, Columbia, South Carolina, USA.

Abstract

Objective: To examine the geospatial variation of retention in care (RIC) across the counties in South Carolina (SC) from 2010 to 2016 and identify the relevant county-level predictors.

Design: Aggregated data on county-level RIC among HIV patients from 2010 to 2016 were retrieved from an electronic HIV/AIDS reporting system in SC Department of Health and Environmental Control. Sociological framework was used to identify county-level predictors from multiple public datasets.

Methods: Geospatial mapping was used to display the spatial heterogeneity of county-level RIC rate in SC. Generalized linear mixed effect regression with least absolute shrinkage and selection operator (LASSO) was employed to identify county-level predictors related to the change of RIC status over time. Confusion matrix and area under the curve statistics were used to evaluate model performance.

Results: More than half of the counties had their RIC rates lower than the national average. The change of county-level RIC rate from 2010 to 2016 was not significant, and spatial heterogeneity in RIC rate was identified. A total of 22 of the 31 county-level predictors were selected by LASSO for predicting county-level RIC status. Counties with lower collective efficacy, larger proportions

Correspondence to Chengbo Zeng, MS, South Carolina SmartState Center for Healthcare Quality, University of South Carolina Big Data Health Science Center, Department of Health Promotion, Education, and Behavior, Arnold School of Public Health, University of South Carolina, Columbia, SC 29208, USA. Tel: +1 803 777 8615; czeng@E-Mail.sc.edu.

of males and/or persons with high education were more likely to have their RIC rates lower than the national average. In contrast, numbers of accessible mental health centers were positively related to county-level RIC status.

Conclusions: Spatial variation in RIC could be identified, and county-level factors associated with accessible healthcare facilities and social capital significantly contributed to these variation. Structural and individual interventions targeting these factors are needed to improve the county-level RIC and reduce the spatial variation in HIV care.

Keywords

Spatial variation; Retention in care; LASSO regression; HIV/AIDS; South Carolina

---

Introduction

Retention in care (RIC), referring to regular attendance at scheduled HIV medical appointments, is one of the most important components in HIV care continuum (i.e., diagnosis of HIV infection, linkage to HIV medical care, receipt of HIV medical care, retention in medical care, and achievement and maintenance of viral suppression).[1] Retaining to HIV medical care can promote the achievement and maintenance of viral suppression, control the disease progression, and improve the life expectancy of people living with HIV (PLWH).[1] Poor retention in HIV medical care serves as a major barrier of optimal HIV care, correlates with poor clinical outcomes, and increases onward HIV transmission. World Health Organization (WHO) estimated that among PLWH who received HIV medical care, nearly 35% of them were lost to follow-up within 12 months after antiretroviral therapy (ART) initiation.[2] The US Centers for Disease Control and Prevention (CDC) estimated that only 58.0% of the PLWH retained for HIV medical care in 2018.[3, 4] Due to the late diagnosis and poor RIC rate, South Carolina (SC) consistently ranked among the top 10 states in the US with the highest number of AIDS cases for more than a decade.[5] According to the South Carolina's Epidemiologic Profile in 2019, among the 20,166 PLWH, there were only 53.0% retaining to HIV care, which was lower than the national average retention in care rate in 2018 (58.0%).[4, 6] Given the large number of HIV population and low rate of RIC, it is imperative to investigate factors associated with retention in HIV medical care among PLWH in SC.

Most existing research on retention in HIV care has focused on individual-level factors. [1, 7, 8] For instance, Bulsara and colleagues summarized the predictors of RIC among PLWH in both developed and developing countries.[1] These predictors included substance use, sociodemographic characteristics (e.g., gender, age, race/ethnicity, and literacy), HIV-disease progression and physical comorbidities, psychiatric comorbidities, social welfare, and health beliefs.[9–13] However, findings on individual-level factors may be limited to reveal the barriers of RIC at macro level and could not adequately inform future community or structural interventions which could effectively target these barriers and address the geospatial disparities in RIC among PLWH.

The existing literature suggested a spatial variation in RIC.[8] Although geospatial mapping and analysis have been applied into HIV research, most studies focused on spatial

heterogeneity of HIV morbidity and mortality but few on RIC.[14–17] Among the few, Rebeiro and colleagues described the spatial variation in RIC among PLWH in the US from 2000 to 2010 using data derived from the 12 cohorts of the North American AIDS Cohort Collaboration on Research and Design, but they did not investigate the factors associated with this variation.[18] Understanding county-level predictors of spatial variation in RIC is imperative in developing targeted interventions for improving retention in HIV care, especially in the areas with a large number of HIV population and low rate of RIC, such as SC.[8] For example, the information on county-level predictors could help decision-makers to identify areas that require for additional resources and inform tailored community or structural efforts to improve retention in HIV medical care in specific geolocations.

The predictors of spatial variation in RIC could be conceptualized using established conceptual frameworks, such as the sociological framework of health suggested by Macintyre and colleagues.[19] According to the sociological framework of health, factors in compositional, contextual, and collective dimensions could predict the socio-physical environment, social norm, and healthcare access at local areas, which in turn might be associated with personal physical and psychological well-being.[19, 20] Macintyre and colleagues used the sociological framework to propose the paths from geolocations to health outcomes.[19]

The utilization of sociological framework of health could provide a comprehensive understanding of community or structural barriers related to spatial variation in RIC. Compositional factors represent the local population sociodemographic characteristics, such as population density, gender, race/ethnicity, income, unemployment, and health insurance coverage. These factors could reflect the regional economic deprivation (e.g., proportions of persons in low working class, with less than high school education, and/or in poverty), which had been found to be associated with the geospatial disparities in RIC.[21] Compared with the compositional dimension, factors in contextual dimension reflect the regional resource accessibility, such as the socioeconomic opportunities (e.g., Gini index, vacant houses) and healthcare resources (e.g., accessible healthcare facilities, distance to healthcare centers, and healthcare provider rate).[19, 20] Particularly, the accessible healthcare facilities at local areas and distance to these facilities are closely related to RIC.[21] Finally, collective factors represent the social norm, social culture, and social capital at a region. The typical factor in this dimension is religious culture and environment, which could provide people at local areas with social support, social interaction, and social network and thus result in good health outcomes.[22] However, to the best of our knowledge, limited research has used this framework and select the relevant county-level predictors to help conceptualize the spatial variation in RIC among PLWH.

To address these knowledge gaps, the current study examined the spatial variation of RIC in SC through 1) adopting the sociological framework of health in variable selection and hypothesizing that compositional, contextual, and collective factors might contribute to the county-level RIC among PLWH in SC; and 2) leveraging multiple public datasets, and applying geospatial mapping and machine learning approach to investigate county-level factors associated with RIC in SC from 2010 to 2016. This study aimed to generate

empirical evidence that could inform community or structural interventions to improve county-level RIC in SC.

