



# **Association of longitudinal patterns of CD4, Viral load and Non-AIDS Related Cancer**

**Shujie Chen**

Department of Epidemiology and Biostatistics  
Norman J Arnold School of Public Health  
University of South Carolina



## Outline:

- Background
- Data source
- Statistical analysis
- Results
- Conclusion
- Future Study



## Outline:

- Background
- Data source
- Statistical analysis
- Results
- Conclusion
- Future Study



# HIV and Cancer

- A growing of aged People Living with HIV/AIDs (PLWH) population.
  - Attributed to the increasing efficacy of combination antiretroviral therapy (ART), there is a change of epidemiology of the HIV population. (Engels et al., 2008).
- With the growing aged People Living with HIV/AIDs (PLWH) population
  - The burden of cancer among them is also projected to evolve in the nation (Dekeen et al., 2012).
  - Even AIDS defining cancers (ADC) are decreasing in prevalence, Non-AIDS defining cancers (NADC) are increasing (Meredith et al., 2017).



# Non-AIDS Defining Cancers (NADCs)

- Non-AIDS-defining cancers (NADCs)
  - All cancers except the ADC including Hodgkin lymphoma and cancers of the mouth, throat, liver, lung, and anus.
- Complexity NADC risk among PLWH
  - HIV infection
  - Traditional cancer risk factors
  - Other factors: infection with other viruses (such as hepatitis B or C virus), and heavy alcohol or tobacco use

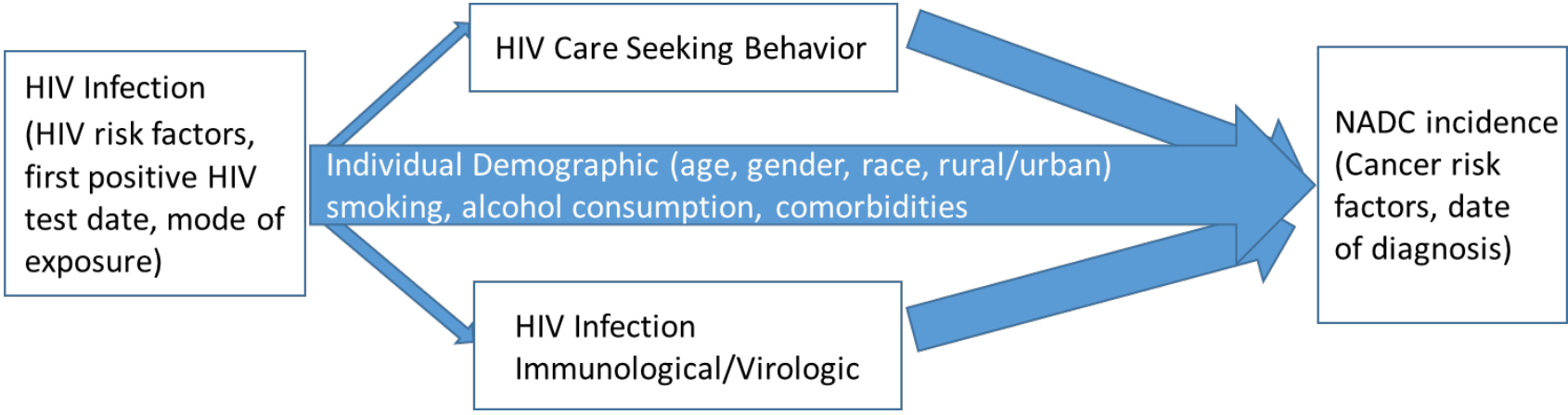


# Rationale

- South Carolina has a high incidence of HIV, as well as high rates of AIDs cases.
  - SC ranks 8th in HIV/AIDs incidence and 11th in HIV/AIDs prevalence rates nationwide;
  - the death rate of PLWH was the 7th highest in the US in 2015;
  - the cancer death rate in SC was the 14th highest in the US in 2018.
- Lack of study on NADC among PLWH in SC.



# Conceptual Model





## Outline:

- Background
- **Data source**
- Statistical analysis
- Results
- Conclusion
- Future Study





# Feasibility

- The availability of multiple key administrative and electronic health records (EHR) data sources in South Carolina (SC) provides a unique opportunity for an integrative analysis.



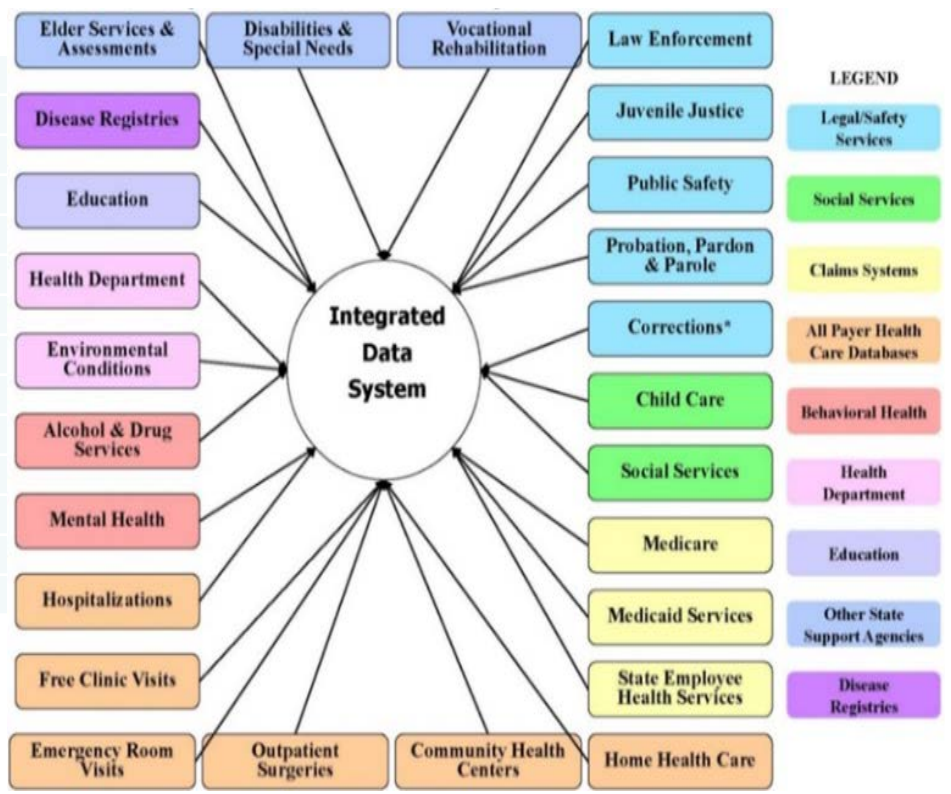
# Data Source

- S.C. Department of Health and Environmental Control (DHEC) is the government agency responsible for public health and the environment in the U.S. state of South Carolina. (Main Data source to identify PLWH)
- S.C. Revenue and Fiscal Affairs (RFA) is data warehouse in SC. RFA is responsible for providing a diverse set of economic data and research and analysis reports for the Governor, General Assembly, state and local government entities, and the private sector.



