Invisible Information, Unseen Connections:
An Exploration of the Hidden Relationships that can Shape Data

by

Emily Catherine Andrus

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Epidemiological Science)
in the University of Michigan
2023

Doctoral Committee:

Associate Professor Marisa C. Eisenberg, Chair
Assistant Research Scientist Andrew F. Brouwer
Associate Professor David W. Hutton
Professor Rafael Meza
Associate Professor Jonathon L. Zelner

Emily Catherine Andrus

ecandrus@umich.edu

ORCID iD: 0000-0002-5272-7258

Dedication

To Michael and Ginger

Acknowledgements

Even though my name is on this dissertation, its contents could not have been written without the help of a community of people. I'd like to acknowledge the members of my committee, Dr. Marisa C Eisenberg, Dr. Andrew F Brouwer, Dr. David Hutton, Dr. Rafael Meza, and Dr. Jonathon L Zelner for their insight, critiques, and unending support. Dare I say I had *fun* working on my dissertation.

I would like to thank Dr. Sanyu A Mojola for her counsel on my first aim. This project would not have been possible without her expertise, and I am grateful for the time we spent collaborating. I would like to acknowledge Beth Moran, for her work compiling the Demographic Health Survey Data used in Aim 1; and I want to thank Dr. Lisa Lau, and fellow Nevadan, Dr. Matthew Boulton for their insight on economic measures.

I would like to thank Dr. Joshua G Petrie for inviting me to join the project that became my third and fourth aims, and for his generous funding support; and Dr. Robert Woods for his counsel on Aims 3 and 4. Understanding the layout and day to day operations of University Hospital would have been impossible otherwise. I would like to acknowledge Jess Millar, for cleaning and compiling the hospital transfer data used in my third and fourth aims.

I want to thank Rob Trangucci for his guidance on statistics; and Betsy Salzman for her laboratory acumen, patience, unconditional support and wisdom. I like to thank Dr. Michael AL Hayashi for his mentorship, guidance, and friendship, and Drs. Katherine M Begley and John T Kubale, and Viktoryia A Kalesnikava for their support and friendship.

Lastly, I want to thank Michael Menard for seeing my potential and being my biggest cheerleader. *This* was my goal.

Preface

Chapter 2 has been published with the following authors and citation:

Andrus E, Mojola SA, Moran E, Eisenberg M, Zelner J. Has the relationship between wealth and HIV risk in Sub-Saharan Africa changed over time? A temporal, gendered and hierarchical analysis. *SSM-population health*. 2021 Sep 1;15:100833.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Datasets often reflect complex and nuanced relationships which can be difficult to detect or fully represent with traditional epidemiological methods. This may be problematic as it can hinder further analyses, or give the investigator an incomplete picture of the outcome being studied. In this dissertation I explored three analytic contexts in which important relationships can go undetected and examined several methods that can be used to ascertain hidden or latent relationships in the data, drawing from meta-regression, latent class analysis, network analysis, and transmission modeling.

In Aim 1, we used meta-regression to ascertain how the association between individual wealth, country level wealth, and Human Immunodeficiency Virus (HIV) burden has changed over time across a set of Sub-Saharan African (SSA) Countries. It has been assumed that like, The West, HIV is also a disease of poverty in SSA. However newer research suggests that this assumption may not be true. Here, we show that HIV may be positively associated with wealth in urban but not rural contexts, and that this association has waned over time.

Aim 2 identifies patterns of sexual behavior and substance use across the life course, and examines the association between these patterns and sexually transmitted infection risk. Risk factors for sexually transmitted infections have proven challenging to study due to their tendency to be highly correlated or even collinear with one another. This collinearity is problematic because it inhibits the ability of statistical software to detect the effect of covariates in a regression model, rendering the coefficients of the variables uninformative. Consequently,

alternative approaches are needed in order to identify behaviors that put individuals at risk for infection. This aim uses Latent Class Analysis which, unlike regression, uses collinearity to its advantage to identify response patterns. Our results reveal the existence of 5 archetypes that serve as the basis for the profiles present across our four age strata. However, the exact composition of each strata's profiles varies in the magnitude that particular behaviors are endorsed, which we attribute to a combination of age, period, and cohort effects.

Aim 3 constitutes the first part of a two-part analysis that uses network methodology to characterize and quantify patient movement and disease transmission. In this aim, a descriptive analysis of network structure was undertaken to describe the underlying interrelationship between hospital units and patient movement, using patient transfer data from the University Hospital at Michigan Medicine. We then characterized the resulting network to understand key structural features, including node centrality, graph centralization, degree distributions, and community structure. As a network, University Hospital is decentralized but highly transitivity .

In Aim 4 we used an SEIR compartmental model to simulate COVID-19 in a hospital setting, to examine the relationship between the hospital network structure and disease transmission dynamics. The purpose of this analysis was to illustrate how the network relationship between locations can be an underlying structure that informs transmission dynamics within the hospital.

In summary, the chapters of this dissertation illustrate contexts in which latent variable associations exist in data and provide tools researchers can use to extract them. It is our hope that this work provokes thought and sparks new lines of inquiry.

1 Introduction

Sometimes an otherwise quotidian dataset may contain complex and nuanced relationships that are undetectable with traditional epidemiological methods. Depending on the questions being asked, these hidden relationships may be problematic as they can hinder further analyses or give the investigator an incomplete picture of the outcome being studied. In this dissertation, I provide three analytic contexts in which important relationships can go undetected, and I demonstrate methods that can be used to ascertain this "hidden" information. In the first aim, I use meta-regression to reexamine commonly held assertions about the relationship between HIV and wealth in sub-Saharan Africa. In the second aim, I use latent class analysis to circumvent issues with collinearity and identify inter-generational differences in sexually transmitted infection risk factors. In the third and fourth aims, I use network analysis to capture the meta-effects a hospital's layout and movement of patients throughout the structure drive infectious disease transmission.

1.1 Aim 1: Has the Relationship Between Wealth and HIV risk in Sub-Saharan Africa Changed over time? A Temporal, Gendered, and Hierarchical Analysis.

Historically it has been canon that Human Immunodeficiency Virus (HIV) is associated with poverty and marginalization.[1–4] It has been assumed that since HIV is a disease of poverty in wealthier western countries, then it must be a disease of poverty in Low- and Middle-Income Countries, with poorer countries often assumed to be at higher risk of HIV.[2,3] However, this rationale implicitly assumes that markers of country level wealth are sufficient to reflect the

burden of disease amongst individuals, and is thus ecologically fallacious. Indeed, recent research suggests that this long held assumption may not be true for Sub-Saharan African Countries.[5–9] Instead, these new data propose that HIV may be positively associated with individual and country-level wealth. Aim 1 re-investigates this relationship, to uncover how the masked association between wealth, and HIV burden has changed over time in Sub-Saharan Africa. [5–9]

Specifically, this aim re-analyzes the relationship between country level wealth, individual level wealth and HIV positivity in sub-Saharan Africa. Drawing upon insights gleaned from recent studies we employ meta-regression in a novel context to separate the effects of individual level wealth from country level wealth on HIV prevalence. Doing so counters an occurrence of the ecological fallacy that has historically impacted prior analyses . [5–9]

1.1.1 Meta-regression

Meta-regression is an analytic technique traditionally used to combine findings from multiple studies to obtain an overall or "meta" effect measurement.[10–12] Meta-regression is multi-level modeling technique, and as such can account for fixed and mixed effects. [10–12] The first level represents the individual findings from each study individually, and the second level uses the effect measures from the first level to find an overall effect measure for all of the studies survey. [10–12] However, meta-regression differs from traditional multi-level models in that meta-regression includes an additional error term that can account for between cluster variance, also known as cluster heterogeneity. [10–12] Simply put, traditional multi-level models assume that the effect across all clusters is the same, whereas meta-regression doesn't hold this assumption; assuming instead that a distribution of effect sizes exists across clusters. [10–12]

In the context of this aim, the first step modeled the univariate relationship between HIV, and gender, urbanicity and wealth for each of the 29 countries, individually. In the second step, aggregated effect measures and standard errors from step one were combined into a single dataset then used to estimate the relationship between country level economic indicators and HIV burden, as well as the country level HIV burden over time.

Since this methodology models individual level and country level data separately, the individual effect sizes and variances of each country can be accounted for in the second level model; as opposed to a multi-level model that would use the mean of all 29 countries in the second level. As a result, we avoid the additional pitfalls encountered by prior analyses that assume the relationship between HIV and wealth is the same across Sub-Saharan Africa, and we can increase precision of our estimates by accounting for the individual variances of each country. In short, this aim uses meta-regression to tease apart the individual relationships between HIV, individual level wealth and population wealth that heretofore have remained intertwined.

1.1.2 Economic measures:

Country-level economic indicators were used to reflect the effects of country level wealth in our analyses.. We elected to use four economic measures.

1.1.2.1 Gini Index:

Named after Italian economist Carrado Gini, the Gini Index measures the extent to which the distribution of income across individuals in an economy deviates from an equal distribution.[13,14] The basis of the Gini Index is the Lorenz curve of an economic model, that first splits an economy's population into quantiles then plots the cumulative income of an economy

against the cumulative population quantiles.[13,15] As a control/comparator this curve is plotted next to a 45-degree line that represents perfect income parity.[13,15,16] The Gini Index is defined by the area between the Lorenz curve and the 45-degree line.[13,16] Thus, a Gini Index of 0 would indicate an economy with no wealth inequality, while a Gini index of 100 indicates maximum wealth inequality. [13]

1.1.2.2 Human Development Index (HDI):

The Human Development Index is a composite measure of three indices.[17] The first dimension is "health" and is measured by a population's life expectancy at birth.[17] The second dimension is "education" and is measured in two ways. First by the average amount of schooling in years, adults 25 and over have received, and second by the expected years of schooling for school age children.[17] The final dimension is "standard of living" which is measured by the gross national income purchasing power parity. [17]To calculate the HDI, these indices are normalized, then the geometric mean of the three is calculated.[17]

1.1.2.3 Gross Domestic Product (GDP):

A country's gross domestic product is the total revenue added to a country's economy from the sale of goods manufactured domestically.[13] Revenue includes taxes but does not include product subsidies.[13] This analysis considered both a countries total GDP, as well as the proportion of the GDP attributable to health care spending.

1.2 Aim 2: Latent Class Analysis of Generational Trends In STI Risk Factors

Historically, sexuality and STI research has focused on younger populations, and as a consequence there is a paucity of research on sexual health in older adults or across the life course.[18] This is problematic because STIs continue to be an issue across all age groups.[18–20]

It is known that substance use and sexual activity are risk factors for STI positivity.[21–24] However, these behaviors are heavily influenced by externalities such as socio-cultural norms, responsibilities, and age effects, leading one to question if it is safe to assume that the effects of sexual activity and substance use are the same across age groups? To complicate things further there is no one standard way to operationalize these constructs, and regardless of operationalization these constructs can be highly collinear.

Such collinearity is problematic because it inhibits the ability of statistical software to detect the effect of covariates in a regression model, rendering the coefficients of the variables uninformative. Consequently, alternative approaches are needed in order to identify behaviors that put individuals at risk for infection. This aim uses one such approach called Latent Class Analysis (LCA). Unlike regression, LCA uses collinearity to its advantage to identify response patterns that might otherwise remain undetected by regression.