Materials and methods

Participants

All PLWH who received a diagnosis of HIV infection in SC between January 2010 and December 2016 were included in this study. Their de-identified medical records including sociodemographic characteristics and laboratory reports were extracted from the electronic HIV/AIDS reporting system in the SC Department of Health and Environmental Control (DHEC). The study protocol was approved by the Institutional Review Boards at both University of South Carolina and SC DHEC.

County-level retention in care status

Based on the definition recommended by CDC [23], individual-level RIC was calculated at each calendar year and aggregated to the county-level to define the county-level RIC status as "above the national level (> 56.0%)" or "below the national level (56.0%)". According to CDC, the average national level from 2010 to 2016 was 56.0% ranging from 54.7% to 57.6%.[23]

County-level predictors

County-level predictors associated with RIC were selected using the sociological framework of health.[19] Based on this framework, a total of 31 predictors were organized as compositional, contextual, and collective factors.

Compositional factors

Compositional factors referred to the socio-demographic characteristics of individuals living in a certain place. In the current study, 5-year estimated county-level factors, such as population density, proportions of male, individuals who were older than 18 years (%), living alone (%), Black (%), no health insurance coverage (%), and individuals who received public assistance (%). The compositional factors were extracted from American Community Survey (ACS) and were aggregated to counties of SC.[24, 25] To assess the socioeconomic status (SES) of individuals in these areas, 5-year estimated county-level factors including median household annual income, poverty, wealth (proportion of owner-occupied homes worth >= \$750,000), low income (proportion of households with an income < \$25,000), high income (proportion of households with an income > \$200,000), low education (proportion of persons with less than 12th grade education), high education (proportion of persons with at least 4 years of college), unemployment, and low working class were employed.[24, 25] For each calendar year, the 5-year estimation referred to data collected over the last 5-year. For example, in 2010, its 5-year estimation referred to data collected from 2006 to 2010. These variables have shown good validity in capturing the sociodemographic characteristics of individuals at local areas.[9, 26]

Contextual factors

Contextual factors were defined as the broader social, economic, and physical opportunities in a region. To assess the socioeconomic status and opportunities, 5-year estimated factors, such as Gini index, vacant houses, transportation accessibility, crowding, and rental houses, were also extracted from ACS.[24, 27] In addition, geolocations of Ryan White HIV healthcare centers and community mental health centers were extracted from US Department of Health and Human Services (DHHS) Data Warehouse and health department websites in SC and its neighboring states.[28–31] According to the definition of accessible healthcare facility provided by the US DHHS[32], these data were used to develop a measure of healthcare access opportunity in terms of the numbers of accessible HIV healthcare and mental health centers within 25 miles radius of each county in SC.

Collective factors

According to the sociological framework of health, collective factors referred to the socio-cultural features of a region, including safety, religion, and social capital.[19] Percentages of violent crime (i.e., murder, rape, and assault) and non-violent crime (i.e., robbery, burglary, theft, and motor vehicle theft) by each county between 2010 and 2014 were extracted from the US Crime Data (UCR) of Social Explorer website.[20, 33, 34] Missing values in 2015 and 2016 were imputed using neighboring data in 2014. In terms of the religious environment, percentage of religious adherents by county in 2010 was obtained from the US Religion Data.[19, 20, 35] Finally, county-level social capital was assessed using the social capital index from the 2018 US Congress Joint Economic Committee.[36] The social capital index includes four dimensions, such as family unity, community health, institutional health, and collective efficacy.[36] All these four dimensions of social capital index by each county were extracted from the U.S. Congress Joint Economic Committee and used as constant values during the analysis.[36]

Table 1 shows the detailed description and cutoff of each predictor. The cutoffs of regional SES and economic opportunities (e.g., unemployment, working class, low income, high income, and wealth) were determined based on the 2010 US census.[37] As US DHHS suggested that the 25-mile was the low travel distance of geographic access to primary care, we used it as a cutoff to define the accessible healthcare facilities at local areas.[32]

Data linkage and management

First, county-level RIC rate was calculated in each calendar year according to the CDC's definition. The county-level RIC rate was categorized according to the national average level and to create the county-level RIC status (> 56.0% or 56.0%). Second, all proposed county-level predictors were extracted from multiple public datasets with the Federal Information Processing Standards (FIPS) as unique identity for each county. Predictors, such as low income, low education, transportation accessibility, and crowding, were extracted from the relevant public datasets based on their definitions and cutoffs recommended by previous studies.[37] Other predictors, such as accessible healthcare facilities, were calculated using the extracted predictors from the official webpages. Finally, all county-level predictors were linked to the RIC rate by the FIPS code and calendar year. Missing values were imputed using the neighboring data in previous year.

Statistical analysis

First, the geospatial distribution of RIC was examined by every other year (i.e., 2010, 2012, 2014, and 2016) through mapping the county-level RIC rate among PLWH. Maps were created using ArcGIS version 10.2 (Environmental Systems Research Institute, CA). Heat map was used to describe the temporal trend of RIC across the 46 counties during the study period. Second, descriptive statistics were reported on all county-level factors using three quantiles (i.e., 25th percentile, 50th percentile, and 75th percentile) and interquartile range (IQR) for relevant time frames. Changes of key variables were examined using Kruskal-Wallis test. Third, generalized linear mixed effect regression with least absolute shrinkage and selection operator (LASSO) was employed to examine whether these variables of interest had significant impacts on the change of county-level RIC status from 2010 to 2016. [38] With the longitudinal observations at 46 counties across 7 years, the Backtest approach was employed to split the dataset into training and testing sets according to the temporal order.[39] Therefore, data between 2010 and 2015 were used as training set for model preparation while that in 2016 was used as testing set for model evaluation.

In generalized linear mixed effect regression with LASSO, the random effect was used to account for the repeated measures at the same county. As logit function could provide prediction for binary outcome, we used this link function during the data analysis. Tuning parameter was chosen based on the small prediction error rate using 10-fold cross-validation (Table 1 in Appendix). Model with smaller prediction error rate and better conceptual meaning was chosen as the final model. Confusion matrix and area under the curve (AUC) statistics were used to evaluate the model performance using data of 2016.

All analyses were conducted using R version 3.6.3 (The R Foundation). Generalized linear mixed effect model with LASSO was conducted using the R package "glmmLasso". Statistical significance was set at $p < 0.05$, and all tests were two tailed.