# Illustration of Raw data sources (2019)

Data source	Linker
TESTS (MAIN)	\
HSSC_COHORT	RFA_ID
HSSC_PATIENT	STUDY_ID (HSSC_LINKER)
RW	RFA_ID
CAPSS	RFA_ID
CHIP_BENEFIT	RFA_ID
CHIP_CLIENT	RFA_ID
CHIP_PARTICIPATION	RFA_ID
ALLPAYER	RFA_ID
CASES	RFA_ID
SCDC_ADMISSION	RFA_ID
SCDC_RELEASES	RFA_ID
SCDMH	RFA_ID



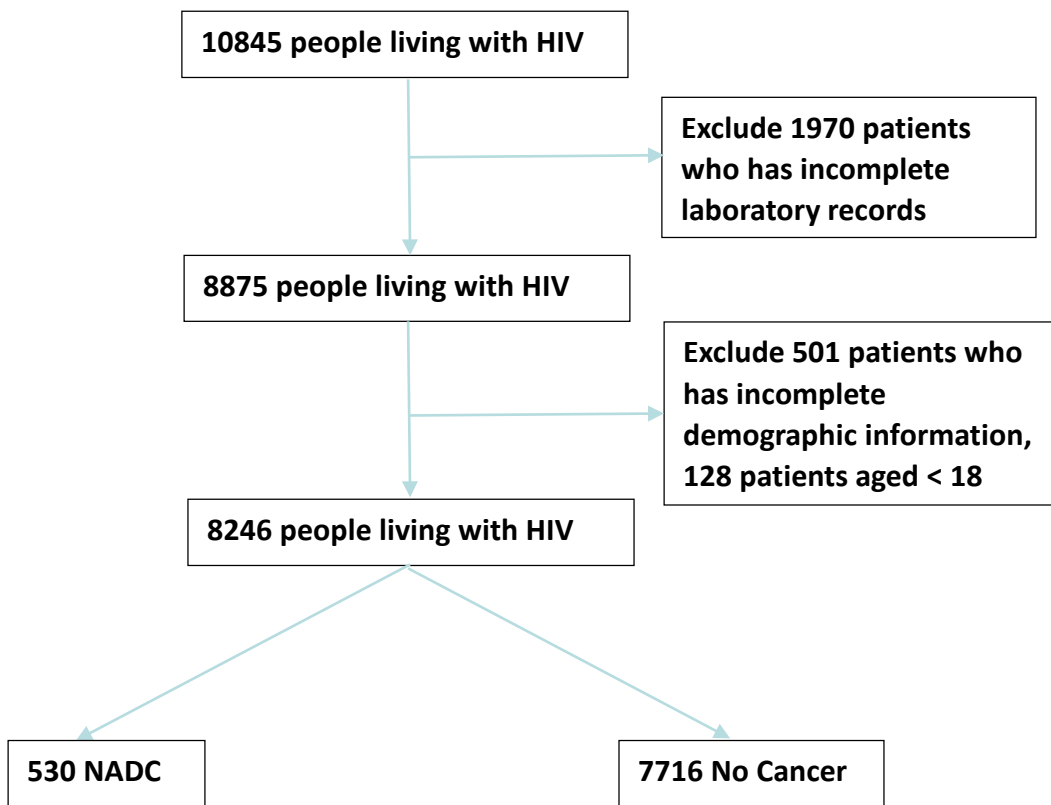


## Data Extraction for Current Study

- First, we will use demographic data from the electronic HIV/AIDS Reporting System (eHARS) from SC DHEC to define a unique PLWH cohort whose HIV/AIDS diagnosis is in 2005-2016 in SC.
- Second, we will define cancer outcome from uniformed billing data from RFA.
- Third, we will link eHARS with the outcome dataset and include longitudinal CD4 and Viral Load measure from HIV diagnosis to the event.



# Study Population



After exclusion due to missing information (Lab results of Viral load), the current study included total sample size of 8,246 over the period of 2005-2016



## Outline:

- Background
- Data source
- **Statistical analysis**
- Results
- Conclusion
- Future Study



# Hypothesis

- Determine the association between CD4 or viral load with NADC while adjusting other variables.
  - Outcome: diagnosis of NADC based on ICD-9 or ICD-10
  - Main exposure: CD4 and viral load at the initial, recent and lowest/peak



# Definition of NADC

31 specific cancer types (ICD-9 and ICD-10 Codes) were used to define NADC.

- Bone and joint, esophagus, Hodgkin lymphoma, kidney and renal pelvis, leukemias, melanoma, myeloma, ovary, testis, thyroid, uterus, anus, bladder, brain and other nervous system, breast, bronchus and lung, colorectum, gall bladder, head and neck, larynx, liver and intrahepatic bile duct, other female genital organs, pancreas, penis and other male genital organ, prostate, skin, small intestine, stomach, trachea, vagina, vulva cancer





# Variables

- Demographics
  - Age: 18-29, 30-39, 40-49, 50-59, 60+
  - Gender: females, males
  - Race/ethnicity: White, Black, Hispanic, other/unknown
  - Residence: urban/rural
    - Urban: Aiken, Anderson, Beaufort, Berkeley, Calhoun, Charleston, Chester, Darlington, Dorchester, Edgefield, Fairfield, Florence, Greenville, Horry, Jasper, Kershaw, Lancaster, Laurens, Lexington, Pickens, Richland, Saluda, Spartanburg, Sumter, Union, York
    - Rural: Abbeville, Allendale, Bamberg, Barnwell, Cherokee, Chesterfield, Clarendon, Colleton, Dillon, Georgetown, Greenwood, Hampton, Lee, Marion, Marlboro, McCormick, Newberry, Oconee, Orangeburg, Williamsburg



# Variables

- HIV disease factors:
  - Type of transmission risk: MSM, IDU, MSM & IDU, heterosexual, other/unknown.
  - CD4 count is a continuous variable, which is categorized into three levels: <200, 200-350, >350 copies/mL.
    - Initial/recent/nadir CD4: the first/last/lowest CD4 across observations for each subject.
  - Detectable is defined as viral load level (VL) $\geq$ 200 copies/mL, otherwise is undetectable.
    - Initial/recent: the first/last viral load across observations for each subject.



# Variables

- Behaviors (ICD-9 and ICD-10 code):
  - Tobacco use
  - Drug use
  - Alcohol use
  
- HIV care seeking behavior
  - Timely linkage: gap between HIV diagnosis and the first lab visit is less than or equal to 30 days.
  
- Comorbidities (ICD-9 and ICD-10 code):
  - Hepatitis B or C, Hypertension, Diabetes, Obesity, Hypothyroidism, Dyslipidemia, etc.