1.2.1 What is latent class analysis?

At its core, latent class analysis is a form of unsupervised machine learning that can detect hidden or "latent" heterogeneous subgroups or "classes" in seemingly homogeneous data.[25–27] In order to detect a specific number of latent classes, LCA uses measured categorical variables, or "manifest variables", as proxies for the different characteristics of the latent classes [FIGURE 1.1]. Whereas regression systematically analyzes the outcome measure with each coefficient individually, LCA takes a more holistic approach. First the data are analyzed for recurrent configurations of variables, irrespective of an individual's particular response pattern.[28] These recurrent configurations represent the characteristics of each latent class. [28] Once the latent classes have been identified, individual observations are binned into the class they have the

highest probability of belonging to. [28]What results are classes defined by the probability that a member of the class will have a certain response to one of the manifest variables. [28]

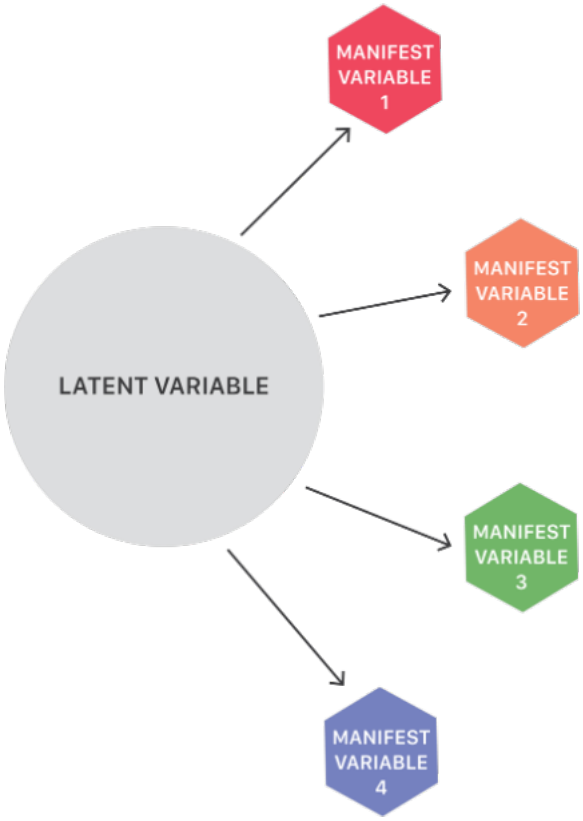


Figure 1-1 A Latent Variable is Expressed as Dimensions of Manifest Variable

1.3 Aim 3: Characterization of Network Structure and Patient Volume at Michigan Medicine - University Hospital

Understanding how infections spread through physical space is essential for day-to-day infection control and can aid in planning for future epidemics. However, the self-contained nature of units or wards in a hospital can pose challenges for traditional mathematical models. This is because typical compartmental models assume that people in a population mix uniformly and randomly. While this assumption may hold for individual wards, it is not necessarily true for movement between units.

One solution is to render the hospital as a network of units connected by patient traffic. The benefits of doing this are twofold. Not only can a network accommodate within and between unit movement, but networks come with a set of statistics which can help us identify potential hotspots, describe the interconnectivity of units, and identify clusters of units that share a common patient base.

Aim 3 constitutes the first part of a two-part analysis that uses network methodology to characterize and quantify patient movement and disease transmission within University Hospital at Michigan Medicine. In this aim, a descriptive analysis of network structure is undertaken to identify and describe the underlying interrelationship between hospital units and patient movement. Since network terminology may be unfamiliar to some, a short glossary of terms is included below, as well as a description of the centrality measures used in this aim.

Network Terminology

The terms below are presented in alphabetical order.

1.3.1 Attribute:

A characteristic or feature of a node or edge.[29–31]

1.3.2 Degree:

The number of edges connecting a node to other nodes.[29,30] For a directed graph (see definition below), *in-degree* of a node is defined as the number of incoming edges from other nodes, and *out-degree* of a node is defined as the number of outgoing edges directed toward other nodes. Similarly, for a weighted graph, the *weighted degree* (or weighted in/out degree) of a node is the sum of the edge weights for each edge connected to the given node (or in/outgoing weights respectively).

1.3.3 Directed Graph:

A network in which edges indicate the direction of movement.[29,30] The hospital network developed for this aim has directed edges. We differentiate between transfers from Node A to Node B and transfers from Node B to Node A. In an undirected graph these transfers would be combined and represented as one link instead of two.

1.3.4 Edge:

The link between two nodes.[29] In this analysis the edges represent number of people being moved from one unit to another.

1.3.5 Network:

A group of points linked together by lines.[29] Sometimes also termed a *graph*.

1.3.6 Node/Vertex:

Point or object in a network (i.e. one of the objects connected by edges).[29] Nodes can represent myriad things: computers connected through internet, children in a peer group, or

checkpoints on a marathon course. In this analysis the hospital units comprise our network's nodes.

1.3.7 Weighted Graph:

A network in which edges indicate the strength of connection.[29–31] The network for this aim is a weighted network. Edges have the added attribute size, which is visually represented by the thickness of the links and is scaled to match the number of people moving from one unit to another. In unweighted graphs edges do not have such a component. Instead, they are present if a connection exists between two nodes, or are absent if a connection does not exist.

1.3.8 Centrality:

Centrality is a characteristic of nodes that quantifies how central, connected or important that node is relative to other nodes. There are numerous methods for calculating centrality, all of which define importance in a distinct way. In this aim we used three of the more general and commonly used centrality measures, described below.[29–31]

1.3.9 Betweenness Centrality:

A form of centrality that assess a node's importance by the frequency with which it lies on the shortest path connecting two other nodes. [29–31]The number of times a node is found to lie on the shortest path between two others represents the non-normalized betweenness centrality of the node. [29–31]This score can be normalized by dividing it by the theoretically highest score a graph with the same number of nodes could have.[29–31] Normalized values can take a range of [0,1], with 0 indicating low betweenness centrality, and 1 indicating high betweenness centrality.[30]

1.3.10 Closeness Centrality:

A form of centrality that uses a node's proximity to other nodes as an indicator of its importance. Closeness is calculated by first identifying the shortest paths between a node and all other nodes.[29] The number of edges on each path are summed, if the graph is unweighed each edge has a value of 1, otherwise their value is equal to their weight. Then, the mean length is taken by dividing this sum by the number of nodes.[29] Finally, to aid interpretability the inverse of this mean is taken so that low values represent low centrality and high values represent high centrality.[29–31] Closeness Centrality values range from [0,1].

1.3.11 Degree Centrality:

A form of centrality that uses the node's degree as a measure of its importance. Nodes with a low degree are considered less central or less important compared to those with a high degree. Degree centrality values range from [0,∞) (or for a network with N nodes, [0,N)).[29–31]

1.3.12 *Modularity:*

A measure of the strength of a given proposed community structure for a network. For a proposed set of communities on a network, modularity measures the extent to which nodes within a community tend to be connected to other nodes within their community vs. to nodes outside their community, compared to a network in which the edges are assigned randomly. Higher modularity tends to indicate a better proposed community structure, and although finding the true maximum modularity is an NP-hard problem, community detection algorithms exist which have been shown to perform well in practice on most networks.[29]

1.3.13 Neighbor/Neighborhood:

The *neighborhood* of a node is defined as the set of all nodes which are directly connected to the given node. Nodes within the neighborhood of a node *n* are known as *neighbors* of *n*.

1.3.14 Transitivity:

A measure of the interconnectedness of nodes, also known as the *global clustering coefficient*.[29] Transitivity is defined as the number of triangles in the network (i.e. situations in which two neighbors of a node are themselves neighbors, a measure of clustering), divided by the total number of possible triangles in the network. Transitivity ranges from 0 to 1.

1.3.15 Centralization:

This is the extent to which the nodes in a graph are centered around a common node or hub, and can be conceptually thought of as graph level centrality. Centralization is obtained by first summing the differences in centrality measures between a node and all of the other nodes in a graph, then dividing this sum by highest possible centrality value for a graph with the same number of nodes. Thus, centralization is the ratio between the summed differences in centrality between nodes, and the highest theoretically possible centrality value for a node given a graph of the same size. Values of centralization will range from [0,1], with 0 indicating no centralization and 1 high centralization.[32,33]

1.4 Aim 4: Simulation of Respiratory Virus Outbreak on a Network

In the second part of this two-part analysis, disease outbreaks were simulated on the network built in Aim 3. Since the motivation for this study was to explore COVID-19 transmission, an SEIR compartmental model was used to capture both asymptomatic and symptomatic carriers. The purpose of this analysis is to show how the existing meta spatial

relationship between locations can provide additional structure that shapes the transmission of an infectious agent.

1.4.1 SEIR Compartmental Model

A compartmental model is a type of model that uses subgroups or "compartments" to simulate population level transitions between tangible or conceptual states. [34] They have wide application, and in epidemiology are used to study the dynamics of an infection as it passes through a population. Here, the compartment represents the states one experiences when catching an infection, falling ill, and then recovering. The "S" compartment represents the state of being "Susceptible" to an illness.[34] Most of the population in a simulation will begin here. The "E" compartment represents being "Exposed" to a pathogen (traditionally termed "exposed", but perhaps more accurately capturing latent infection) but not being able to transmit it to others.[34] This compartment may not be needed in other types of infections. The "I" compartment represents the state of being "Infectious" and able to transmit the pathogen to other people.[34] Finally, the "R" compartment represents "Recovered" (or in some cases, "Removed" if this class includes individuals who are deceased due to the disease).[34] This is the category individuals pass to once they are no longer infectious and no longer showing signs or symptoms. Movement between these compartments is governed by rates that are typically either estimated from data or pulled from the literature.

When put together, the compartments, rates and the process they simulate can be expressed as a series of ordinary differential equations.[34] Where S,E,I, and R represent the number of people in that compartment at any given time point. β represents the probability (or in the differential equation setting, rate) of two people coming into effective contact with one another (i.e. a contact of sufficient closeness so as to transmit the illness if one is susceptible and

the other is infectious), and marks the transition from "Susceptible" to "Exposed".[34] δ represents the transition from "Exposed" to "Infected" and implies that a person is now able to spread the illness to others.[34] γ represents the transition from Infected to Recovered and is typically estimated using the inverse of the mean duration of illness.[34] For instance, if we assume that it takes a mean of 14 days to recover from COVID-19 then γ would equal 1/14 or 0.071. [34]

Equations for the standard SEIR model are given below:

$$dS/dt = -\beta SI$$
$$dE/dt = \beta SI - \delta E$$
$$dI/dt = \delta E - \gamma R$$
$$dR/dt = \gamma R$$

Note that the transition from Susceptible to Exposed is controlled by an interaction term. This is to reflect that transmission can only happen when a susceptible person and an infected person come into contact with one another.[34]

In this aim, we give each node of our network their own set of these equations to track the number of susceptible, exposed, infected, and recovered individual are in each unit at any given time. Then we use transition rates between units, given by the inverse of the mean waiting time for transition from one unit to another, to model the gross movement of people between compartments.

2 Has the Relationship Between Wealth and HIV risk in Sub-Saharan Africa Changed over time? A Temporal, Gendered, and Hierarchical Analysis.

2.1 Introduction

At a global level, HIV is widely understood to be a disease of inequality in which poorer individuals experience disproportionate morbidity and mortality from HIV infection [1–4]. This association has often been assumed to be consistent across social and geographic contexts, with poorer individuals in poorer countries at greatest risk of HIV infection [2]. However, it is now well established that in Sub-Saharan Africa (SSA), both wealthier individuals and wealthier countries have higher HIV prevalence than their poorer counterparts (Gillespie et al., 2007; Hajizadeh et al., 2014).

The relative vulnerability of wealthier people runs contrary to findings from North America and Western Europe over the last two decades, where wealth is often strongly protective against infectious disease [2,3,36–39].

A positive relationship between wealth and HIV risk also appears to violate key understandings of how health disparities are generated and maintained. For example, Link and Phelan's fundamental cause framework suggests that as knowledge and the availability of new tools for treatment and prevention of HIV have developed, wealthier individuals and countries should be able to better avoid illness and prevent death compared to poorer individuals and countries [36,39,40]. One potential explanation for this paradox is that the wealthy and educated may

feel insulated from the risk of HIV infection and therefore do not take the same precautions others do [2,31].

Similarly, despite most residents of SSA living in rural rather than urban settings, when HIV prevalence is stratified by urbanicity urban areas have a higher prevalence of HIV than rural ones [41]. This suggests that HIV prevalence cannot be fully explained by differences in urbanicity [1–9,42–44].

Lastly, it has also been observed that women in SSA experience a greater risk of HIV infection compared to men. This is again, a mirror image of results from North America and Western Europe and has been attributed to greater biological vulnerability to acquisition [45–47] as well as the greater presence of quid-pro-quo sexual relationships in which wealthy men have multiple female partners [5,7,43,48,49].