Results

Descriptive statistics and changes of key variables

Among the 46 counties in SC, the median of RIC rate increased from 2010 (55.23%) to 2012 (57.75%) and decreased from 2012 (57.75%) to 2016 (54.40%). The change of county-level RIC rate across the time was not significant. There were 20, 28, 22, and 18 counties with their RIC rates higher than the national average level reported by the CDC in 2010, 2012, 2014, and 2016, respectively.

Table 2 also describes the information and changes of county-level predictors in SC. Across the counties in SC, nearly half of the population were male, and more than 70.0% of the population were older than 18 years. Large county variation existed in the proportion of Black population, with the IRQs being 22.43% (25th percentile: 24.99%; 75th percentile: 47.42%) in 2010 and 21.72% (25th percentile: 24.76%; 75th percentile: 46.48%) in 2014. The variation (IRQs) of poverty, low income, low education, high education, low working class, vacant houses, and rental houses were lower than 10.0% but higher than 5.0%.

Spatial distribution of county-level retention in care

Figure 1 shows the spatial distribution of county-level RIC rates in SC in 2010, 2012, 2014, and 2016, while Figure 2 illustrates the temporal change of RIC rates by county from 2010 to 2016. Spatial variation in RIC across the 46 counties were identified. Counties of the northern (i.e., Spartanburg, Union, Chester, and Laurens), central (i.e., Calhoun), eastern (i.e., Florence and Darlington), and western (i.e., McCormick and Barnwell) areas in SC had consistently higher RIC rates than the national average level in 2010, 2012, 2014, and 2016, respectively. Anderson, Cherokee, Newberry, Orangeburg, Chesterfield, Horry, Georgetown, and Marlboro counties had consistently increased RIC rates across the time. Edge-field, Aiken, Beaufort, and Saluda counties had consistently lower than 50.0% RIC rates in the selected time points.

Generalized linear mixed effect model with least absolute shrinkage and selection operator

Before the multivariate analysis, time effect (i.e., linear and quadratic trends) was tested in the crude model. Quadratic trend of county-level RIC status was significant (crude OR=0.91, 95%CI: 0.83~0.99) although the linear trend was not significant (crude OR=1.90, 95%CI: 0.95~3.77). In multivariate analysis, 22 county-level predictors were selected by LASSO. Particularly, both linear and quadratic trends were not detected by the LASSO, which meant that time effect did not explain the variation in RIC status from 2010 to 2015. Results of confusion matrix (Table 2 in Appendix) indicated that the sensitivity, specificity, and prediction accuracy were 0.83, 0.71 and 0.76, respectively. Testing this model with 2016 data, the AUC statistics was 0.84 (Figure 1 in Appendix).

Results of generalized linear mixed effect model with LASSO found that accessible numbers of mental health centers (AOR=2.00, 95%CI: 1.37~2.91) was significantly and positively associated with county-level RIC status. Proportions of male (AOR= 0.05, 95%CI: 0.01~0.47), persons with at least four-year college (AOR=0.71, 95%CI: 0.58~0.87), and collective efficacy (AOR=0.04, 95%CI: 0.01~0.36) were negatively related to county-level RIC status.

Discussion

The current study investigated the spatial variation in county-level RIC among PLWH in SC from 2010 to 2016 and identified predictors contributing to these variation using the integrated multiple public datasets and machine learning approach. Results from our study revealed that the change of county-level RIC from 2010 to 2016 was not significant, but there was clear spatial heterogeneity across the counties in SC. A total of 22 of the 31 county-level predictors were selected by the generalized linear mixed effect model with LASSO for prediction of county-level RIC status. Predictors related to healthcare access and social capital could explain the county-level variation in RIC status in SC. To the best of our knowledge, this was the first study applying sociological framework of health to select social, economic, and cultural factors associated with county-level RIC and using machine learning technique to identify the influential predictors. Findings in the current study had some strong implications in informing future community or structural efforts to improve RIC among PLWH in SC.

Clear spatial variation of RIC was identified across the 46 counties in SC. Generally, counties located at the border of SC were inclined to report low RIC rates during the study period. Additionally, counties located at the northern and southeastern SC reported higher RIC than those at the western and southern parts. There were two possible explanations. First, PLWH who lived in the counties on the border of SC could had their CD4 and viral load tested from the HIV healthcare centers of neighboring states, and their records would not be documented at the SC DHEC. For instance, PLWH in Edge-field and Aiken counties frequently go to the Medical College of George (Augusta) for their clinical check-ups, and the SC DHEC could not obtain their CD4 and viral load information from the facilities in Georgia. Therefore, these counties would have lower RIC rates than others based on the data from SC. Second, county-level disparities in healthcare resource and social capital could contribute to the spatial variation of RIC in SC.

Healthcare access could significantly promote county-level RIC status in SC. We used accessible mental health centers within 25-mile radius of each county to reflect one aspect of local healthcare access and found that RIC status could be promoted by the county-level accessible healthcare facilities, especially for those with accessible mental health centers. Indeed, counties (e.g., Richland and Pickens) with more accessible mental health centers inclined to have higher RIC rates among PLWH. In addition to the importance of healthcare environment at local areas, this finding emphasized the importance of mental health resources as it could provide PLWH with psychological counseling and service to intervene mental health problems (e.g., stigma, depression, and stress) pertinent to HIV infection and ART treatment, which could promote the county-level RIC.[12, 13, 40, 41]

We also found that social capital could account for the spatial disparities in county-level RIC status in SC. Collective efficacy, one dimension of social capital index, was measured using the numbers of violent crimes, with lower violent crimes indicating better collective efficacy at local areas. Counties with better collective efficacy (fewer violent crimes) reported higher RIC than the national average level. Collective efficacy is defined as the shared belief held by local community members that together they could work toward a common good.[42, 43] Extant research had found that collective efficacy had positive effect on health-seeking behavior and retention in medical appointment.[44] Additionally, collective efficacy could reflect social capital at local areas. With better collective efficacy or social capital in general, PLWH in these counties could have more social interaction with others, better access to social resources, and higher retention in HIV care.[22]

Our study showed that proportion of male at local area was negatively associated with the RIC status. This finding was consistent with that of previous studies which were conducted at the individual level.[45, 46] These studies found that males were less likely than females to retain in HIV care. The possible explanation could be that the masculine norms negatively influence males' utilization of HIV services and their RIC.[45, 46] Although some studies identified the protective effect of masculinity on health behaviors among male PLWH, the majority of other studies demonstrated that the norm of masculine could deter males' engagement in and maintaining to HIV care.[46] At the county-level, masculinity could shape the social norm at local areas and influence PLWH's health behaviors, especially in Deep South states.[47] Using qualitative study design, Hooker and colleagues found that

African American males in SC mostly expressed views of masculinity, which were featured as the leader of a family, role model, strong work ethic, and masculine.[47] Males with these features believed that they should make a living, be a provider, maintain order, and assure protection for the whole family.[47] Additionally, these features were related to other personal attributes, such as principled, compassionated, selfless, and wise.[47] However, the norms of masculine would deter males' perception of health risk, belief of internal control over health, change of unhealthy habits, and utilization of healthcare resources.[47] Therefore, under atmosphere of this culture, counties with large proportion of males were more likely to have low RIC rates.