# Analysis Plan

- **Summary of Demographics**

- Relationship between categorical variables and NADC
  - Frequency (proportion)
  - P-value (chi-square test)

- **Statistical Modelling**

- Logistic regression
  - Compare models including initial VL along with initial/nadir/recent CD4 count
  - Compare models including initial CD4 count along with initial/recent VL level

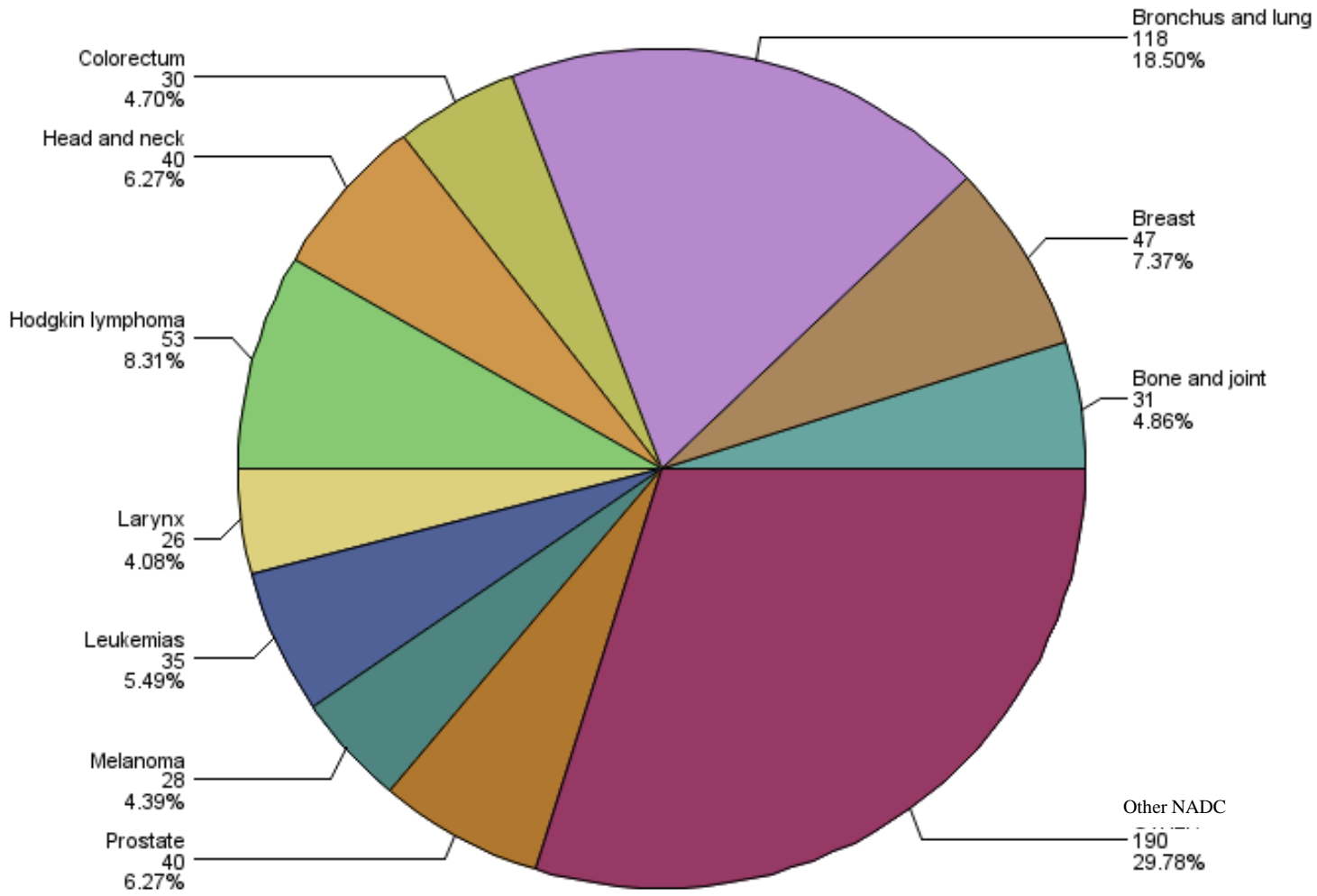


## Outline:

- Background
- Data source
- Statistical analysis
- **Results**
- Conclusion
- Future Study



# Percentage of cancer subtypes among PLWH with NADC in SC





# Results: HIV Patient Characteristics

Characteristic n (%)	Overall (n=8246)	NADC (n=530)	No NADC (n=7716)	p-value
<b>Gender</b>				<b>0.0016</b>
F	1990 (24.1)	158 (29.8)	1832 (23.7)	
M	6256 (75.9)	372 (70.2)	5884 (76.3)	
<b>Race</b>				<b>0.0004</b>
White	1863 (22.6)	134 (25.3)	1729 (22.4)	
Black	5779 (70.1)	380 (71.7)	5399 (70)	
Hispanic	438 (5.3)	8 (1.5)	430 (5.6)	
Other/unknown	166 (2)	8 (1.5)	158 (2)	
<b>Transmission Risk</b>				<b>&lt;.0001</b>
MSM	4122 (50)	174 (32.8)	3948 (51.2)	
IDU	316 (3.8)	30 (5.7)	286 (3.7)	
MSM and IDU	172 (2.1)	2 (0.4)	170 (2.2)	
Heterosexual	1741 (21.1)	170 (32.1)	1571 (20.4)	
Other/unknown	1895 (23)	154 (29.1)	1741 (22.6)	
<b>Age at HIV diagnosed</b>				<b>&lt;.0001</b>
18-29	3126 (37.9)	48 (9.1)	3078 (39.9)	
30-39	1752 (21.2)	70 (13.2)	1682 (21.8)	
40-49	1866 (22.6)	196 (37)	1670 (21.6)	
50-59	1101 (13.4)	146 (27.5)	955 (12.4)	
60+	401 (4.9)	70 (13.2)	331 (4.3)	
<b>Residence</b>				<b>0.0018</b>
Rural	1455 (17.6)	120 (22.6)	1335 (17.3)	
Urban	6791 (82.4)	410 (77.4)	6381 (82.7)	



# Results: HIV Patient Characteristics

Characteristic n (%)	Overall (n=8246)	NADC (n=530)	No NADC (n=7716)	p-value
<b>Initial VL</b>				<b>0.5633</b>
undetectable	450 (5.5)	26 (4.9)	424 (5.5)	
detectable	7796 (94.5)	504 (95.1)	7292 (94.5)	
<b>Recent VL</b>				<b>&lt;.0001</b>
undetectable	4423 (53.6)	231 (43.6)	4192 (54.3)	
detectable	3823 (46.4)	299 (56.4)	3524 (45.7)	
<b>Initial CD4 count</b>				<b>&lt;.0001</b>
<200	2678 (32.5)	232 (43.8)	2446 (31.7)	
200-350	1734 (21)	107 (20.2)	1627 (21.1)	
>350	3834 (46.5)	191 (36)	3643 (47.2)	
<b>Nadir CD4 count</b>				<b>&lt;.0001</b>
<200	3803 (46.1)	307 (57.9)	3496 (45.3)	
200-350	2059 (25)	118 (22.3)	1941 (25.2)	
>350	2384 (28.9)	105 (19.8)	2279 (29.5)	
<b>Recent CD4 count</b>				<b>&lt;.0001</b>
<200	1295 (15.7)	151 (28.5)	1144 (14.8)	
200-350	1076 (13)	107 (20.2)	969 (12.6)	
>350	5875 (71.2)	272 (51.3)	5603 (72.6)	