In this analysis, we advance the literature in several directions: Existing individual-level studies of HIV, gender and wealth have typically employed data from a specific year for a single nation or a small subset of SSA nations [2,3,5–9,50]. Further, although these studies enable between-country comparisons for a particular year, such analyses do not allow for an assessment of how this relationship varies across time. However, there have been important changes over the last two decades which have likely affected this relationship. The widespread roll-out of free anti-retroviral therapy across the continent has both improved survival of especially poorer people with HIV who may not have had prior access, as well as lowered HIV incidence by reducing the national viral load, and thus reducing the likelihood of new acquisition among new cohorts of individuals [51,52]. Additionally, over the last two decades, knowledge about HIV prevention has increased along with HIV testing and condom use, especially among young women and men [53–

[58]. The achievement of gender parity in education in many African countries [59–62] may also have increased women's leverage to demand safe sex in relationships, thereby changing the rate at which women acquire HIV from their male partners [63,64]. Finally, the percentage of sub-Saharan Africans living in urban environments has dramatically increased, from a mean of 27.3% in 1990 to a mean 40.7% in 2019 [41]. Overall, these studies suggest that drivers of the long-established relationship between wealth and HIV in Africa have shifted.

In this study, we draw on DHS and AIS data from 27 sub-Saharan countries to investigate whether and how the relationship between wealth, gender and urbanicity has changed over time. We also account for hierarchical dimensions of wealth, examining both individual wealth and country level wealth, and assessing how these relationships have changed across time.

2.2 Methods

2.2.1 Data

We performed a secondary data analysis on a secondary de-identified dataset from 43 nationally representative cross-sectional Demographic and Health Surveys (DHS) and AIDS Indicator Surveys (AIS), covering 27 countries with linked HIV test results [65].

These data spanned 14 years; from 2003-2016 (Table 1)[65]. Although data were available, Burkina Faso 2003, Sierra Leone 2013, Guinea 2005 & 2012, Niger 2006, Togo 2013, Mali 2012, Zimbabwe 2005 & 2015, but not Zimbabwe 2010, were dropped from this analysis due to low or 0 cell counts for HIV positivity when stratified by wealth tertiles and urbanicity.

To obtain nationally representative statistics DHS and AIS follow a two-cluster sampling design. Countries are first broken into enumeration areas based on national census data. An effort is made to use existing census enumeration areas if they are available. A subset of these

enumeration areas is then selected with a probability proportional to their population size [65]. Within these areas, households are randomly selected to participate [65].

Country with Abbreviation	HIV Positive (%)	Women (%)	Urban (%)
eSwatini (SZ)	25.88	54.03	27.38
Lesotho (LS)	23.66	54.48	29.61
South Africa (SA)	20.98	49.26	67.95
Zimbabwe (ZW)	15.69	52.76	35.59
Namibia (NM)	14.33	53.51	54.62
Zambia (ZM)	13.68	50.7	45.19
Malawi (MW)	10.25	51.56	18.89
Kenya (KE)	6.52	51.84	24.68
Uganda (UG)	6.37	55.24	14.36
Tanzania (TZ)	5.36	54.7	25.57
Cameroon (CM)	4.68	50.62	55.05
Gabon (GA)	4.24	49.67	87.65
Côte D'Ivoire (CI)	3.97	50	51.26
Rwanda (RW)	3.05	52.85	17.34
Ghana (GH)	2.1	51.77	49.63
Angola (AO)	1.95	54.22	70.79
Gambia (GM)	1.95	52.63	58.74
Liberia (LB)	1.83	54.08	48.81
Tchad (TD)	1.56	52.73	25.74
São Tomé & Príncipe (ST)	1.54	50.49	52.15
Sierra Leone (SL)	1.47	53.26	36.04
Burundi (BU)	1.43	52.78	12.55
Mali (ML)	1.34	52.48	35.39
Ethiopia (ET)	1.19	51.5	22.04
D.R. Congo (CD)	1.17	51.67	39.6
Burkina Faso (BF)	1.02	53.96	27.78
Senegal (SN)	0.69	54.31	52.6

Table 2-1 Weighted Statistics of Study Population by Country. Weighted Population, Weighted HIV Prevalence Weighted Proportion of Men

2.2.2 Informed Consent and Data Privacy

Survey procedures and questionnaires were reviewed by both the ICF Institutional Review Board (IRB) and an IRB within the country being surveyed [66]. The ICF review ensured that protocols were compliant with US regulations on human subjects research, and the within country IRB review ensured that the survey complied with the country's laws [66]. Prior to participation in the survey and biomarker sampling, adult respondents underwent an informed consent process that emphasized the voluntary nature of the survey [66].

Identification numbers that included enumeration areas and household, were used in lieu of names on surveys and biospecimens [66]. Once survey data processing was completed, the section of the survey containing this number was destroyed and the enumeration area and house number codes were randomized and reassigned [66]. All survey and biomarker data was strictly confidential [66]. More detailed information can be found in the Methodology Section of the DHS website [66].

2.2.3 Variables

In our analysis, we investigated the relationships between three variables – wealth, gender, and urbanicity – and HIV infection.

2.2.3.1 HIV.

HIV infection status at the time of the survey was operationalized to a binary variable, 'infected' or 'not infected'. 'Infected' status reflects a positive test result for HIV-1, HIV-2, or both, while "not infected" reflects an individual with a negative result for all HIV types. Individuals with 'invalid" or "indeterminate" test results were dropped from the study. These

observations comprised less that 1 percent of the data. HIV specific weights calculated by DHS were used to estimate measurements of association for the entire population.

2.2.3.2 Individual Wealth.

An individual's wealth was quantified using the DHS Wealth index. The Wealth Index is a composite measure of an individual's assets. These include money, livestock, transportation (bicycles, cars, motorbikes), and home appliances (radios, refrigerators), among other items [67]. The index is mean-centered, with zero representing the mean level of wealth within a country during a survey period [67]. This makes the wealth index a more useful tool for understanding how differences in relative socioeconomic position within each country relate to within-country differences in HIV risk rather than for understanding the impact of country-independent absolute differences in wealth. These are instead captured by the relationship between national GDP and country-level average incidence (see below). To facilitate interpretation of the wealth index in these terms, we collapsed the continuous wealth score in the survey data into lower, middle, and upper wealth index tertiles for each year and country combination in an approach similar to Magadi and colleagues (Magadi et al 2017).[2]

It is important to note that this standardized measure is contextual. A potential drawback is that an individual within the lower wealth tertile of one country, may fall into the middle or upper wealth tertile of another country. This may thus limit comparability across countries. However, the relative nature of this measure could also be a benefit. The wealth tertiles become a proxy for the *relative* lived experience and access to resources afforded to someone based upon

[2] We elected to create tertiles instead of using the quintiles calculated by DHS, for parsimony and because the DHS indicates that this measure is amenable to recategorization into different quantiles.

their wealth within a given country, instead of a mere categorization of the assets they have. This allows us to get at the experience of being in the lower, middle, and upper wealth tertile of each country, independent of the total value of one's assets.

2.2.3.3 Country-Level Wealth.

We used three measures of wealth and inequality at the country level downloaded from the World Bank's online data portal at *https://databank.worldbank.org/* [41]: the GINI coefficient, which measures wealth disparity; the human development index (HDI) which is a composite measure of a country's gross domestic product (GDP) per Capita, life expectancy and education level, [69,70]; and health expenditures as a percentage of gross domestic product, which is a measure of the amount of money a country is spending on health care. These country-level economic metrics were used to ascertain if a relationship exists between a country's economic strength and its HIV prevalence.

2.2.3.4 Gender.

Since DHS data does not have a discrete variable representing gender identity, the biological sex of each survey participant was used as a proxy.

2.2.3.5 Urbanicity.

Urban or rural designation is assigned to an individual based upon the location of an individual's *de facto* residence. There is no standard definition of what constitutes an urban versus rural area. Instead, these designations are assigned based upon the country's particular characteristics. Country-specific definitions can be found via Integrated Public Use Microdata Series DHS website [71].

2.3 Statistical Analysis

To estimate gendered, temporal, and hierarchical risks of HIV infection while accounting for variation in the magnitude and direction of these effects across settings, we employed a two-step modeling approach, often used in meta-analyses[10].

2.3.1 Step 1- Individual data

In the first step, we modeled the odds of HIV infection for each country/year combination using a logistic regression model adjusted for wealth and gender. For each country/year combination, the effect of living in a rural vs. urban setting on overall prevalence as well as the impact of wealth and gender was captured by stratifying by setting. Step 1 analysis was completed using the *survey* package in R 3.5, which allowed us to include DHS-provided sample weights to obtain population-level inferences from DHS's complex survey data [72,73].

Once effect measures for gender and wealth on HIV infection by urbanicity were obtained in Step 1, in Step 2 we employed a meta-regression model to measure relationships between country-level measures and HIV infection risks associated with being in the middle and upper wealth tertiles or being a woman across all the countries in our analysis.

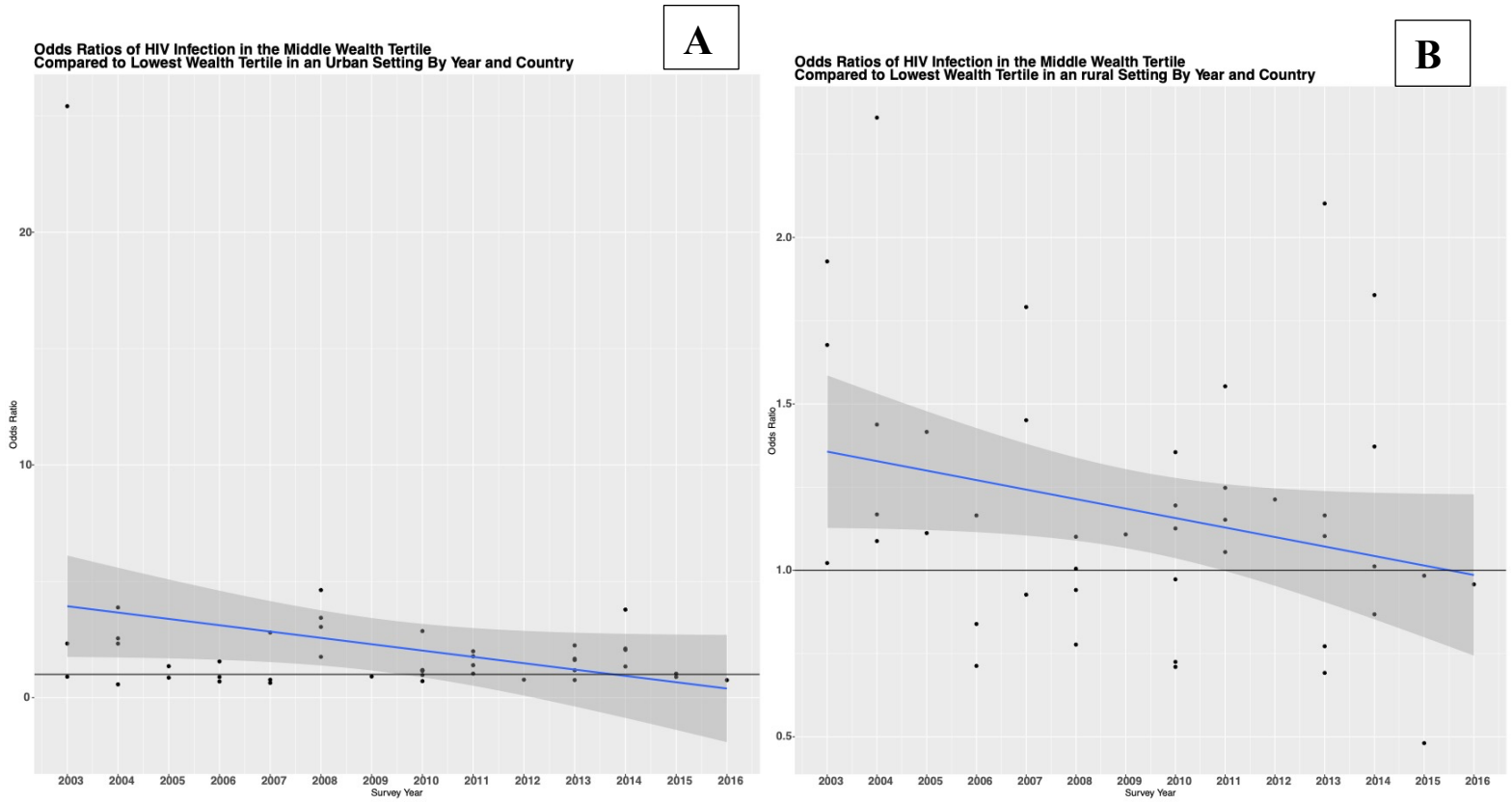


Figure 2-1 Odds Ratios of HIV infection by Middle Wealth Tertile per Country and Survey Year, With Meta-Regression Trend Overlay