Unexpectedly, we found that counties with lower proportion of persons with at least four-year college were more likely to have higher RIC rates than the national average. The possible explanation for this unexpected finding might be that some high RIC counties (e.g., Barnwell and Cherokee) had a small number of PLWH and low proportion of persons with a bachelor degree. For instance, Barnwell was one of the counties with the lowest proportion of persons with a bachelor degree, but the RIC rates in this county were 65.1% (71/109), 68.2% (73/107), 66.4% (75/113), and 69.8% (81/116) in 2010, 2012, 2014, and 2016, respectively. These rates were higher than the third quantile (75th percentile) of RIC rates and closed to the maximum rates across the 46 counties in the selected time points. Future studies investigating this relationship in counties with a larger population of PLWH are needed.

The findings in this study had some strong implications in developing structural and individual interventions to improve county-level RIC among PLWH. At the structural level, policy efforts and extra funding are needed to build up healthcare facilities, especially for the mental health centers. Accessible healthcare facilities could ensure PLWH to engage in and retain to HIV healthcare, and they could also provide PLWH with mental health services and intervene their psychological distress related to HIV infection and ART treatment. Additionally, structural efforts are needed to build up regional social capital, which could increase social cohesion.[43] Social cohesion is a key component of collective efficacy and could promote health behaviors as well as RIC among PLWH.[42, 43] As suggested by Butel and Braun[42], community activities, such as volunteering and community projects, could increase collective efficacy and social capital, which could lead to improved health outcomes. At the individual level, the positive relationship between accessible mental health centers and county-level RIC status also indicated the importance of mental health services, psychological counseling, and psychiatric treatment for PLWH who were in needs of help. These services should be ensured at local healthcare facilities. Additionally, due to the masculine norms in SC, mental health services and behavioral interventions should be tailored by gender and incorporate the social culture into their design as well as implementation.

Although the current study was innovative in aggregating multiple public datasets and using machine learning approach to explore county-level factors associated with county-level RIC status, there were some limitations that need to be noted. First, there might be small sample size of HIV cases in some counties which might limit the statistical power to appropriately identify factors significantly associated with county-level RIC status. Second, although we

used sociological framework of health to select the county-level factors in this study, there are still other factors which were not considered in this study. For instance, as alcohol use are closely related to mental health problems and missed medical appointments, distribution of alcohol outlet by county may be associated with RIC and should be explored in the future research.[48] Third, we defined county-level social, economic, and cultural predictors based on the sociological framework. However, population subtype factors pertinent to HIV epidemic (e.g., routes of HIV infection, proportions of survival or death) could also predict county-level RIC status and provide further implications on HIV prevention specificity at the structural and geographic levels across SC. Future studies investigating the prediction efficacy of these factors are needed. Fourth, the PLWH who lived in the counties on the border of SC might seek healthcare from Georgia and North Carolina and SC DHEC does not have access to their medical records in other states, which could lead to low county-level RIC rates in those counties. Fifth, some of the county-level predictors were only available at specific time points, which would bias the parameter estimations of final model. For example, proportions of adherents at local areas were assessed every 10-year, and we used data from 2010 and designed it as a constant value during the analysis. Finally, the current study only explored county-level factors associated with RIC. In the future, studies incorporating both individual-level and county-level factors are needed, which could explore how county-level factors interact with individual-level factors to influence RIC among PLWH.

Conclusion

Although much efforts had been made on curtailing HIV epidemic in the past decades, there were still more than half of the counties in SC where RIC rates were lower than the national average level. Spatial variation in RIC could be identified using geospatial mapping, and county-level factors associated with accessible healthcare facilities and social capital made significant contribution to these variation. To improve county-level RIC in SC and other Southern US states, structural and individual interventions targeting these influential factors are needed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Reference

1. Bulsara SM, Wainberg ML, Newton-John TRO. Predictors of adult retention in HIV care: A systematic review. AIDS Behav 2018; 22(3):752–764. [PubMed: 27990582]

2. World Health Organization. Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection: Recommendations for A Public Health Approach. 2016. Available at: http://apps.who.int/iris/bitstream/10665/208825/1/9789241549684_eng.pdf?ua=1

3. Centers for Disease Control and Prevention. Vital signs: HIV prevention through care and treatment —United States, 2016. Morbidity and Mortality Weekly Report (MMWR) 2019; 68(11):267–271. [PubMed: 30897075]

4. Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data United States and 6 dependent areas, 2018. HIV Surveillance Supplemental Report 2020; 25(2).

5. Edun B, Iyer M, Albrecht H, Weissman S. The South Carolina HIV cascade of care. South Med J 2015; 108(11):670–674. [PubMed: 26539946]

6. South Carolina Department of Health and Environment Control. An Epidemiologic Profile of HIV and AIDS in South Carolina 2019. Available at: https://www.dhec.sc.gov/sites/default/files/media/document/2019-Epi-Profile.pdf

7. Momplaisir FM, Storm DS, Nkwihoreze H, Jayeola O, Jemmott JB. Improving postpartum retention in care for women living with HIV in the United States. AIDS 2018; 32(2):133–142. [PubMed: 29194122]

8. Anderson AN, Higgins CM, Haardorfer R, Holstad MM, Nguyen MLT, Waldrop-Valverde D. Disparities in retention in care among adults living with HIV/AIDS: A systematic review. AIDS Behav 2020; 24(4):985–997. [PubMed: 31555931]

9. Althoff AL, Zelenev A, Meyer JP, Fu J, Brown SE, Vagenas P, et al. Correlates of retention in HIV care after release from jail: results from a multi-site study. AIDS Behav 2013; 17 Suppl 2`:S156–170. [PubMed: 23161210]

10. Nosyk B, Lourenco L, Min JE, Shopin D, Lima VD, Montaner JS, et al. Characterizing retention in HAART as a recurrent event process: insights into 'cascade churn'. AIDS 2015; 29(13):1681–1689. [PubMed: 26372279]

11. Rebeiro P, Althoff KN, Buchacz K, Gill J, Horberg M, Krentz H, et al. Retention among North American HIV-infected persons in clinical care, 2000–2008. J Acquir Immune Defic Syndr 2013; 62(3):356–362. [PubMed: 23242158]