# Results: HIV Patient Characteristics

Characteristic n (%)	Overall (n=8246)	NADC (n=530)	No NADC (n=7716)	p-value
<b>Tobacco use</b>				<b>&lt;.0001</b>
No	4917 (59.6)	144 (27.2)	4773 (61.9)	
Yes	3329 (40.4)	386 (72.8)	2943 (38.1)	
<b>Drug use</b>				<b>&lt;.0001</b>
No	7176 (87)	398 (75.1)	6778 (87.8)	
Yes	1070 (13)	132 (24.9)	938 (12.2)	
<b>Alcohol use</b>				<b>&lt;.0001</b>
No	7396 (89.7)	420 (79.2)	6976 (90.4)	
Yes	850 (10.3)	110 (20.8)	740 (9.6)	
<b>Timely linkage</b>				<b>0.2532</b>
No	3290 (39.9)	199 (37.5)	3091 (40.1)	
Yes	4956 (60.1)	331 (62.5)	4625 (59.9)	



# Results: HIV Patient Characteristics

Characteristic n (%)	Overall (n=8246)	NADC (n=530)	No NADC (n=7716)	p-value
<b>Hepatitis B or C</b>				<.0001
No	7739 (93.9)	448 (84.5)	7291 (94.5)	
Yes	507 (6.1)	82 (15.5)	425 (5.5)	
<b>Hypothyroidism</b>				<.0001
No	6817 (82.7)	281 (53)	6536 (84.7)	
Yes	1429 (17.3)	249 (47)	1180 (15.3)	
<b>Hypertension</b>				<.0001
No	6041 (73.3)	215 (40.6)	5826 (75.5)	
Yes	2205 (26.7)	315 (59.4)	1890 (24.5)	
<b>Diabetes</b>				<.0001
No	7487 (90.8)	396 (74.7)	7091 (91.9)	
Yes	759 (9.2)	134 (25.3)	625 (8.1)	
<b>Obesity</b>				<.0001
No	7909 (95.9)	488 (92.1)	7421 (96.2)	
Yes	337 (4.1)	42 (7.9)	295 (3.8)	
<b>Dyslipidemia</b>				<.0001
No	7562 (91.7)	397 (74.9)	7165 (92.9)	
Yes	684 (8.3)	133 (25.1)	551 (7.1)	



# Summary of Characteristic table

- PLWH with NADC tend to be older, rural residence.
- PLWH with NADC are more likely to have low CD4 count and detectable viral load during follow-up.
- PLWH with NADC are more likely to have behavior like tobacco use, alcohol use, and drug use, and have comorbidities like diabetes, hypertension.



# Comparison I

- Model 1: Logistic model including initial VL level and nadir CD4 count adjusting for other covariates.
- Model 2: Logistic model including initial VL level and initial CD4 count adjusting for other covariates.
- Model 3: Logistic model including initial VL level and recent CD4 count adjusting for other covariates.

# Predictors of Diagnosis with NADC-comparison I



Characteristic	Model 1		Model 2		Model 3	
	OR	95% C.I.	OR	95% C.I.	OR	95% C.I.
<b>Gender</b>						
Females	Ref.		Ref.		Ref.	
Males	1.39	(1.09,1.78)	1.39	(1.08,1.77)	1.29	(1.01,1.65)
<b>Race</b>						
Black	Ref.		Ref.		Ref.	
Hispanic	0.56	(0.27,1.16)	0.56	(0.27,1.15)	0.52	(0.25,1.08)
Other/unknown	1.2	(0.55,2.62)	1.19	(0.55,2.59)	1.1	(0.5,2.42)
White	1.31	(1.03,1.66)	1.32	(1.04,1.67)	1.3	(1.02,1.65)
<b>Transmission risk</b>						
Heterosexual	Ref.		Ref.		Ref.	
IDU	0.6	(0.38,0.94)	0.6	(0.38,0.94)	0.61	(0.39,0.96)
MSM	0.74	(0.55,0.99)	0.74	(0.56,0.99)	0.78	(0.59,1.05)
MSM and IDU	0.15	(0.04,0.61)	0.15	(0.04,0.62)	0.16	(0.04,0.66)
Other/unknown	0.77	(0.6,0.99)	0.77	(0.6,0.99)	0.76	(0.59,0.98)
<b>Age at HIV diagnosis</b>						
18-29	Ref.		Ref.		Ref.	
30-39	2.07	(1.41,3.05)	2.06	(1.4,3.03)	2.03	(1.38,2.98)
40-49	4.7	(3.32,6.67)	4.66	(3.28,6.61)	4.58	(3.24,6.49)
50-59	5.62	(3.87,8.16)	5.56	(3.82,8.1)	5.38	(3.7,7.81)
60+	8.97	(5.77,13.95)	8.86	(5.69,13.8)	8.49	(5.46,13.19)
<b>Residence</b>						
Urban	Ref.		Ref.		Ref.	
Rural	1.22	(0.96,1.54)	1.21	(0.96,1.53)	1.17	(0.92,1.48)
<b>Tobacco use</b>						
No	Ref.		Ref.		Ref.	
Yes	2.83	(2.26,3.54)	2.82	(2.25,3.52)	2.76	(2.2,3.46)
<b>Drug use</b>						
No	Ref.		Ref.		Ref.	
Yes	1.24	(0.96,1.6)	1.24	(0.96,1.6)	1.18	(0.91,1.53)
<b>Alcohol use</b>						
No	Ref.		Ref.		Ref.	
Yes	0.89	(0.68,1.16)	0.89	(0.68,1.16)	0.87	(0.66,1.13)
<b>Timely linkage</b>						
No	Ref.		Ref.		Ref.	
Yes	1.09	(0.89,1.33)	1.09	(0.89,1.32)	1.08	(0.89,1.32)