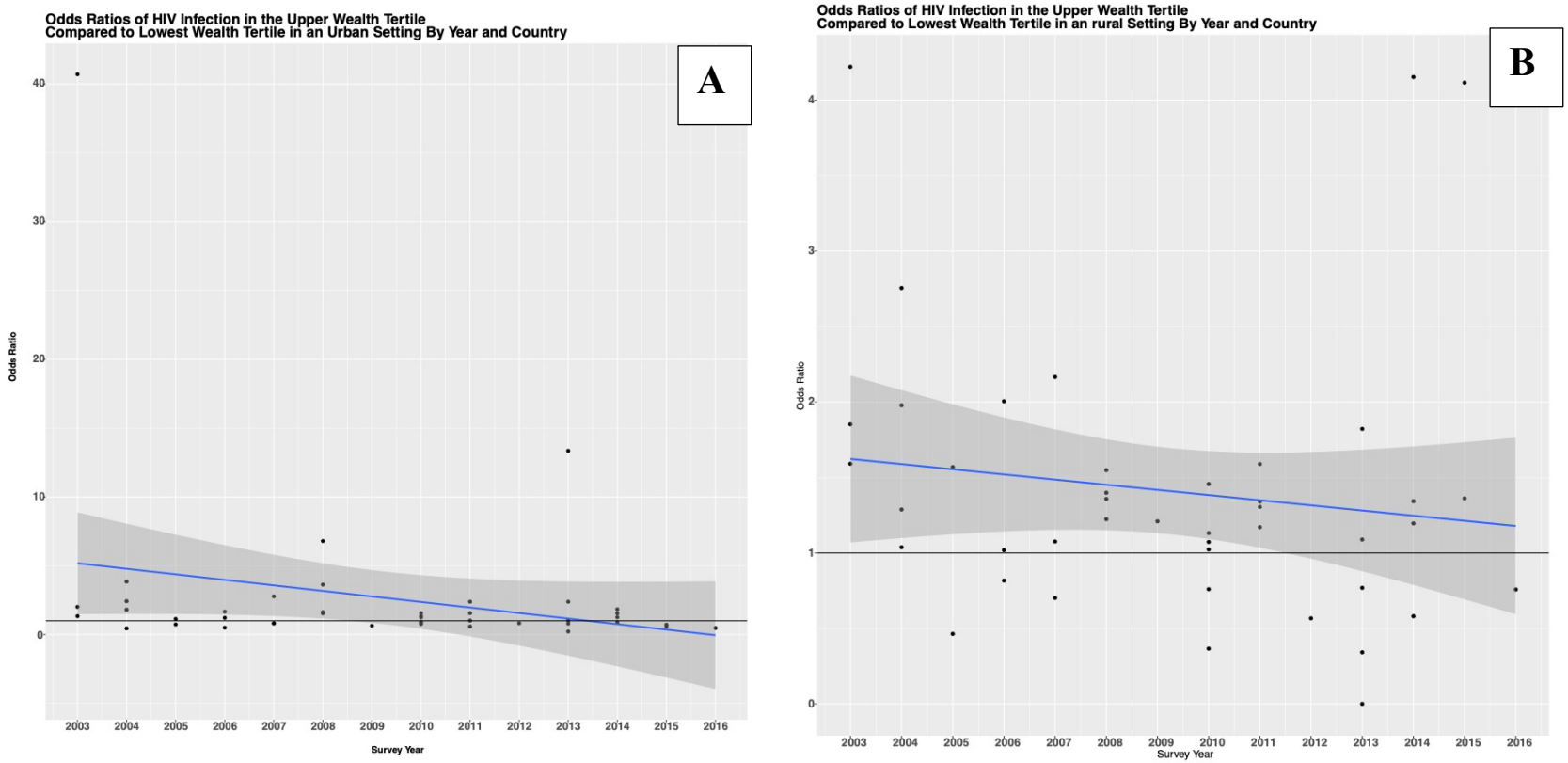


Figure 2-2 Odds Ratios of HIV infection by Upper Wealth Tertile per Country and Survey Year, With Meta-Regression Trend Overlay

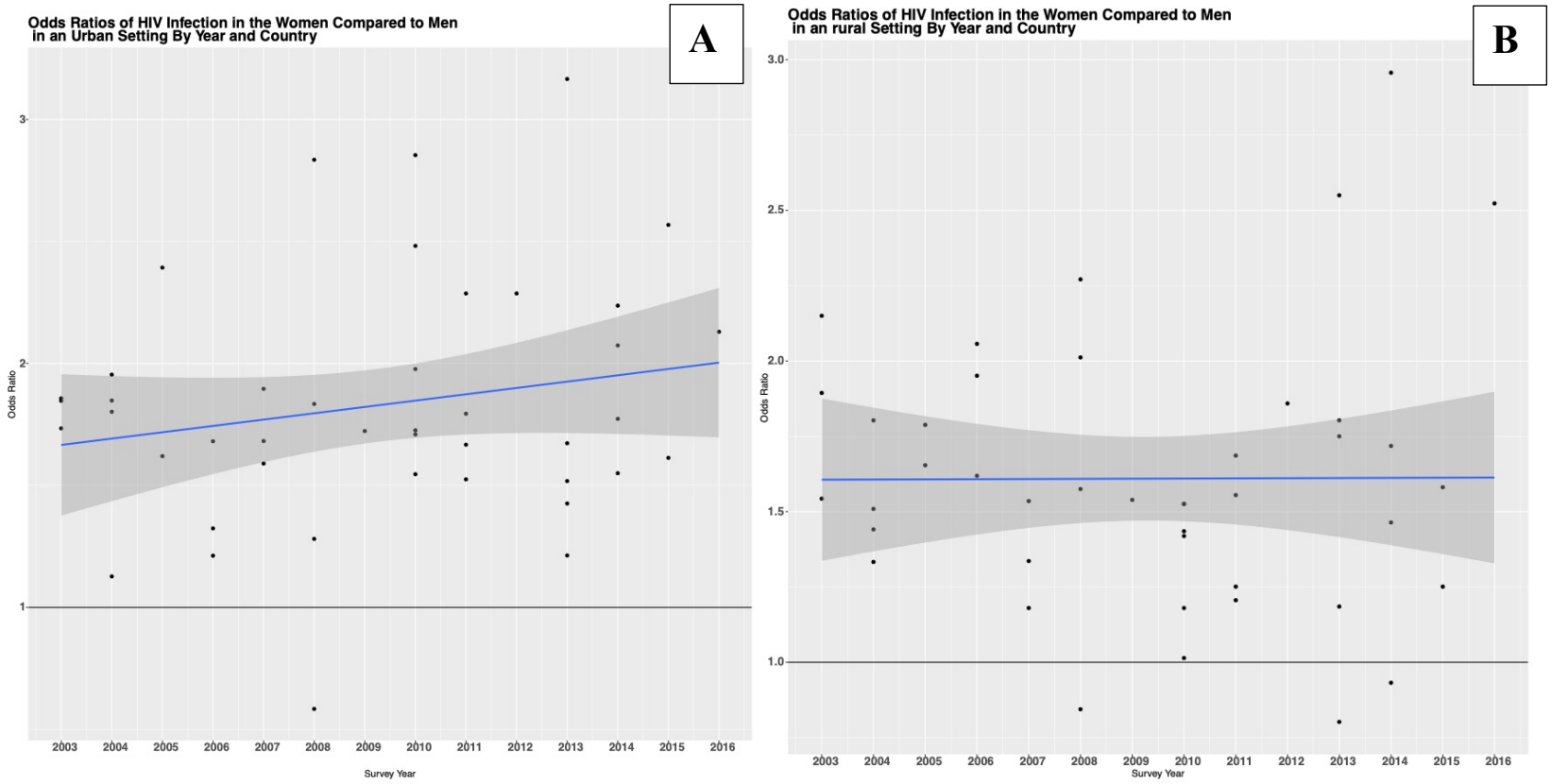


Figure 2-3 Odds Ratios of HIV infection by Gender per Country and Survey Year, With Meta-Regression Trend Overlay

2.3.2 Step 2- Country-Level data

The objective of the second step was to assess if a country's GDP, HDI, or GINI was predictive of HIV burden in either an urban or rural setting for the groups mentioned above. Since these economic measures capture different dimensions of wealth distribution and economic robustness, statistically significant relationships between them and HIV infection might offer insight into why certain groups have a higher burden of disease compared to others.

To do this, a mixed-effects meta-regression model was fit using the parameter estimates and confidence intervals from the Step 1 analysis [74]. Meta-regression was specifically used in this context as it can adjust the parameter results to account for any uncertainty in the effect estimates obtained from the models in Step 1. A mixed-effects model was used to account for residual between-country and survey\ variability not captured by included covariates [74]. An additional benefit of using a mixed-effects model is that it does not assume that covariate effects are uniform across countries [74]. Meta-regression analyses were implemented with the *Metafor* package for R [75].

2.4 Results

2.4.1 Descriptive Statistics

2.4.1.1 HIV.

The prevalence of HIV across the countries analyzed ranged from 0.69% in Senegal, to 25.88% in eSwatini. 6 other countries had an HIV prevalence of greater than 10%: Lesotho, 23.66%; Malawi, 10.25%; Namibia, 14.33%; South Africa, 20.98%; Zambia, 13.68%;

Zimbabwe, 15.69%. All other countries had an HIV prevalence of greater than 1% but less than 10% (Table 2.1).

2.4.1.2 Urbanicity.

The urban populations within the countries studied range from 12.6% in Burundi to 87.7% in Gabon. Seven other countries had an urban population greater than 50%: Angola, 70.8%; Cote D'Ivoire, 51.3; Cameroon, 55.1%; Gambia, 58.7%; Namibia, 54.6%; Senegal 52.6%; São Tomé & Príncipe, 52.2%; and South Africa, 68% (Table 2.1).

2.4.1.3 Gender.

The proportion of women in each survey varied from 49% to 55.25%. South Africa had the lowest proportion at 49.26%, and Uganda, the greatest with 55.24%. The median proportion of women in each country is 52.7% (Table 2.1).

2.4.2 Wealth

After stratifying by urbanicity and controlling for gender we found that, similar to prior stratified findings, those living in urban settings were at greater risk of infection compared to those in the same wealth tertile living in a rural setting (Figure 2.1 A, B, 2.2 A,B)[8,9,35]. In 20 of 27 countries, urban dwellers in the middle wealth tertile had a higher odds of HIV infection compared to their rural counterparts. Of these, urban dwellers in 14 of these 20 countries had 50% greater odds of infection than rural dwellers. Congo 2011 was the only survey in which urban dwellers in the upper tertile's odds of HIV was 50% than that of rural upper tertile dwellers. (Figure 2.1 A,B). Likewise, when comparing urban and rural individuals in the upper wealth tertile, we also found that the odds of HIV were higher for those in urban vs. rural settings

in 19 countries, with urban odds ratios that were at least 50% greater than their rural counterparts in 10 countries. (Figure 2.1 A,B).

2.4.2.1 Meta-analysis of urban strata over time

Results of our cross-country meta-model indicate that, on average, the odds of HIV decreased by 3.4% per year for the urban middle versus the urban lowest wealth tertile, and by 7.1% per year for urban upper versus the urban lowest wealth tertile (Figures 2.1,2.2 A). Both meta regression lines and their confidence intervals began and remained above the null for several years. This suggests that within our data, the odds of HIV infection seen in the urban middle tertile or urban upper wealth tertile were significantly greater than those of the reference group. As suggested by the slope these relationships have attenuated over time. From 2008 onward the lower confidence bound for the urban upper wealth tertile estimate included the null (Figure 2.1A). Likewise, from 2011 onward the lower confidence bound for the urban middle wealth tertile included the null (Figure 2.2A). These findings suggest that over time, the association between wealth and HIV infection has weakened over time in urban areas.