12. Tobias CR, Cunningham W, Cabral HD, Cunningham CO, Eldred L, Naar-King S, et al. Living with HIV but without medical care: barriers to engagement. AIDS Patient Care STDS 2007; 21(6):426–434. [PubMed: 17594252]

13. Blank AE, Fletcher J, Verdecias N, Garcia I, Blackstock O, Cunningham C. Factors associated with retention and viral suppression among a cohort of HIV+ women of color. AIDS Patient Care STDS 2015; 29 Suppl 1:S27–35. [PubMed: 25458205]

14. Hanna DB, Selik RM, Tang T, Gange SJ. Disparities among US states in HIV-related mortality in persons with HIV infection, 2001–2007. AIDS 2012; 26(1):95–103. [PubMed: 22008659]

15. Hixson BA, Omer SB, del Rio C, Frew PM. Spatial clustering of HIV prevalence in Atlanta, Georgia and population characteristics associated with case concentrations. J Urban Health 2011; 88(1):129–141. [PubMed: 21249526]

16. Namosha E, Sartorius B, Tanser F. Spatial clustering of all-cause and HIV-related mortality in a rural South African population (2000–2006). PLoS One 2013; 8(7):e69279. [PubMed: 23935972]

17. Tanser F, Barnighausen T, Cooke GS, Newell ML. Localized spatial clustering of HIV infections in a widely disseminated rural South African epidemic. Int J Epidemiol 2009; 38(4):1008–1016. [PubMed: 19261659]

18. Rebeiro PF, Gange SJ, Horberg MA, Abraham AG, Napravnik S, Samji H, et al. Geographic variations in retention in care among HIV-infected adults in the United States. PLoS One 2016; 11(1):e0146119. [PubMed: 26752637]

19. Macintyre S, Ellaway A, Cummins S. Place effects on health: How can we conceptualise, operationalise and measure them? Soc Sci & Med 2002; 55(1):125–139. [PubMed: 12137182]

20. Collins J, Ward BM, Snow P, Kippen S, Judd F. Compositional, contextual, and collective community factors in mental health and well-being in Australian rural communities. Qual Health Res 2017; 27(5):677–687. [PubMed: 26848083]

21. Eberhart MG, Yehia BR, Hillier A, Voytek CD, Fiore DJ, Blank M, et al. Individual and community factors associated with geographic clusters of poor HIV care retention and poor viral suppression. J Acquir Immune Defic Syndr 2015; 69 Suppl 1:S37–43. [PubMed: 25867777]

22. Medved Kendrick H Are religion and spirituality barriers or facilitators to treatment for HIV: a systematic review of the literature. AIDS Care 2017; 29(1):1–13. [PubMed: 27410058]

23. National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention. Understanding the HIV Care Continuum. 2019. Available at: https://www.cdc.gov/hiv/pdf/library/factsheets/cdc-hiv-care-continuum.pdf

24. Niyonsenga T, Trepka MJ, Lieb S, Maddox LM. Measuring socioeconomic inequality in the incidence of AIDS: Rural-urban considerations. AIDS Behav 2013; 17(2):700–709. [PubMed: 22711226]

25. Roux AVD, Kiefe CI, Jacobs DR, Haan M, Jackson SA, Nieto FJ, et al. Area characteristics and individual-level socioeconomic position indicators in three population-based epidemiologic studies. Ann Epidemiol 2001; 11(6):395–405. [PubMed: 11454499]

26. Card KG, Lachowsky NJ, Althoff KN, Schafer K, Hogg RS, Montaner JSG. A systematic review of the geospatial barriers to antiretroviral initiation, adherence and viral suppression among people living with HIV. Sex Health 2019; 16(1):1–17 [PubMed: 30409243]

27. Centers for Disease Control and Prevention. HIV Surveillance Report: Social Determinants of Health among Adults with Diagnosed HIV Infection, 2016. HIV Surveillance Supplemental Report 2018; 23(6): pt A& pt B. Available at: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2016-vol-23-6.pdf

28. U.S. Department of Health and Human Services Data Warehouse. Find a Ryan White HIV/AIDS Program Medical Provider. Available at: https://findhivcare.hrsa.gov/

29. The Georgia Collaborative ASO. Find Agencies and Programs. Available at: https://www.georgiacollaborative.com/

30. NC Department of Health and Human Services. Available at: https://www.ncdhhs.gov/

31. South Carolina Department of Mental Health. Mental Health Centers and Clinics. Available at: https://scdmh.net/welcome/mental-health-centers/

32. Serban N Healthcare System Access: Measurement, Inference, and Intervention. John Wiley & Sons, Inc. 2020.

33. U.S. Crime Data. Retrieved from the Social Explorer. Available at: https://www.socialexplorer.com/explore-tables.

34. Ojikutu BO, Bogart LM, Klein DJ, Galvan FH, Wagner GJ. Neighborhood crime and sexual transmission risk behavior among Black men living with HIV. J Health Care Poor Underserved 2018; 29(1):383–399. [PubMed: 29503307]

35. U.S. Religious Data. Retrieved from the Social Explorer. Available at: https://www.socialexplorer.com/explore-tables.

36. U.S. Congress Joint Economic Committee. The Geography of Social Capital in America 2018. Available at: https://www.jec.senate.gov/public/index.cfm/republicans/2018/4/the-geography-of-social-capital-in-america

37. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter?: the Public Health Disparities Geocoding Project. Am J Epidemiol 2002; 156(5):471–482. [PubMed: 12196317]

38. Schelldorfer J, Meier L, Bühlmann P. GLMMLasso: An algorithm for high-dimensional generalized linear mixed models using ℓ_1-penalization. Journal of Computational and Graphical Statistics 2014; 23(2):460–477.

39. Brownlee J How to backtest machine learning models for time series forecasting? Time Series 2016. Available at: https://machinelearningmastery.com/backtest-machine-learning-models-time-series-forecasting/

40. Lall P, Lim SH, Khairuddin N, Kamarulzaman A. Review: an urgent need for research on factors impacting adherence to and retention in care among HIV-positive youth and adolescents from key populations. J Int AIDS Soc 2015; 18(2 Suppl 1):19393. [PubMed: 25724503]

41. Sevelius JM, Patouhas E, Keatley JG, Johnson MO. Barriers and facilitators to engagement and retention in care among transgender women living with human immunodeficiency virus. Ann Behav Med 2014; 47(1):5–16. [PubMed: 24317955]

42. Butel J, Braun KL. The role of collective rfficacy in reducing health disparities: A systematic review. Fam Community Health 2019; 42(1):8–19. [PubMed: 30431465]

43. Cain D, Pitpitan EV, Eaton L, Carey KB, Carey MP, Mehlomakulu V, et al. Collective efficacy and HIV prevention in South African townships. J Community Health 2013; 38(5):885–893. [PubMed: 23660646]