# Predictors of Diagnosis with NADC-comparison I

Characteristic	Model 1		Model 2		Model 3	
	OR	95% C.I.	OR	95% C.I.	OR	95% C.I.
<b>Hepatitis B or C</b>						
No	Ref.		Ref.		Ref.	
Yes	1.25	(0.93,1.67)	1.26	(0.94,1.68)	1.22	(0.91,1.64)
<b>Hypothyroidism</b>						
No	Ref.		Ref.		Ref.	
Yes	2.73	(2.2,3.37)	2.68	(2.17,3.3)	2.44	(1.98,3.02)
<b>Hypertension</b>						
No	Ref.		Ref.		Ref.	
Yes	1.5	(1.19,1.88)	1.5	(1.19,1.88)	1.55	(1.24,1.95)
<b>Diabetes</b>						
No	Ref.		Ref.		Ref.	
Yes	1.29	(1,1.67)	1.29	(1,1.67)	1.26	(0.97,1.64)
<b>Obesity</b>						
No	Ref.		Ref.		Ref.	
Yes	0.79	(0.54,1.15)	0.79	(0.54,1.16)	0.84	(0.57,1.24)
<b>Dyslipidemia</b>						
No	Ref.		Ref.		Ref.	
Yes	1.21	(0.93,1.58)	1.22	(0.93,1.58)	1.28	(0.98,1.68)
<b>Initial viral load</b>						
Undetectable	Ref.		Ref.		Ref.	
Detectable	0.87	(0.55,1.36)	0.83	(0.53,1.29)	0.72	(0.46,1.13)
<b>Nadir CD4 count</b>						
>350	Ref.					
<200	0.96	(0.72,1.29)	.		.	
200-350	0.93	(0.72,1.2)	.		.	
<b>Initial CD4 count</b>						
>350			Ref.			
<200	.		1.16	(0.89,1.51)	.	
200-350	.		1.05	(0.84,1.31)	.	
<b>Recent CD4 count</b>						
>350					Ref.	
<200	.		.		1.81	(1.4,2.34)
200-350	.		.		1.86	(1.47,2.35)



## Comparison II

- Model 1: Logistic model including initial CD4 count and initial VL level adjusting for other covariates.
- Model 2: Logistic model including initial CD4 count and recent VL level adjusting for other covariates.

# Predictors of Diagnosis with NADC-comparison II



Characteristic	Model 1		Model 2	
	OR	95% C.I.	OR	95% C.I.
<b>Gender</b>				
Females	Ref.		Ref.	
Males	1.39	(1.08,1.77)	1.36	(1.06,1.74)
<b>Race</b>				
Black	Ref.		Ref.	
Hispanic	0.56	(0.27,1.15)	0.56	(0.27,1.16)
Other/unknown	1.19	(0.55,2.59)	1.19	(0.54,2.59)
White	1.32	(1.04,1.67)	1.33	(1.05,1.69)
<b>Transmission risk</b>				
Heterosexual	Ref.		Ref.	
IDU	0.60	(0.38,0.94)	0.59	(0.38,0.93)
MSM	0.74	(0.56,0.99)	0.75	(0.56,1.01)
MSM and IDU	0.15	(0.04,0.62)	0.15	(0.04,0.61)
Other/unknown	0.77	(0.6,0.99)	0.77	(0.59,0.99)
<b>Age at HIV diagnosis</b>				
18-29	Ref.		Ref.	
30-39	2.06	(1.4,3.03)	2.10	(1.42,3.09)
40-49	4.66	(3.28,6.61)	4.83	(3.4,6.86)
50-59	5.56	(3.82,8.1)	5.87	(4.03,8.56)
60+	8.86	(5.69,13.8)	9.27	(5.95,14.44)
<b>Residence</b>				
Urban	Ref.		Ref.	
Rural	1.21	(0.96,1.53)	1.21	(0.96,1.52)
<b>Tobacco use</b>				
No	Ref.		Ref.	
Yes	2.82	(2.25,3.52)	2.79	(2.23,3.49)
<b>Drug use</b>				
No	Ref.		Ref.	
Yes	1.24	(0.96,1.6)	1.19	(0.92,1.53)
<b>Alcohol use</b>				
No	Ref.		Ref.	
Yes	0.89	(0.68,1.16)	0.85	(0.65,1.12)
<b>Timely linkage</b>				
No	Ref.		Ref.	
Yes	1.09	(0.89,1.32)	1.08	(0.89,1.32)





# Predictors of Diagnosis with NADC-comparison II

Characteristic	Model 1		Model 2	
	OR	95% C.I.	OR	95% C.I.
<b>Hepatitis B or C</b>				
No	Ref.		Ref.	
Yes	1.26	(0.94,1.68)	1.24	(0.93,1.66)
<b>Hypothyroidism</b>				
No	Ref.		Ref.	
Yes	2.68	(2.17,3.3)	2.6	(2.11,3.21)
<b>Hypertension</b>				
No	Ref.		Ref.	
Yes	1.5	(1.19,1.88)	1.5	(1.2,1.89)
<b>Diabetes</b>				
No	Ref.		Ref.	
Yes	1.29	(1,1.67)	1.31	(1.02,1.7)
<b>Obesity</b>				
No	Ref.		Ref.	
Yes	0.79	(0.54,1.16)	0.81	(0.55,1.2)
<b>Dyslipidemia</b>				
No	Ref.		Ref.	
Yes	1.22	(0.93,1.58)	1.27	(0.97,1.65)
<b>Initial viral load</b>				
Undetectable	Ref.			
Detectable	0.83	(0.53,1.29)	.	
<b>Recent viral load</b>				
Undetectable			Ref.	
Detectable	.		1.55	(1.27,1.88)
<b>Initial CD4 count</b>				
>350	Ref.		Ref.	
<200	1.16	(0.89,1.51)	1.12	(0.86,1.46)
200-350	1.05	(0.84,1.31)	1.01	(0.81,1.26)



## Outline:

- Background
- Data source
- Statistical analysis
- Results
- **Conclusion**
- Future Study



## Conclusions

- Several factors affect NADC in PLWH in SC.
- Aged PLWH, with low recent CD4 count or detectable recent viral load level should be the focus population alongside with PLWH who has comorbidities like Hypothyroidism, diabetes and hypertension.



## Outline:

- Background
- Data source
- Statistical analysis
- Results
- Conclusion
- **Future Study**