2.4.2.2 Meta-analysis of rural strata over time

Similarly, results of our cross-country meta-model indicate that, on average, the odds of HIV decreased by 2.2% per year for the middle versus the lowest wealth tertile and by 7.6 % per year for the upper versus lower wealth tertiles (Figures 2.1 B, 2.2 B). The meta regression line and confidence intervals for the middle wealth tertile odds ratios remain above the null between the years of 2003- 2008. This finding suggests that within our data those individuals in the middle tertile had significantly higher odds of HIV infection compared to the reference group during the 2003-2008 time interval. From 2008 onward, however, the two groups had similar

odds of HIV infection. Conversely, the regression confidence intervals for the upper wealth tertile included the null for all years analyzed. This suggests that within our study population, wealth has weakened as a predictor of HIV infection amongst rural populations in a manner that is similar to what has occurred in more-urban areas. Finally, the prevalence of HIV amongst the rural poor and the urban poor has remained largely unchanged over the years surveyed, with the urban poor bearing a higher burden of disease compared to the rural poor overall. Likewise, the ratio of these two prevalence has also remained fairly constant over time.

The meta-regression findings suggest that the relationship between wealth and HIV may be starting to move in the direction typically observed outside of SSA, where an inverse relationship is seen between assets and the risk of HIV infection. We found a decreasing odds of HIV among middle and upper tertile Africans, regardless of urbanicity, with the wealthiest SSA residents experiencing the largest declines. While one may conclude that this relationship is simply the result of survival bias, wealthier individuals in SSA are less likely to die from AIDS, even when anti-retroviral therapy is freely available, suggesting that these declines are unlikely to be explained by this mechanism alone. [9,76].

2.4.3 Gender

Figures 2.3 A and 2.3 B illustrate the odds ratios of HIV for women compared to men across time and country by urbanicity. Overall women typically had a higher HIV burden than men (>1 odds ratios), and urban women had a greater burden relative to rural women. 19 of the 27 countries surveyed had a higher odds of HIV infection among urban women compared to rural women. In 6 out of 19 countries, the odds of infection for urban women were at least 50% greater than that of rural women.

Our regression results indicate that women across all countries in our analysis experienced greater odds of HIV infection compared to men after controlling for wealth. The odds of HIV among rural women were 52% greater than the odds of HIV infection in rural men. Similarly, the odds of HIV for urban women are 73% greater than for urban-dwelling men. These finding are consistent with those of Lakew, Barankanira and Hajizedah [5–7]. The change in odds of HIV infection in women versus men in both urban and rural areas, across the surveys used in this analysis, was less than 1% per year. This may indicate that across Sub-Saharan African countries, the relative odds of HIV infection in women, regardless of environment has remained unchanged over the last decade.

2.4.4 Effect Modification and Confounding

When we compare the results from our analysis stratified on urbanicity to the results of an unstratified analysis, we see evidence of effect modification or confounding by urbanicity depending on the variable. Effect modification is indicated by a crude odds ratio falling between the two stratified odds ratios, while confounding is indicated by a crude odds ratio that is higher or lower than both stratified odds ratios [77]. Following these criteria, urbanicity appears to modify the effect of gender on HIV infection, but confound the relationship between wealth and HIV infection. (Figures 2.4-6).

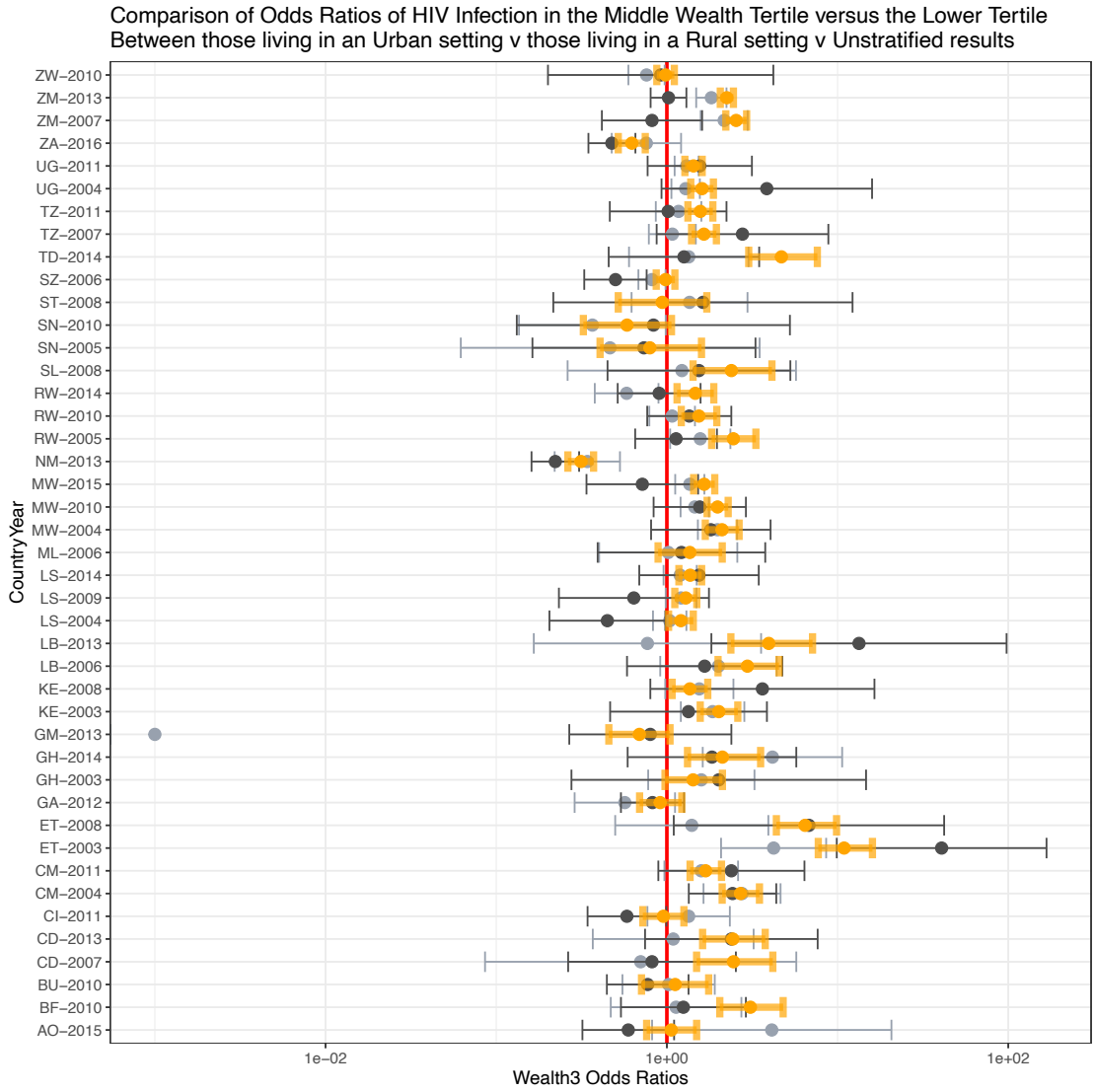


Figure 2-4 Comparison of Urban v Rural Odds Ratios for Upper Wealth Tertile. Orange Confidence intervals reflect unstratified odds ratios. Black confidence intervals reflect Urban ORs, Grey reflect Rural ORs

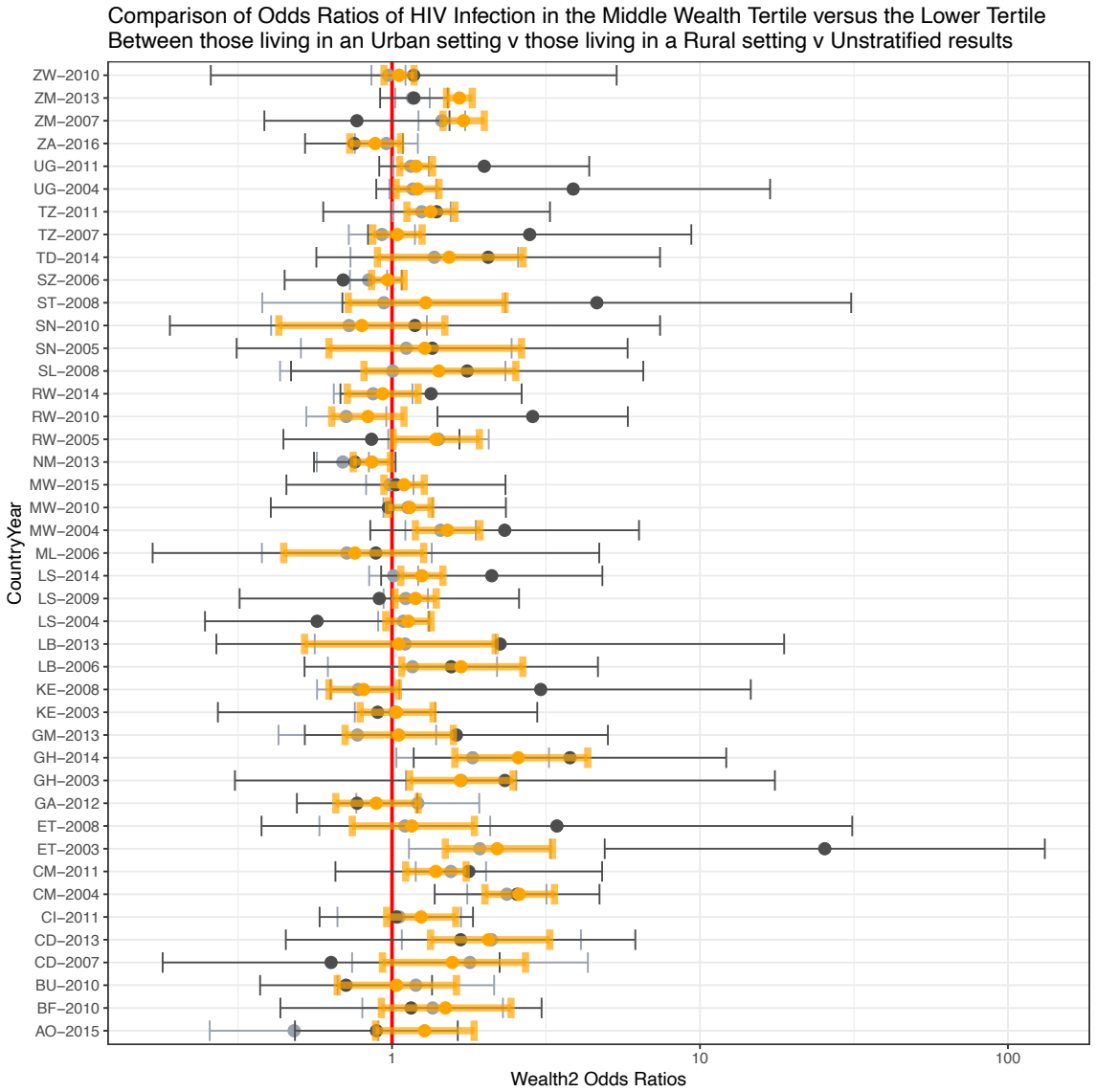


Figure 2-5 Comparison of Urban v Rural Odds Ratios for Middle Wealth Tertile Orange Confidence intervals reflect unstratified odds ratios. Black confidence intervals reflect Urban ORs, Grey reflect Rural ORs.