44. Browning CR, Burrington LA, Leventhal T, Brooks-Gunn J. Neighborhood structural inequality, collective efficacy, and sexual risk behavior among urban youth. J Health Soc Behav 2008; 49(3):269–285. [PubMed: 18771063]

45. Skovdal M, Campbell C, Madanhire C, Mupambireyi Z, Nyamukapa C, Gregson S. Masculinity as a barrier to men's use of HIV services in Zimbabwe. Global Health 2011; 7:13. [PubMed: 21575149]

46. Sileo KM, Fielding-Miller R, Dworkin SL, Fleming PJ. A scoping review on the role of masculine norms in men's engagement in the HIV care continuum in sub-Saharan Africa. AIDS Care 2019; 31(11):1435–1446. [PubMed: 30909724]

47. Hooker SP, Wilcox S, Burroughs EL, Rheaume CE, Courtenay W. The potential influence of masculine identity on health-improving behavior in midlife and older African American men. J Mens Health 2012; 9(2):79–88. [PubMed: 23459337]

48. Monroe AK, Lau B, Mugavero MJ, Mathews WC, Mayer KH, Napravnik S, et al. Heavy alcohol use is associated with worse retention in HIV Care. J Acquir Immune Defic Syndr 2016; 73(4):419–425. [PubMed: 27243904]
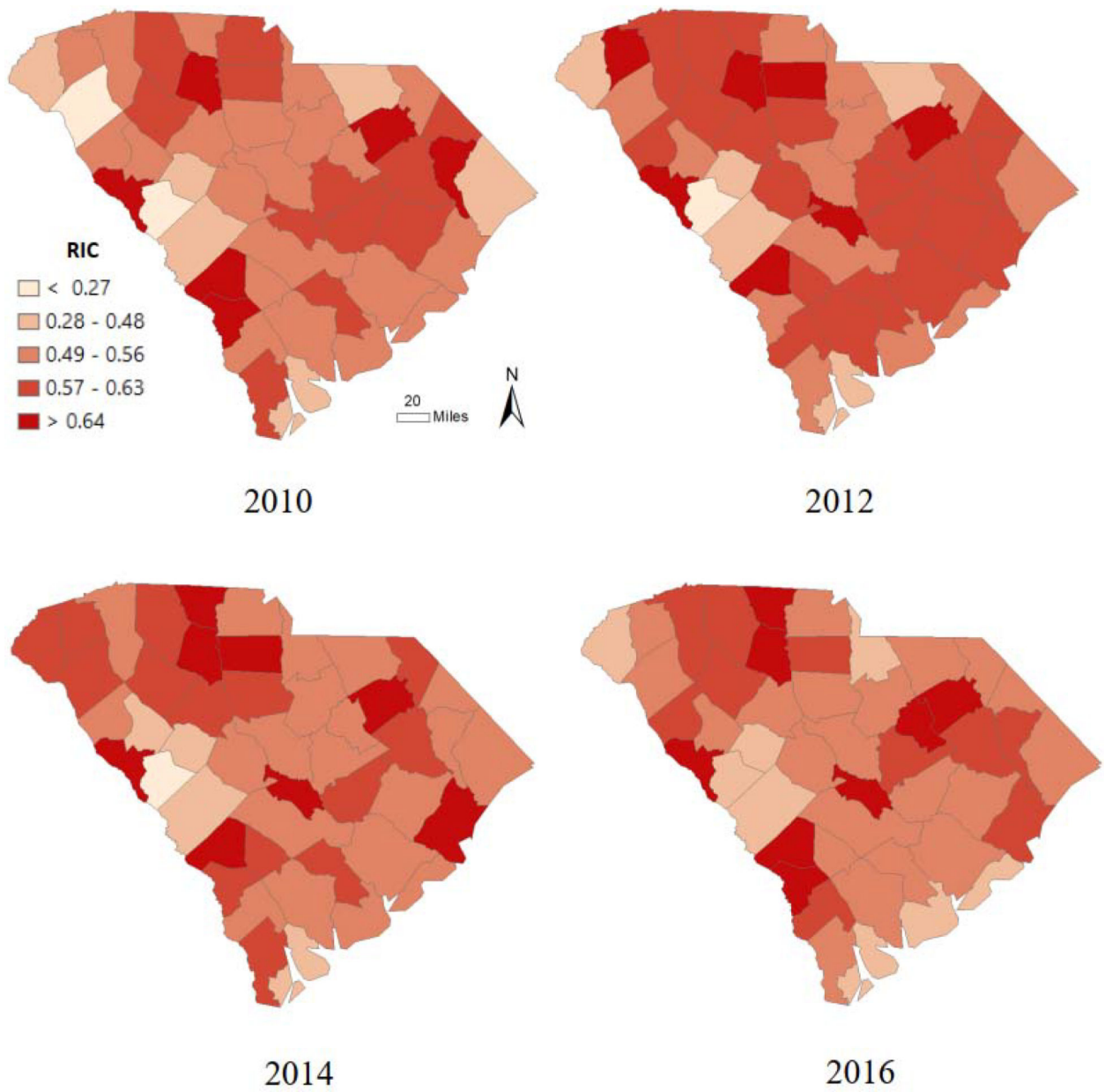
Figure 1.
Spatial distribution and temporal trend of county-level retention in care among PLWH in SC in 2010, 2012, 2014, and 2016
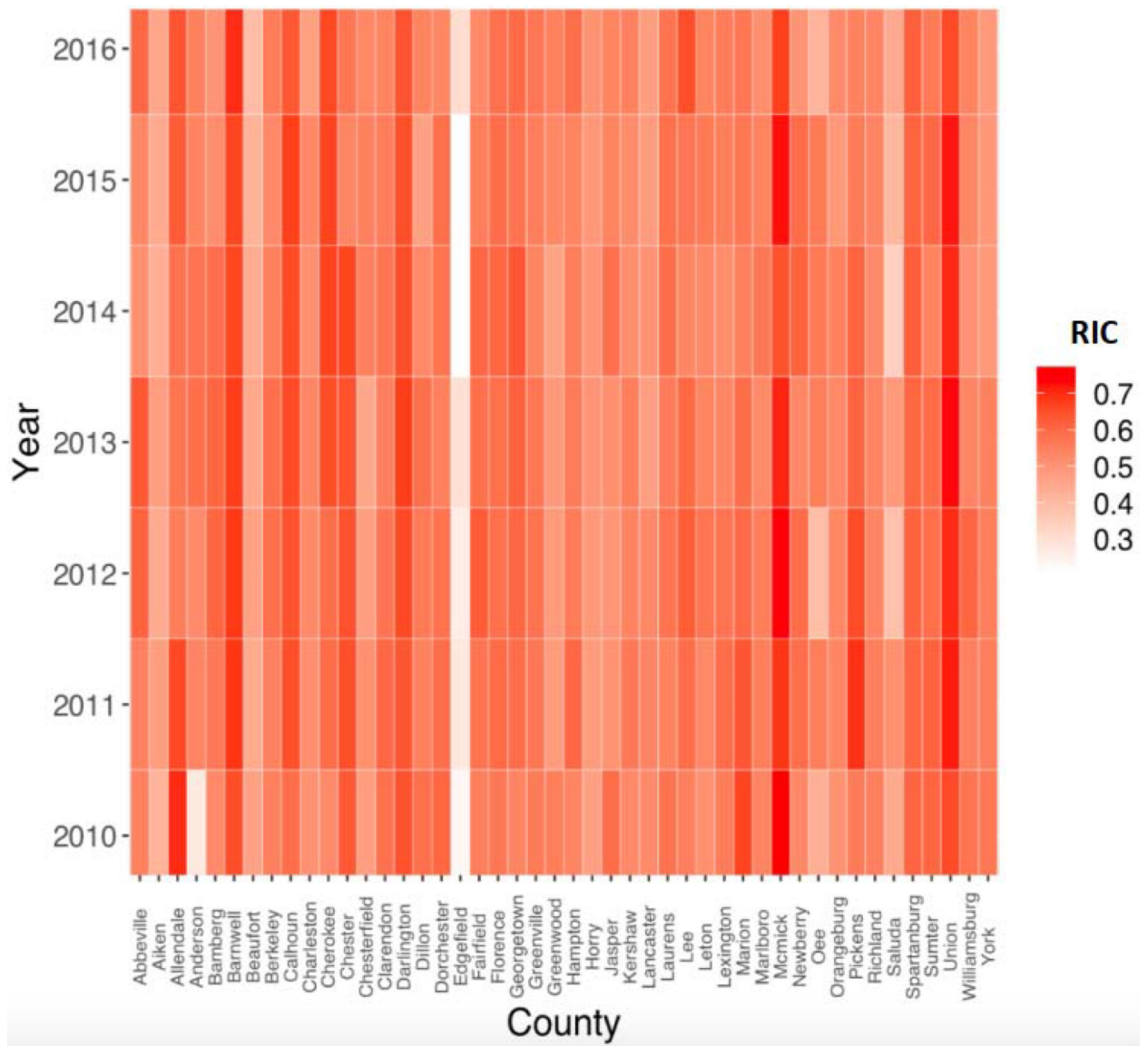Note: PLWH: People living with HIV; RIC: Retention in care