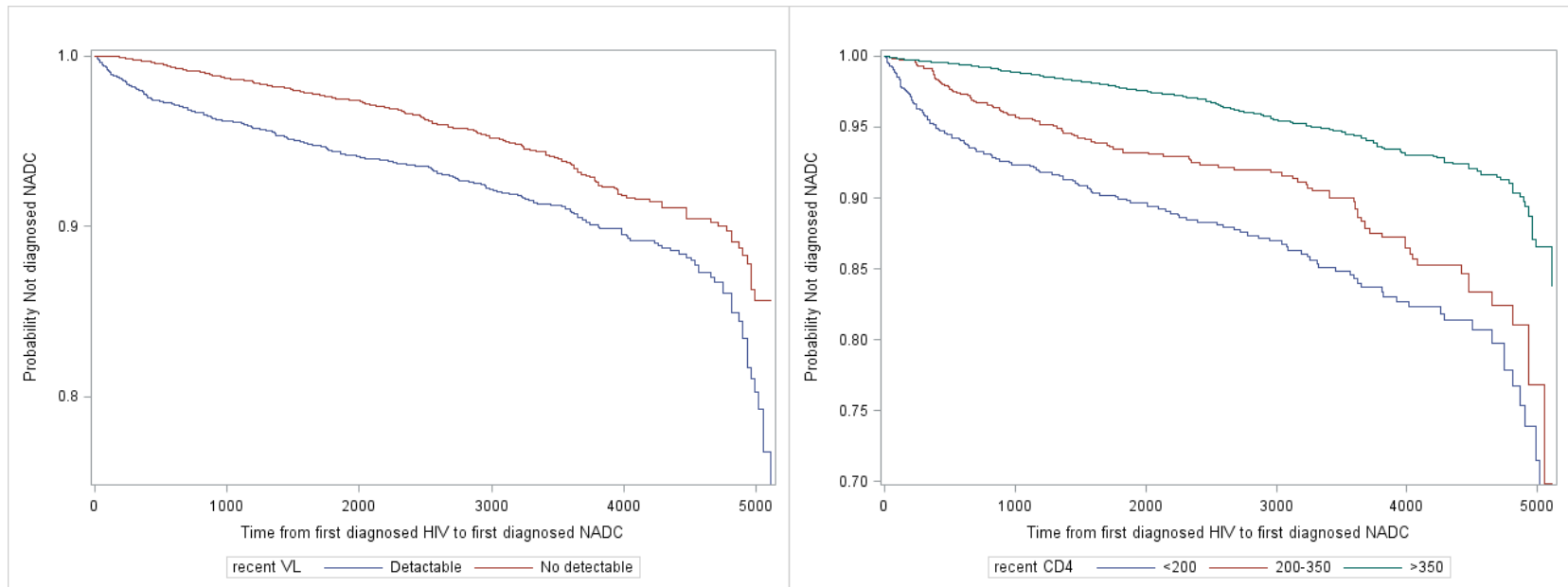
## Future study I

- The longitudinal observation is reflected via the recent/ initial/ lowest CD4 or initial/ recent viral load. The joint modelling approach will be applied later to capture the observation of CD4/viral load over time.



## Future study II

- Investigate the prediction of time to onset of NADC using the longitudinal patterns of CD4 and viral load.



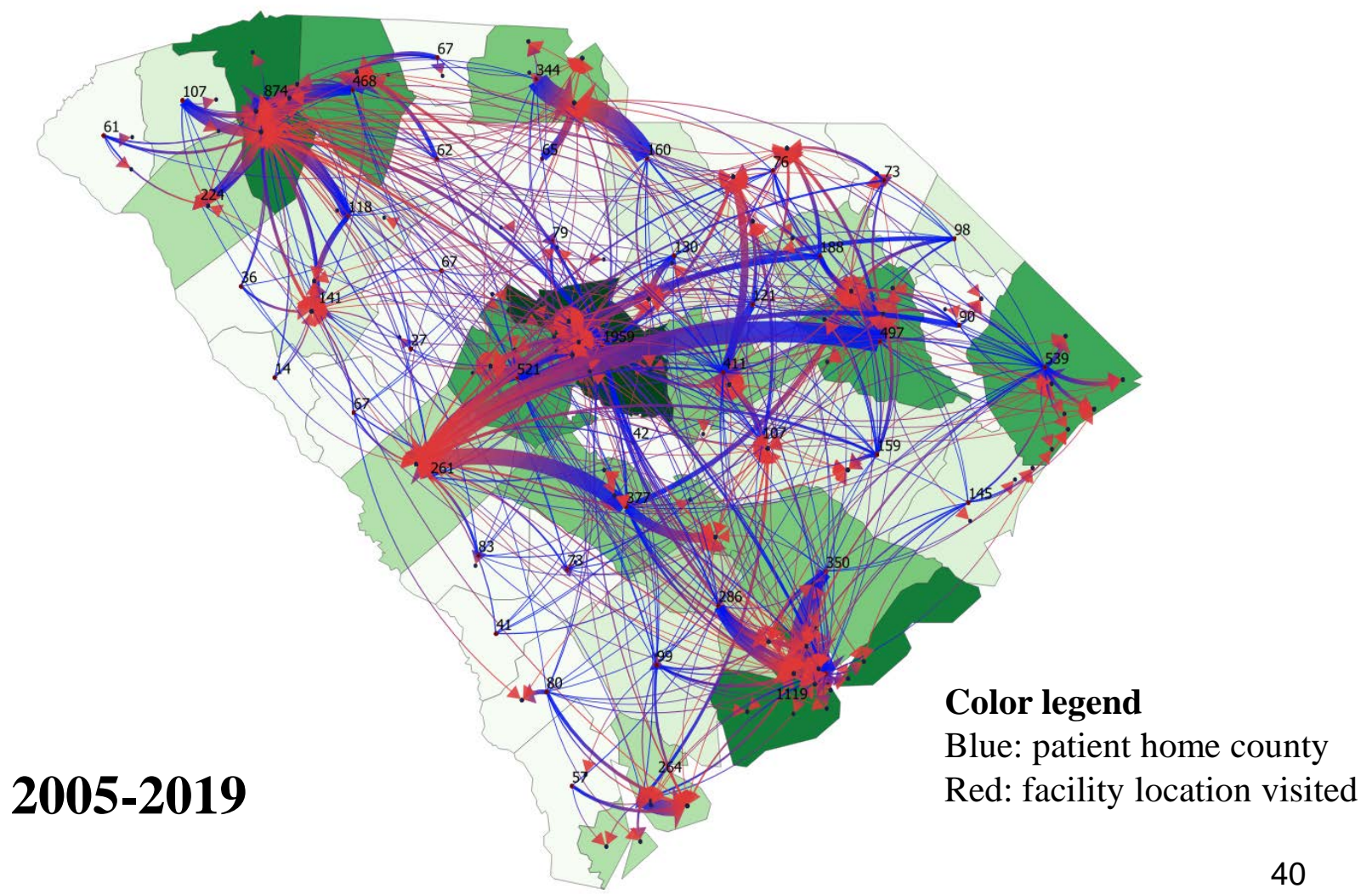


## Future Study III

- Geographical patterns of HIV care seeking behavior, examine geographical difference
  - Linkage to care
  - Retention in care pattern
  - Adherence to care