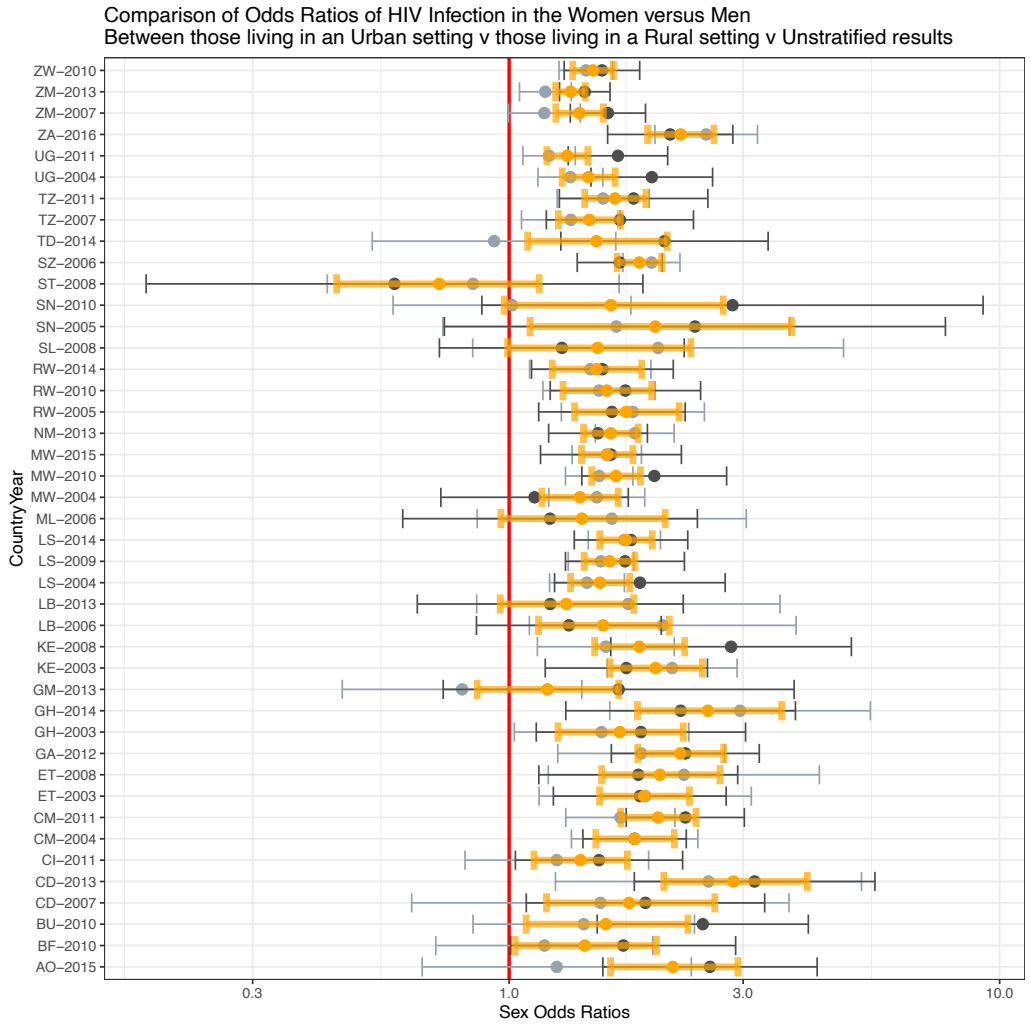


Figure 2-6 Comparison of Urban v Rural Odds Ratios for Gender Orange Confidence intervals reflect unstratified odds ratios. Black confidence intervals reflect Urban ORs, Grey reflect Rural ORs.

Meta-Regression Results for the Middle Wealth Tertile

Metrics	Odds Ratio	95% Confidence Interval	Change in Odds of HIV Infection per Decile Increase of Metric
GDP per Capita	0.6	0.32,1.14	40% Decrease
GINI	1.02	0.69,1.5	2% Increase
HDI	0.8	0.50,1.28	20% Decrease
HE.GDP	0.67	0.45,0.99	33% Decrease

Table 2-2 Predicted Change in the Odds Ratio of HIV in the Middle Wealth Tertile Given World Bank Metric

Meta-Regression Results for the Upper Wealth Tertile

Metrics	Odds Ratio	95% Confidence Interval	Change in Odds of HIV Infection per Decile Increase of Metric
GDP per Capita	0.33	0.10,1.07	67% Decrease
GINI	0.4	0.20,0.81	50% Decrease
HDI	0.2	0.09,0.45	80% Decrease
HE.GDP	0.3	0.14,0.62	71% Decrease

Table 2-3 Predicted Change in the Odds Ratio of HIV by Upper Wealth Tertile Given World Bank Metric

Meta-Regression Results for Gender

Metrics	Odds Ratio	95% Confidence Interval	Change in Odds of HIV Infection per Decile Increase of Metric
GDP per Capita	0.87	0.59,1.28	13% Decrease
GINI	0.86	0.59,1.06	14% Decrease
HDI	0.85	0.66,1.09	15% Decrease
HE.GDP	0.93	0.75,1.15	7% Decrease

Table 2-4 Predicted Change in the Odds Ratio of HIV by Gender Given World Bank Metri

2.4.5 Country-level inequality in wealth and gender risks

The results of our analyses can be found in Tables 2.2 through 2.4. These tables depict how the country-level economic metrics predicted the odds of HIV infection amongst those across Sub-Saharan Africa in the middle wealth tertile for their country compared to those individuals across Sub-Saharan Africa that are in the lowest wealth tertile for their country. For example, those individuals within the Middle wealth tertile for their country will experience a 40% decrease in the odds of HIV infection, compared to those who are in the lowest wealth tertile, for every decile increase in their country's GDP per Capita. Table 2.3 repeats the findings of Table 2.2, except that these reflect how country level economic metrics predict HIV infection in those individuals that fall into the upper wealth tertile for their country of residence, compared to those who fall into the poorest wealth tertile for their country of residence. Table 2.4, reflects the change in the odds of HIV infection predicted by the economic metrics for all women in Sub-Saharan Africa versus all men in Sub-Saharan Africa.

Our findings suggest that a country's development as well as the amount of money spent on health expenditures also have a significant impact on the relationship between wealth and HIV, however this is only for those in the upper wealth tertile. We did not find a significant relationship between gender or middle wealth tertile status and the country level metrics used here.

Two metrics were associated with lower HIV risk for the upper wealth tertile relative to the lowest: HDI and the percentage of the GDP attributed to health expenditures. A 1 unit increase in HDI was associated with a 94% decrease, (95% Confidence Interval (CI): 69%, 99%), in HIV in the upper wealth tertile as opposed to the lowest wealth tertile. A 1% increase in

the percent of the GDP directed to health expenditures was associated with a 16.2% decrease (95% CI: 3%, 28%) in HIV among the upper wealth tertile compared to the lowest wealth tertile.

2.5 Discussion

This study sought to examine gendered, temporal and hierarchical dimensions of the relationship between wealth and HIV over time in sub-Saharan Africa. Our analyses revealed three overall findings: First, we found that while wealthier Africans continue to experience higher odds of HIV infection compared to poorer individuals, the risk of HIV infection among middle and wealthier individuals has reduced over time relative to the poor in both urban and rural areas, with the wealthiest experiencing the largest declines. Second, we found that women continue to have higher odds of HIV infection relative to men, even after accounting for wealth and urbanicity, and there has been little change in this relationship over time. Third, we found that country level GDP, its HDI, and its expenditures on health care reduced the relative risk of HIV infection, but only for the wealthiest Africans in their respective countries.

Our results confirm Magadi's finding that the risk of HIV infection is higher in those living in an urban setting compared to individuals living in a rural setting, although our results regarding wealth and HIV conflict with her findings. This is potentially due to our use of wealth index tertiles as opposed to splitting the wealth index at the median [8]. Our findings regarding wealth are consistent with those by Mishra et. al., and Gillespie et. al; namely that wealthier individuals in SSA continue to experience higher odds of infection compared to poorer individuals, and that relative wealth not poverty predict one's risk of HIV infection [9,35].

Our investigation of temporal trends indicates that the disparity in HIV infection risk between wealth tertiles has been declining over time in both urban and rural areas. Our findings

note that the overall average decline in HIV infection risk amongst individuals in the upper wealth tertiles is steeper than the decline in HIV in the middle wealth tertile, despite increasing availability of free antiretroviral therapy across SSA [51,78] (Figure 2.1,2.2 A-B).

There are several potential explanations for our findings. First, since wealthier individuals are less likely to die from AIDS [76,79–83], these results may suggest a shift in the composition of the HIV-positive population, with wealthier adults who acquired HIV earlier in the epidemic surviving longer, combined with declining incidence among adults who have accumulated wealth as they age, potentially due to higher uptake of HIV preventive practices. In this case, as time progresses, fewer wealthy SSA residents would be HIV positive. However, since the relationship between wealth and HIV is typically examined in relative terms, this could also reflect compositional shifts among the poor, who have benefited relatively more from freely available ART than wealthier Africans, and are now increasingly able to survive and age with HIV [84–86]. In addition, ART contributes to reduced community viral loads for both the wealthy and poor, providing indirect protection that lowers incidence across the population [51]. In combination, this would result in both the poor and the wealthy trending towards each other over time, potentially blunting the effect of poverty or wealth on HIV outcomes.

Our findings should also be interpreted in light of several limitations. First, there are a varying number of observations in each survey, and a varying number of surveys for any given year. Consequently, our temporal trend estimates are vulnerable to bias due to data sparseness. However, these inconsistencies are unlikely to drastically bias our overall temporal trend estimates because meta-regression uses the standard errors of each point estimate to account for uncertainty.

Not all SSA countries are included in our DHS dataset, which lacks key countries like Botswana; our results about between-country variation and change over time should be interpreted in light of this missing information. While the exclusion of some countries would not affect the individual level estimates of other countries, it may impact temporal trends we observed if these countries are systematically different from those included in our analysis. Finally, we were unable to control directly for the specific phase (acceleration, peak, deceleration) each country's HIV epidemic was in at the time of each survey. This shortcoming is unlikely to bias our results greatly, as most included countries have shown a decline or plateau of cases over the study period, with only a handful increasing. Consequently, any bias in our projections would result in an attenuation towards the null and provide a conservative estimate of the temporal trends.

2.5.1 Conclusion

Taken together, our findings indicate that over the last decade, across a large number of SSA countries, the odds of HIV amongst wealthier individuals have decreased relative to poorer individuals. This suggests that the relationship between wealth and HIV in these countries may be converging towards the global norm in which poverty is predictive of increased HIV risk. We found that the relative difference between those living in urban vs. rural contexts was stable, and that both contexts were largely subject to the same temporal trends in the impact of wealth on HIV risk. Our analyses also confirm the stability of previous findings: HIV continues to be more prevalent in the wealthiest group of individuals compared to the middle or lower wealth tertiles, and more prevalent in women compared to men.

Next, our study identifies a set of metrics that allow us to disentangle the impact of country-level vs. individual attributes on HIV risk. Specifically, increases in health spending as a

fraction of GDP, and HDI are all associated with a decrease in HIV amongst the upper wealth tertile compared to the middle or lower wealth tertiles. Taken together, these trends suggest that wealthier countries have a lower overall burden of disease, but that the wealthiest individuals in these countries remain at higher risk of HIV compared to those at the lower end of the wealth distribution.

Finally, our findings hint at a potential mechanism through which urbanicity affects the relationship between HIV infection, and wealth or gender: urbanicity modifies the effect of gender on HIV infection risk but confounds the relationship between wealth and HIV infection risk. In other words, urbanicity is predictive of wealth and of HIV infection status separately, and as such it must be controlled for if one is to investigate the effect wealth has on HIV infection. However, residing in an urban environment increases a woman's vulnerability to HIV infection.

Our results also point the way to future analyses examining the roles played by wealth and gender on patterns of HIV infection in SSA. In particular, more research should examine the mechanistic drivers of reduced HIV vulnerability among the wealthiest Africans in different countries, the stubbornness of the gender disparity in women's HIV risk, and the positive association between individual wealth on HIV burden.