Figure 2.
Heatmap of county-level retention in care among PLWH in SC from 2010 to 2016
Note: PLWH: People living with HIV; RIC: Retention in care

Table 1.

Definitions and cut-offs of county-level predictors

Variables	Definitions
Compositional factors	
Population density (per square mile)	Numbers of population per square mile
Male (%)	Proportion of male
Age (18, %)	Proportion of persons older than 18 years
Living alone (%)	Proportion of persons living alone
Black (%)	Proportion of Black population
Public assistance (%)	Proportion of households with public assistance income
Poverty (%)	Proportion of population aged 18 to 64 for whom living in poverty (below the federally-define poverty line)
Wealth (%)	Proportion of owner-occupied homes worth >=$750,000 (400% of the median value of owned homes) [25]
Median income ($)	Median household annual income
Low income (%)	Proportion of households with an income < $25,000 (50% of the US median household income) [25]
High income (%)	Proportion of households with an income > $200,000 (400% of the US median household income) [25]
Low education (%)	Proportion of persons aged >=25 years and older, with less than 12th grade education [25]
High education (%)	Proportion of persons aged >=25 years and older, with at least 4 years of college [25]
Unemployment (%)	Proportion of persons aged 16 years older in the labor force who are unemployed
Low working class (%)	Proportion of persons in specific occupations, such as service occupations, sales and office occupations, farming, fishing and forestry occupations, construction, extraction and maintenance occupations, production transportation and material moving occupations [24, 25]
No health insurance coverage (%)	Proportion of persons with no health insurance coverage
Contextual factors	
Vacant houses (%)	Proportion of vacant housing units
Transportation accessibility (%)	Proportion of occupied housing units without access to a vehicle
Crowding (%)	Proportion of occupied housing units with more than one occupants per room
Rental house (%)	Proportion of housing units for rent
Gini index	A measure of statistical dispersion represents the income inequality or wealth inequality within a group of people at local areas
Primary care provider rate (per 100,000 people)	Numbers of primary care provider per 100,000 population
Accessible health facilities (*n*)	
Ryan White HIV centers	Numbers of Ryan White HIV centers within 25 miles radius of each county
Mental health centers	Numbers of mental health centers within 25 miles radius of each county

Variables	Definitions
Collective factors	
Violent crime (%)	Proportion of violent crime (i.e., murder, rape, assault)
Non-violent crime (%)	Proportion of non-violent crime (i.e., robbery, burglary, theft, and motor vehicle theft)
Adherents (%)	Proportion of adherents
Family unity	Proportion of births to unmarried women, women current married, and children with single parent [36]
Community health	Numbers of non-religious, non-profit organization, religious congregation per 1,000 people, and informal civic engagement [36]
Institutional health	Presidential election voting rate, mail-back census response rate, and confidence in institutions [36]
Collective efficacy	Violent crimes per 100,000 people [36]

Table 2.

Descriptive statistics and changes of RIC and county-level predictors

Variables	25th percentile	50th percentile	75th percentile	IQR	p-value
Number of county	12	23	35	–	
RIC rate (%)					0.579
2010	12 (52.24)	23 (55.23)	34 (58.82)	6.59	
2012	11 (52.67)	23 (57.75)	35 (60.83)	8.15	
2014	12 (51.43)	23 (55.20)	35 (60.87)	9.44	
2016	12 (51.59)	24 (54.40)	34 (58.82)	7.24	
Predictors					
Compositional factors					
Population density (per square mile)					1.000
2010	48.82	80.51	166.63	120.81	
2014	45.61	80.20	172.58	126.97	
Male (%)					0.999
2010	48.25	48.58	49.58	1.33	
2014	47.85	48.57	49.26	1.41	
Age (18, %)					**<0.001**[***]
2010	75.12	76.23	76.95	1.83	
2014	76.08	77.28	78.81	2.73	
Living alone (%)					0.318
2010	25.89	27.22	29.90	4.01	
2014	26.32	27.66	29.59	3.27	
Black (%)					1.000
2010	24.99	33.51	47.42	22.43	
2014	24.76	33.04	46.48	21.72	
Public assistance (%)					0.075
2010	1.36	1.58	2.04	0.68	
2014	1.27	1.65	2.06	0.79	
Poverty (%)					0.057