# # of HIV and Hospital visit

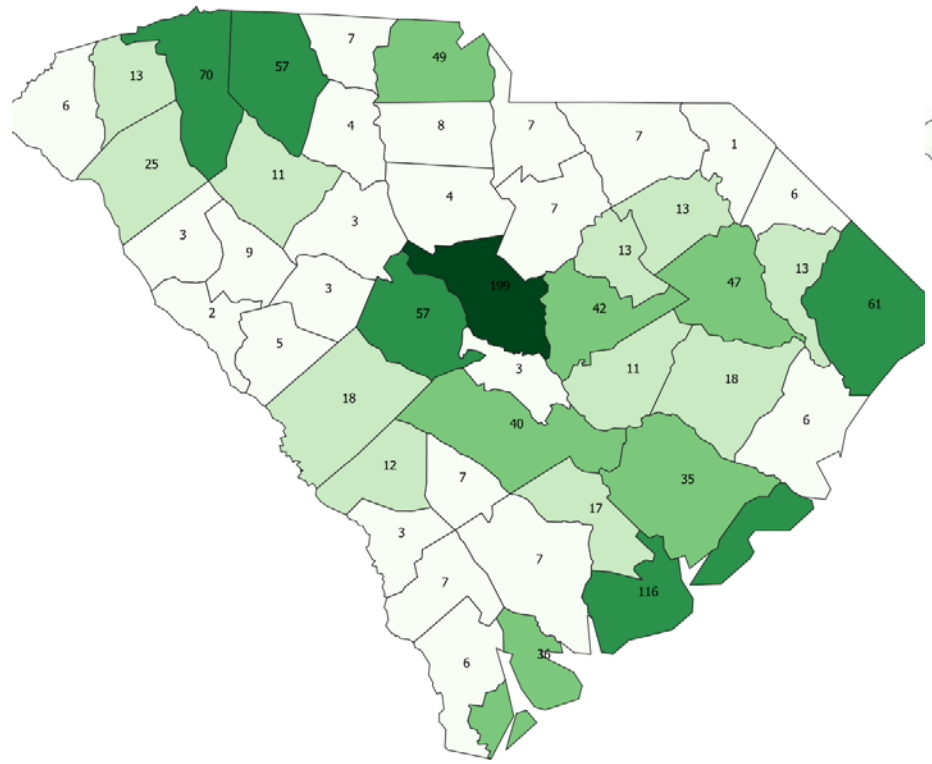




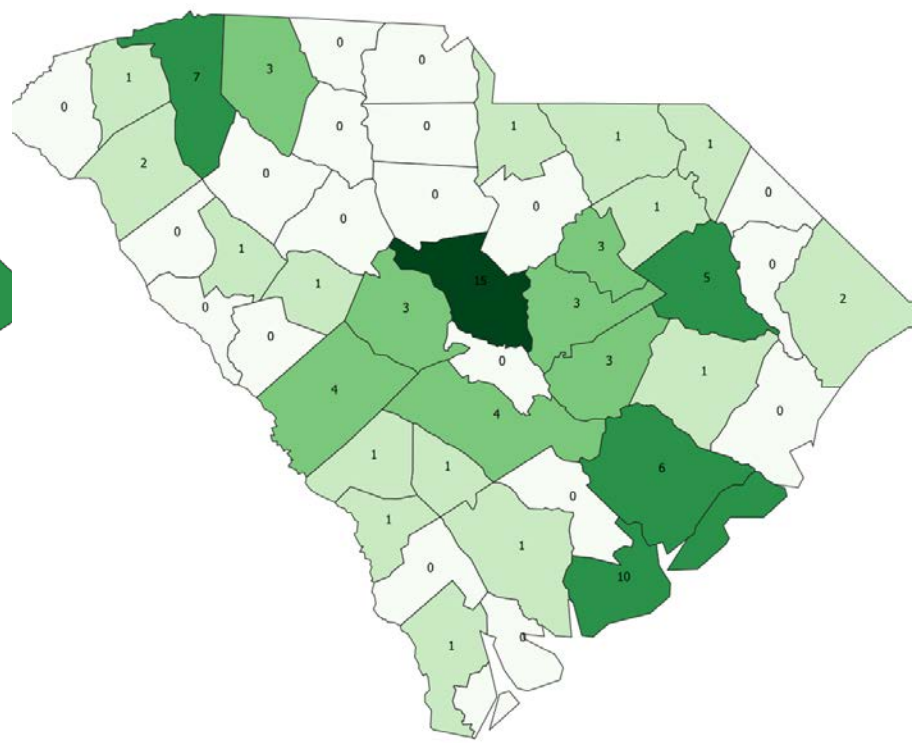


# Spatial Temporal: # of HIV diagnosis

2008



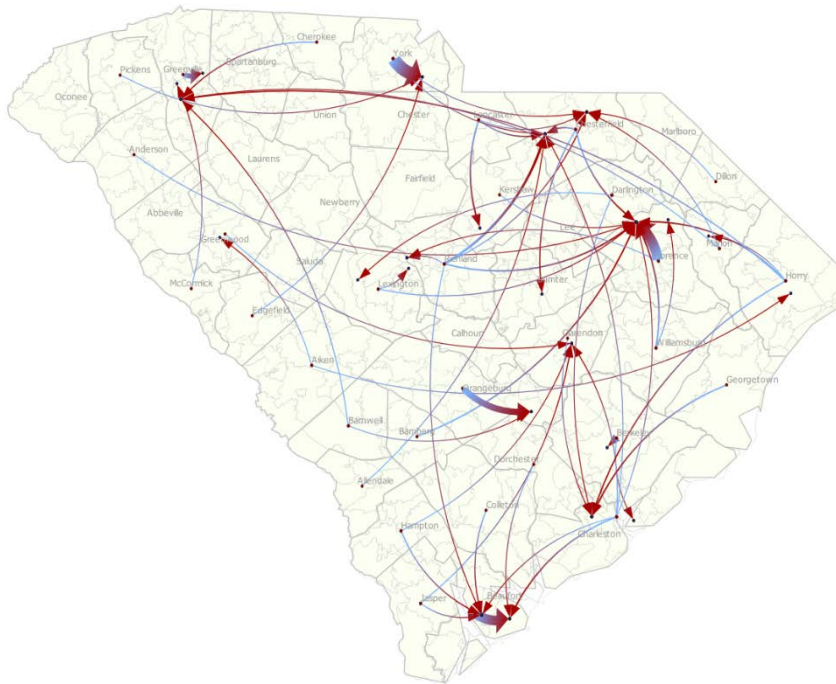
2018



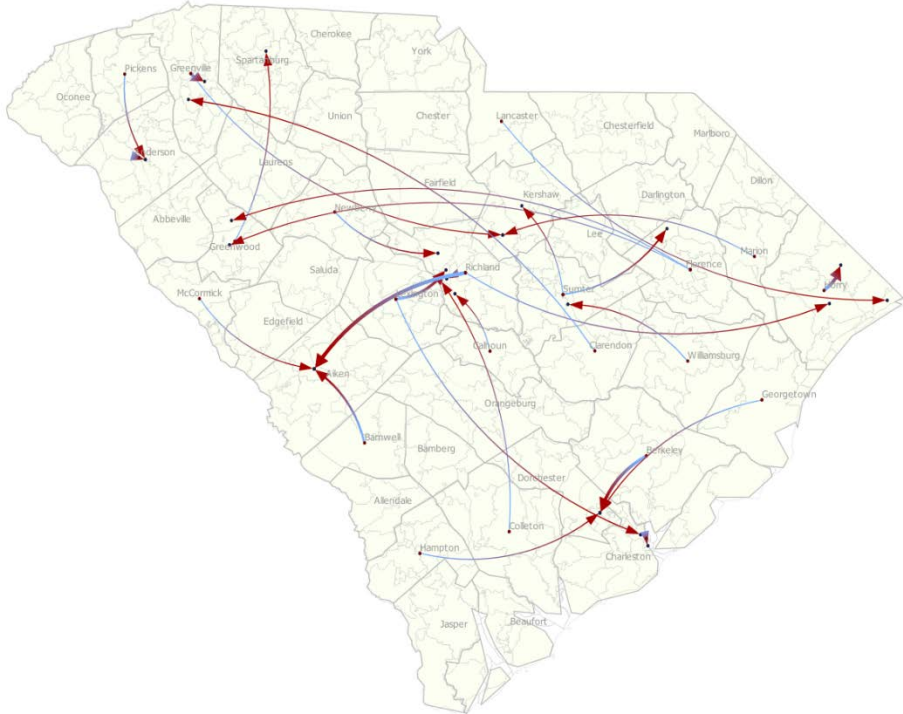


# Hospital visting flow

2008



2018



## Color legend

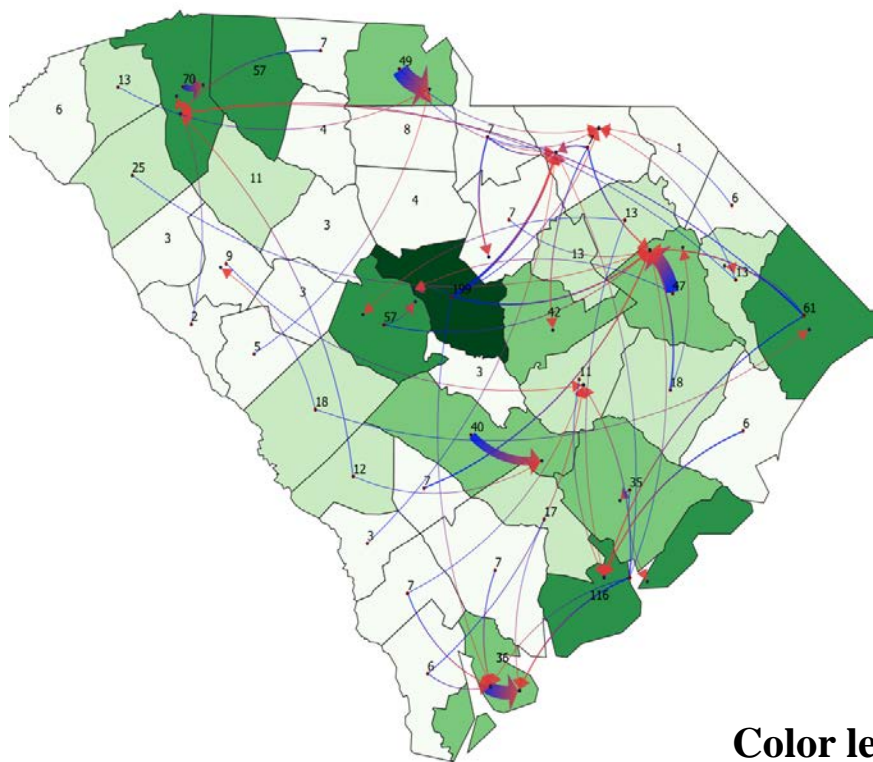
Blue: patient home county

Red: facility location visited

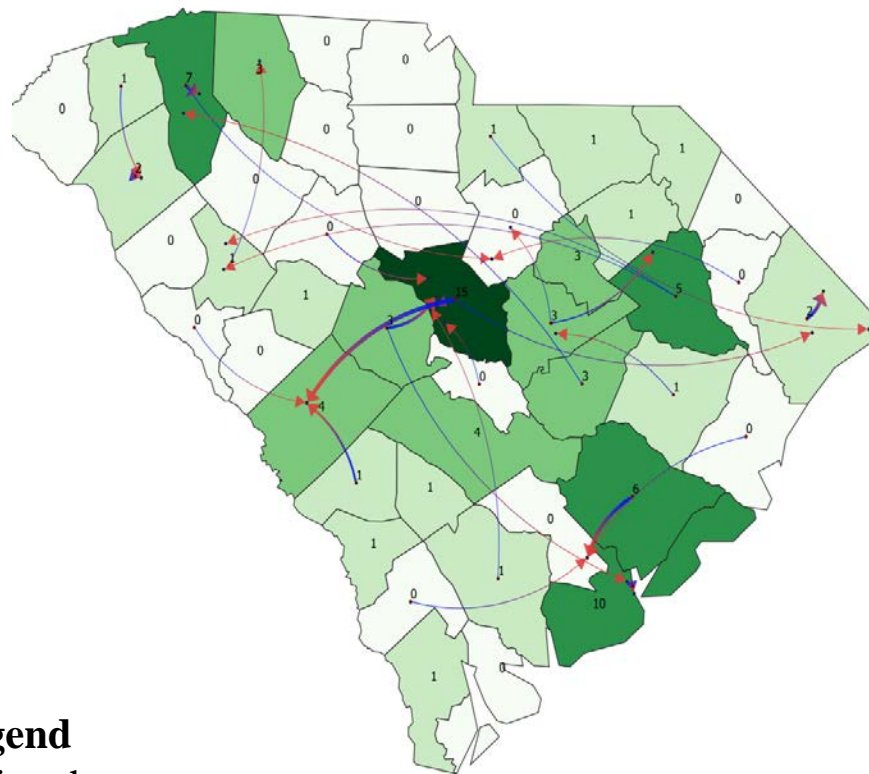


# # of HIV and Hospital visiting flow

2008



2018



## Color legend

Blue: patient home county

Red: facility location visited



# Acknowledgement

- **1R01AI127203-01A1** Big Data Analytics of HIV Treatment Gaps in South Carolina: Identification and Prediction, NIH/NIAID
- SC DHEC, RFA, HSSC, DMH, ...
- Research Team: Xiaoming Li, Banky Olatosi, Jiajia Zhang, Sharon Weissman, Chigozie Nkwonta,



# References

- Deeken, John F., et al. "The rising challenge of non–AIDS-defining cancers in HIV-infected patients." *Clinical Infectious Diseases* 55.9 (2012): 1228-1235.
- Shiels, Meredith S., and Eric A. Engels. "Evolving epidemiology of HIV-associated malignancies." *Current Opinion in HIV and AIDS* 12.1 (2017): 6-11.
- [https://www.scdhec.gov/sites/default/files/media/document/SC-20-year-cancer-report\\_2018.pdf](https://www.scdhec.gov/sites/default/files/media/document/SC-20-year-cancer-report_2018.pdf)
- Engels, Eric A., et al. "Cancer risk in people infected with human immunodeficiency virus in the United States." *International journal of cancer* 123.1 (2008): 187-194.
- South Carolina Department of Health and Environmental Control (SC DHEC). *An Epidemiologic Profile of HIV and AIDS in South Carolina*. 2018.



**Thank You**