3 Latent Class Analysis of Generational Trends In STI Risk Factors

3.1 Introduction

Nationally, one-in-five individuals have had a sexually transmitted infection (STI), and more than 26 million new STIs were estimated to have occurred in 2018 alone.[87] While most of these infections were likely mild or asymptomatic, many are important causes of morbidity and mortality. For instance, congenital syphilis has reemerged as public health concern. Reported cases of congenital syphilis in the US have increased by 261% between 2013-2018 in the US, and 6.8% of cases reported in 2019 resulted in either stillbirth or infant death.[87,88] Similarly, reported cases of gonorrhea have increases from 309,341 in 2010, to 583,405 in 2018.[20]

STIs aren't exclusive to young people.[19,20] According to the CDC, reported cases of primary and secondary syphilis amongst those 55 years or older rose from 600 in 2010 to 3,092 in 2020, likewise reported cases of gonorrhea rose from 2,714 in 2010 to 16,333 in 2019.[20] Indeed, other age brackets have seen a similar increase STI cases as well.[20] STI positivity impacts more than just physical health. STIs are often stigmatized, particularly amongst older age groups, which can lead to feelings of disenfranchisement and isolation among cases. [89–94]

It is understood that an interrelationship between sexual behaviors, substance use, and sexually transmitted infection positivity exists.[21,22,24,95,96] However the continuity of these relationships between birth cohorts and across the life course remains less clear. Personal attitudes regarding sex or substance use are influenced by our family and community.[21,97,98] Social mores and stigma can change over time and vary between generations;[89–93] and the frequency with which people engage in these behaviors changes over the life course, as

well.[92,93,99–101] To complicate things further, many specific risk factors or variables measuring sexual behavior and substance use can be highly collinear in some populations, rendering traditional statistical methods less effective at estimating the individual effects of substance use or sexual behaviors have on STI infection risk.[91,102] Indeed, given the many variables that can be used to measure sexual behavior and substance use, there is a need to develop a framework to integrate these different variables in a way that can be used to understand behavioral and STI risk patterns. It is particularly important to understand how these profiles may change over the life course because people of all ages are at risk of contracting and transmitting sexually transmitted infections.[19,20]

Latent class analysis is an alternative approach to understanding an individual's risk, that views individual behaviors or characteristics as expressions of an unmeasurable, underlying risk profile.[26,103] In brief, LCA moves away from estimating the effects attributable to individual sex acts or other risk factors.[26,103] Instead, LCA exploits variable correlation, by comparing variations in the overall response patterns within the data to create a specified number of classes.[28,104] Next, it calculates an individual's probability of belonging to each class, and assigns individuals the class with the highest probability.[28,104] Results are reported as the probability of an individual in a given class endorsing a particular trait.[28,104] With a sufficiently large population, the probabilities can be interpreted as the estimated prevalence of a trait within its class, amongst the population surveyed.

While this approach may be less informative for targeting specific behaviors for intervention, it better identifies the holistic risk to an individual instead of accounting of their risk as a product of their specific behaviors. LCA's goal then, is not to estimate the direct effect of any specific risk factors. Instead, it is to identify recurrent response patterns across our

observed behavior variables (often termed 'manifest variables') in order to identify subgroups that represent different classes or profiles.

In this analysis, we endeavored to characterize such common response patterns, or profiles, of substance use and sexual activity, and to examine how these profiles change across the life course, represented as decadal age groups. Additionally, while we explored how these profiles predicted the risk of any STI, both overall and by age group. To do this, we performed latent class analysis on a set of substance use and sexual behavior variables take from the 2015-2016 wave of the National Health and Nutrition Examination Survey (NHANES).

3.2 Methods

3.2.1 Data Overview

We used the 2015 to 2016 wave of the National Health and Nutrition Examination Survey (NHANES), as the basis for our analyses. NHANES is nationally representative, biennial survey that assesses a range of health behaviors and outcomes via questionnaire and biospecimen collection. These data contained responses from participants ages 18-59. The 2015-2016 wave was the most recent wave to screen for an array of STIs. These data are de-identified and made available for public access online.[105] This study did not require IRB approval.

3.2.2 Conceptual Model Building and Approach

Our latent variable represents an individual's profile as defined by correlates of STI infection. Levels of our latent variable represented different profiles detected in our population.[28,103,106] These profiles were composed of 5 constructs: "never, non-current, or current use of alcohol", "never, non-current, or current use of cigarettes", "never, non-current, or current use of

marijuana", "number of oral sex partners in the last year", "number of vaginal sex partners in the last year".

Although NHANES did offer a "number of anal sex partners in the last year" variable, this question was only asked to men and had a 1.3% response rate amongst the entire population of 4843 participants surveyed. Consequently, we elected not to include this as an indicator in our analysis. Since LCA requires that all variables be categorical, we recoded count variables as categorical variables.[72,107,108]

The three level operationalizations of alcohol, tobacco, and marijuana use were chosen to assess use history and frequency, while also maintaining model parsimony.[21,22,109] On the other hand, sexual partner count can be directly associated with STI positivity, and as such the increased complexity of our model due to these polychotomous variables is offset by the amount of information gained from their inclusion. [110–112] Lastly, yearly partner counts were chosen because they capture behavioral changes in the rate of unique sexual partnerships over the life course. Indeed, it has been observed that substance consumption and sexual activity peak in early adulthood, then taper off over time.[99,100,113] However, as a sensitivity analysis, we have also included a version using cumulative lifetime partner counts (see Supplementary Information), with broadly similar results.

3.2.3 Latent class analysis:

All latent class analyses were conducted using *Mplus* Version 8.4.[104] We began by obtaining fit criteria for the 2 class through 9 class solutions.[106] To obtain solutions for a nationally representative population, the NHANES survey weights were incorporated via *Mplus*' "WEIGHTS" call.[104] We used the Bayesian Information Criterion to identify the number of classes that best describes our latent variable. BIC was elected over the other information criteria

because it has been shown by Nylund et. al. to accurately detect the optimal solution, across a wide range of samples sizes.[106]

After first conducting an LCA on the entire, unstratified dataset, we wanted to assess if profiles changed with age. The initial dataset was stratified into 4 decadal age groups, 18-29 which at the time the data were collect consisted of "Gen Z", 30-39 or “Millenials”,40-49 or “Gen X” , 50-59 who, at the time of data collection, were known as “Baby Boomers”. The analysis described above was repeated on these individual strata, with the exception that the largest number of classes tested were 6, not 9.

3.2.4 Measurement Invariance:

When making across group comparisons it can be helpful to test if the classes within each age group measure the same construct: in epidemiological terms, do the classes have internal validity? In multi-group LCA this property is known as measurement invariance. Mplus offers a way to check for this property via the “KNOWNGROUP” command, and does so by comparing a model without constraints to one with constraints.[28,104] First a “null” model is run. Here, the number of groups in the data and classes sought are specified, but the proportions of each indicator variable are allowed to vary between groups.[28] Next, an adjusted model is constructed that constrains the indicator variable probabilities for a given class across groups.[28] A likelihood ratio test is conducted using fit statistics from these two models. If the adjusted model fits the data as well as the null model, then one can assert that a class in one group measures the same construct as the corresponding class in a second group.[28] If the adjusted model does not fit the data as well, then one can conclude that item invariance doesn’t hold, and that the underlying construct measured by a class is not the same between groups.[28] (See Supplement B for Calculation).[28]

3.2.5 Regression Analysis:

2015-2016 laboratory results were obtained for chlamydia, trichomonas, herpes simplex virus 1 & 2, oral human papilloma virus, genital human papilloma virus, and HIV tests. We then used the results to derive our outcome: the presence of any STI. The "Any STI" variable was positive if any of the STI tests returned a positive result. Participant class assignments were merged with this composite variable. The lme4 package in R was used to conduct univariate logistic regression, in the stratified and unstratified data sets. Class assignments served as the exposure, and "Any STI" served as the outcome. Odds ratios with Class 1 as the reference group were calculated for each latent class .[114,115]

3.3 Results

Table 3.1 shows the unweighted prevalences of the NHANES variables used in this analysis, both by age group and overall. A weighted table of prevalences can be found in Appendix C.

	ALL		18-29		30-39		40-49		50-59	
	N	%	N	%	N	%	N	%	N	%
Alcohol Consumption										
Never	624	0.16	257	0.22	116	0.12	138	0.15	113	0.13
Current	2411	0.61	699	0.59	600	0.64	547	0.6	565	0.64
Non-Current	454	0.12	113	0.09	93	0.1	128	0.14	120	0.14
Did Not Respond	437	0.11	123	0.1	124	0.13	100	0.11	90	0.1
Cigarette Use										
Never	2513	0.64	888	0.74	564	0.6	604	0.66	457	0.51
Current	802	0.2	202	0.17	215	0.23	170	0.19	215	0.24
Non-Current	608	0.15	101	0.08	152	0.16	139	0.15	216	0.24
Did Not Respond	3	0	1	0	2	0	0	0	0	0
Marijuana Use										
Never	1707	0.43	460	0.39	399	0.43	470	0.51	378	0.43
Current	466	0.12	233	0.2	112	0.12	56	0.06	65	0.07
Non-Current	1249	0.32	361	0.3	288	0.31	270	0.3	330	0.37
Did Not Respond	504	0.13	138	0.12	134	0.14	117	0.13	115	0.13
Oral Sex Partners Past Year										
0	1867	0.48	499	0.42	389	0.42	417	0.46	562	0.63
1	1718	0.44	518	0.43	469	0.5	440	0.48	291	0.33
2 to 4	283	0.07	150	0.13	60	0.06	47	0.05	26	0.03
5 to 9	37	0.01	17	0.01	10	0.01	5	0.01	5	0.01
10 +	17	0	8	0.01	3	0	4	0	2	0
Did Not Respond	4	0	0	0	2	0	0	0	2	0
Vaginal Sex Partners Past Year										
0	1187	0.3	383	0.32	231	0.25	224	0.25	349	0.39
1	2227	0.57	558	0.47	597	0.64	592	0.65	480	0.54
2 to 4	391	0.1	190	0.16	83	0.09	79	0.09	39	0.04
5 to 9	72	0.02	43	0.04	11	0.01	5	0.01	13	0.01
10 +	45	0.01	17	0.01	11	0.01	10	0.01	7	0.01
Did Not Respond	4	0	1	0	0	0	3	0	0	0

Table 3-1 Unweighted Prevalence of NHANES Variables by Age Group

3.3.1 Unstratified Descriptive Latent Class Analysis

A six class solution fit the unstratified dataset the best, and showed an escalation of both substance usage and sexual partnerships. (FIGURE 3.1) Class 1, which represented 29.7% of the population, is titled “Inactive Cigarette Smokers”, as the defining features of this class are no endorsement of sexual partnerships in the last year, and present but relatively low endorsement of cigarette use along with a high probability of refusing to answer questions about alcohol and marijuana consumption. Class 2, accounting for 10.5% of the total population, was titled “Semi-Active Alcohol Drinkers” as this class is defined by current or non-current alcohol consumption (but little tobacco use), and some vaginal sex partnerships, but not oral sex partnerships. Class 3, which accounted for 25.9% of the population, was titled “Semi-Active Substance Users” as we see endorsement of current or non-current alcohol, cigarette and marijuana consumption, and some vaginal sex partnerships, but very low endorsement of oral sex partnerships. Class 4 which represented 10.1% of the population, was titled “Monogamous Alcohol Drinkers” and is distinct from “Semi-Active Alcohol Drinkers” in that this class has a high probability of endorsing one oral sex and vaginal sex partnership in the last year. Class 5, which accounted for 12.8% of individuals, was named “Monogamous Substance Users” because like the prior class, people in this class have a high probability of endorsing 1 oral sex and vaginal sex partnership in the last year. Finally, class 6 which accounted for 11.1% of the individuals in our population, was named “Non-Monogamous Substance Users” as they have a high probability of endorsing current or non-current substance use and having more than 1 oral and vaginal sex partnership in the last year. (FIGURE 3.1)