Variables		25th percentile	50th percentile	75th percentile	IQR	p-value
	2010	14.21	17.31	19.55	5.34	
	2014	16.60	20.39	22.90	6.30	0.954
Wealth (%)						
	2010	0.71	1.58	2.24	1.53	
	2014	1.07	1.55	2.30	1.23	0.959
Median income ($)						
	2010	33,066	38,588	42,871	9,805	
	2014	33,615	38,609.5	43,203	9,588	0.850
Low income (%)						
	2010	29.13	32.67	39.29	10.16	
	2014	28.63	32.41	37.98	9.35	0.360
High income (%)						
	2010	0.72	1.26	2.05	1.33	
	2014	0.97	1.52	2.39	1.42	0.038 *
Low education (%)						
	2010	17.39	21.74	24.68	7.29	
	2014	15.32	19.33	21.84	6.52	0.877
High education (%)						
	2010	12.88	16.52	21.81	8.93	
	2014	13.03	18.06	22.41	9.38	**<0.001** ***
Unemployment (%)						
	2010	8.88	10.47	12.92	4.04	
	2014	10.35	11.97	14.03	3.68	0.899
Low working class (%)						
	2010	67.47	70.00	74.12	6.65	
	2014	66.17	70.05	72.84	6.67	**<0.001** ***
No health insurance coverage‡ (%)						
	2012	15.50	16.95	18.70	3.20	
	2014	15.17	16.34	18.26	3.09	
Contextual factors						

Variables	25th percentile	50th percentile	75th percentile	IQR	p-value
Vacant houses (%)					1.000
2010	12.79	16.67	20.69	7.90	
2014	13.30	16.45	20.78	7.48	
Transportation accessibility (%)					0.968
2010	6.52	8.52	10.32	3.80	
2014	6.30	8.04	10.01	3.71	
Crowding (%)					0.856
2010	1.44	1.76	2.56	1.12	
2014	1.54	1.89	2.49	0.95	
Rental house (%)					0.929
2010	25.56	27.27	31.54	5.98	
2014	25.81	28.89	31.74	5.93	
Gini index					0.294
2010	0.44	0.46	0.47	0.03	
2014	0.45	0.46	0.48	0.03	
Primary care provider rate (per 100,000 people)					**<0.001** ***
2010	61.69	78.76	105.93	44.23	
2014	35.78	48.12	58.93	23.15	
Accessible health facilities [†] (n)					
Ryan White HIV centers	1	1	3	2	–
Mental health centers	3	3	4	1	–
Collective factors					
Violent crime (%)					**0.015** *
2010	0.41	0.49	0.65	0.24	
2014	0.31	0.41	0.50	0.19	
Non-violent crime (%)					0.062
2010	3.17	3.75	4.62	1.45	
2014	2.75	3.32	3.90	1.15	
Adherents [†] (%)	41.20	53.55	59.40	18.20	–

Variables	25th percentile	50th percentile	75th percentile	IQR	p-value
Family unity [†]	−1.81	−1.16	−0.51	1.30	–
Community health [†]	−0.79	−0.55	−0.31	0.48	–
Institutional health [†]	−0.05	0.25	0.38	0.43	–
Collective efficacy [†]	406.7	503.0	629.2	222.5	–

Note:

*
: p<0.05

***
: p<0.001

RIC: Retention in care; IQR: Interquartile range

[‡]: Variable was only available since 2012

[†]: Variables were only available in one-year and used as constant during the analysis.

Table 3.

Factors associated with retention in care status across the counties in South Carolina: LASSO-based predictor selection from multiple datasets

Predictors	Crude OR	Adjusted OR
Compositional factors		
Year (linear trend)	1.90 (0.95~3.77)	N/A
Year *Year (quadratic trend)	**0.91 (0.83~0.99)** [*]	N/A
Population density	0.06 (0.00~2.09)	1.01 (1.00~1.01)
Male	0.50 (0.01~20.14)	**0.05 (0.01~0.47)** [**]
Age	0.10 (0.00~6.71)	N/A
Living alone	0.78 (0.07~8.70)	0.20 (0.01~3.20)
Black	1.03 (0.99~1.08)	0.56 (0.02~14.85)
Public assistance	1.28 (0.12~13.46)	0.19 (0.02~1.49)
Poverty	0.81 (0.02~38.34)	N/A
Wealth	0.02 (0.00~3.14)	1.19 (0.95~1.50)
Median income	**0.02 (0.01~0.69)** [*]	1.00 (0.99~1.00)
Low income	1.09 (0.99~1.19)	0.88 (0.70~1.10)
High income	**0.01 (0.00~0.39)** [*]	1.06 (0.47~2.35)
Low education	**1.13 (1.00~1.28)** [***]	0.38 (0.01~26.82)
High education	**0.85 (0.76~0.96)** [**]	**0.71 (0.58~0.87)** [**]
Unemployment	1.15 (0.98~1.34)	1.12 (0.94~1.32)
Low working class	1.12 (0.98~1.28)	0.05 (0.00~3.27)
No health insurance coverage	2.76 (0.21~36.26)	0.10 (0.01~1.40)
Contextual factors		
Vacant house	1.31 (0.04~42.98)	N/A
Transportation	1.21 (0.98~1.49)	1.14 (0.84~1.55)
Crowding	0.30 (0.03–3.68)	0.89 (0.09~9.25)
Rental house	0.30 (0.01~6.55)	N/A
Gini index	0.46 (0.01~16.60)	N/A
Primary care provider rate	0.25 (0.01~11.43)	N/A
Accessible Ryan White HIV care centers	0.20 (0.02~2.62)	0.28 (0.06~1.42)
Accessible mental health centers	1.69 (0.95~3.02)	**2.00 (1.37~2.91)** [***]
Collective factors		
Violent crime	1.61 (0.93~2.79)	N/A
Non-violent crime	3.59 (0.35~37.31)	N/A
Adherents	1.04 (0.97~1.12)	N/A
Family unity	**0.04 (0.00~0.68)** [*]	0.06 (0.00~1.23)
Community health	0.56 (0.15~2.04)	0.33 (0.00~25.94)
Institution health	2.36 (0.154~36.18)	2.10 (0.31~14.29)

Predictors	Crude OR	Adjusted OR
Collective efficacy	1.19 (0.03~46.10)	**0.04 (0.01~0.36)** [**]

Notes:

OR: Odds ratio. N/A: Variables were not selected by the LASSO penalty in multivariate analysis.

[*] : $p<0.05$

[**] : $p<0.01$

[***] : $p<0.001$.