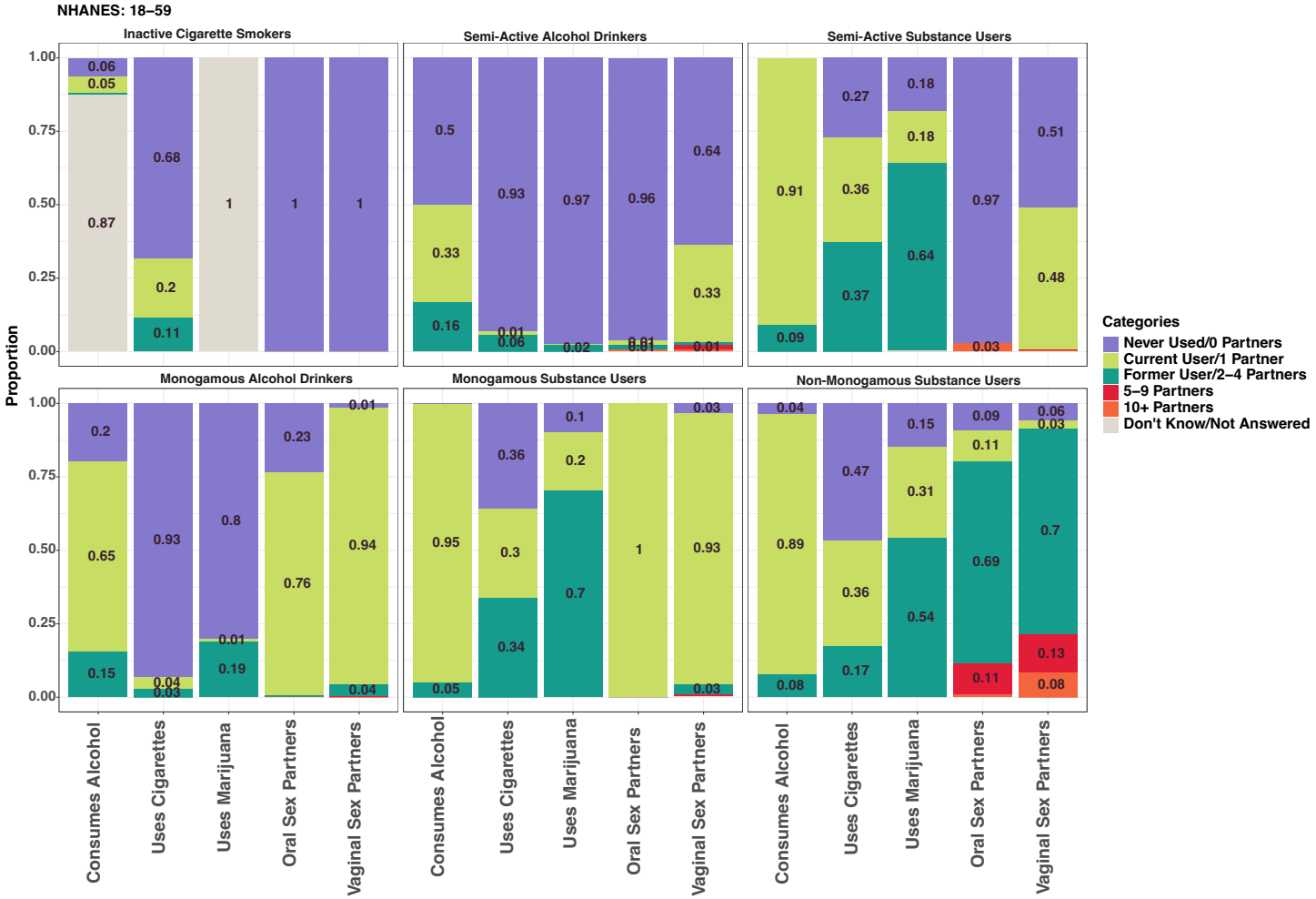


Figure 3-1 6 Class LCA Solution for Unstratified Data

3.3.2 Stratified Descriptive Latent Class Analysis

The five-class solution fit each of the age strata best. The general composition of the 5 classes was broadly consistent across the "18 to 29", "30 to 39", "40 to 49", and "50 to 59" age groups. Class 1 represented individuals with 0 sexual partnerships in the last year who endorse some tobacco use, here called "Inactive Cigarette Smokers". Class 2 composed of people with low probability of having at least one sexual partnership in the last year, and who endorse current alcohol consumption, or "Inactive Alcohol Drinkers"; Class 3 members were dubbed "Monogamous Drinkers", as they had a very high probability of endorsing having 1 relationship in the last year and endorsing current alcohol consumption. and Class 4 represented "Monogamous Substance Users" as these individuals also tended to endorse having 1 sexual partnership in the last year, but unlike "Monogamous Drinkers" also endorsed current tobacco and marijuana usage. The last class given the title "Non-Monogamous Substance Users" as these individuals also had a high probability of endorsing current substance use but had a low probability of having one or no sexual partnerships in the last year. Figure 3.2 is a conceptual diagram comparing the intensity of risk behaviors by class and age group. Figures 3.3-3.6 show the composition of classes within an age group. (Figures 3.2 – 3.6) Note that both substance use and sexual behavior can be represented using multiple variables, so this diagram is a low-dimensional simplification, but captures the basic conceptual relationships between substance use and sexual behavior seen in the classes.

With exception of Inactive Alcohol drinkers, individuals in the 18-29 age group had a higher probability of endorsing current marijuana using than members of the same class in the other age groups. This may reflect changing attitudes towards marijuana consumption, and/or an age effect on marijuana consumption.

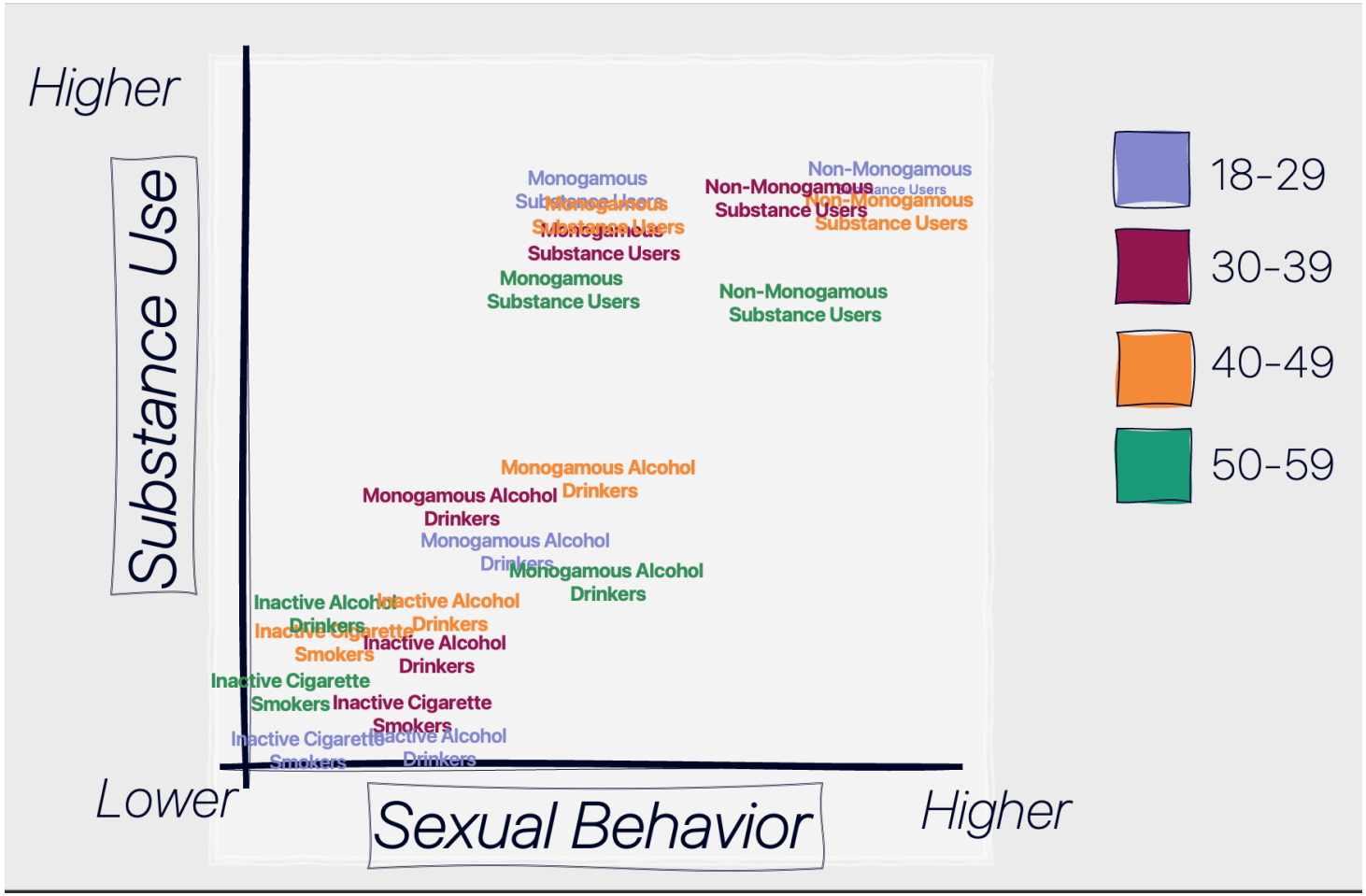


Figure 3-2 Conceptual Diagram Showing the Estimated Intensity of Risk Behaviors by Class and Age Group.

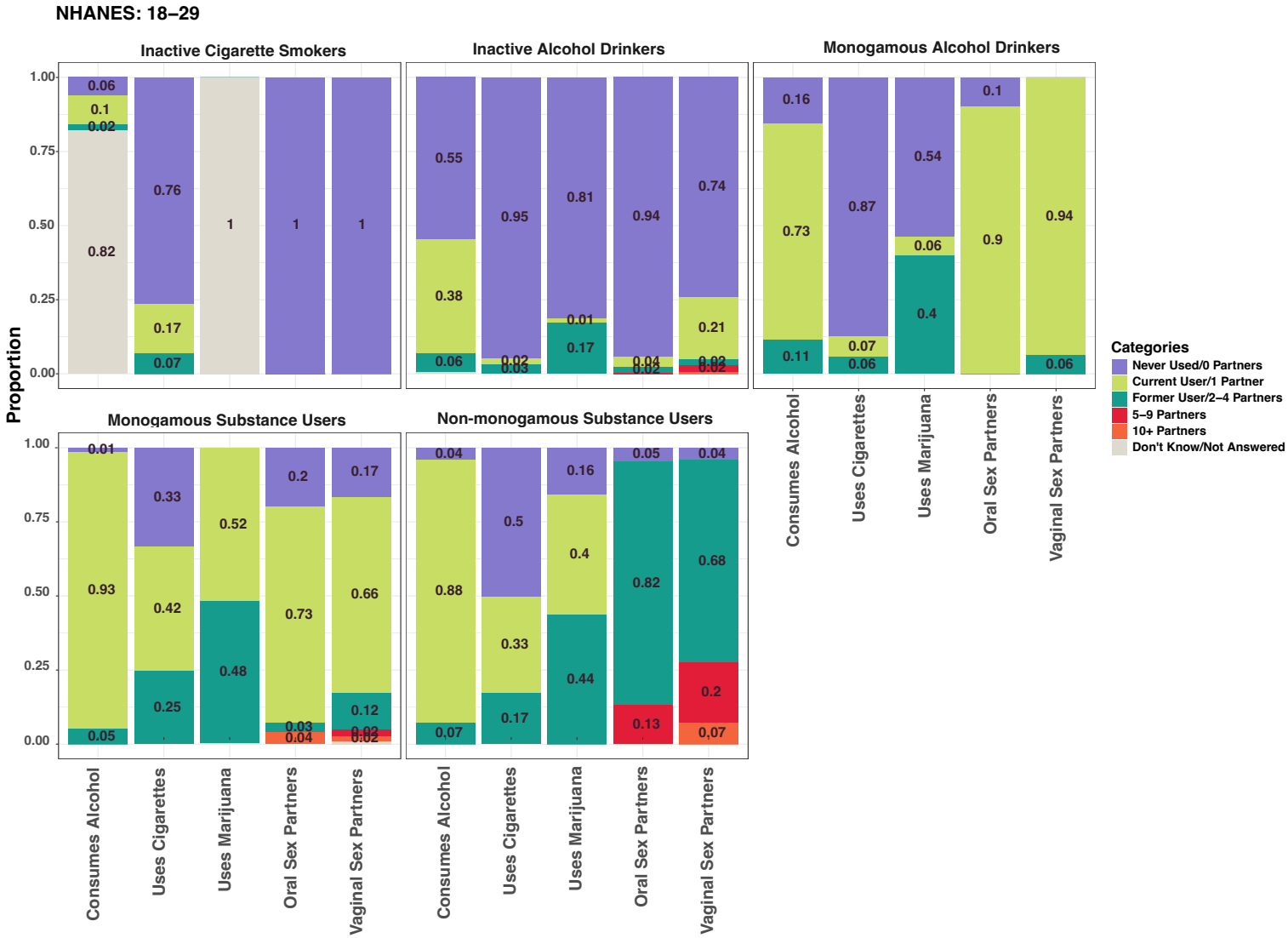


Figure 3-3 5 Class LCA Solution for 18-29 Year Olds

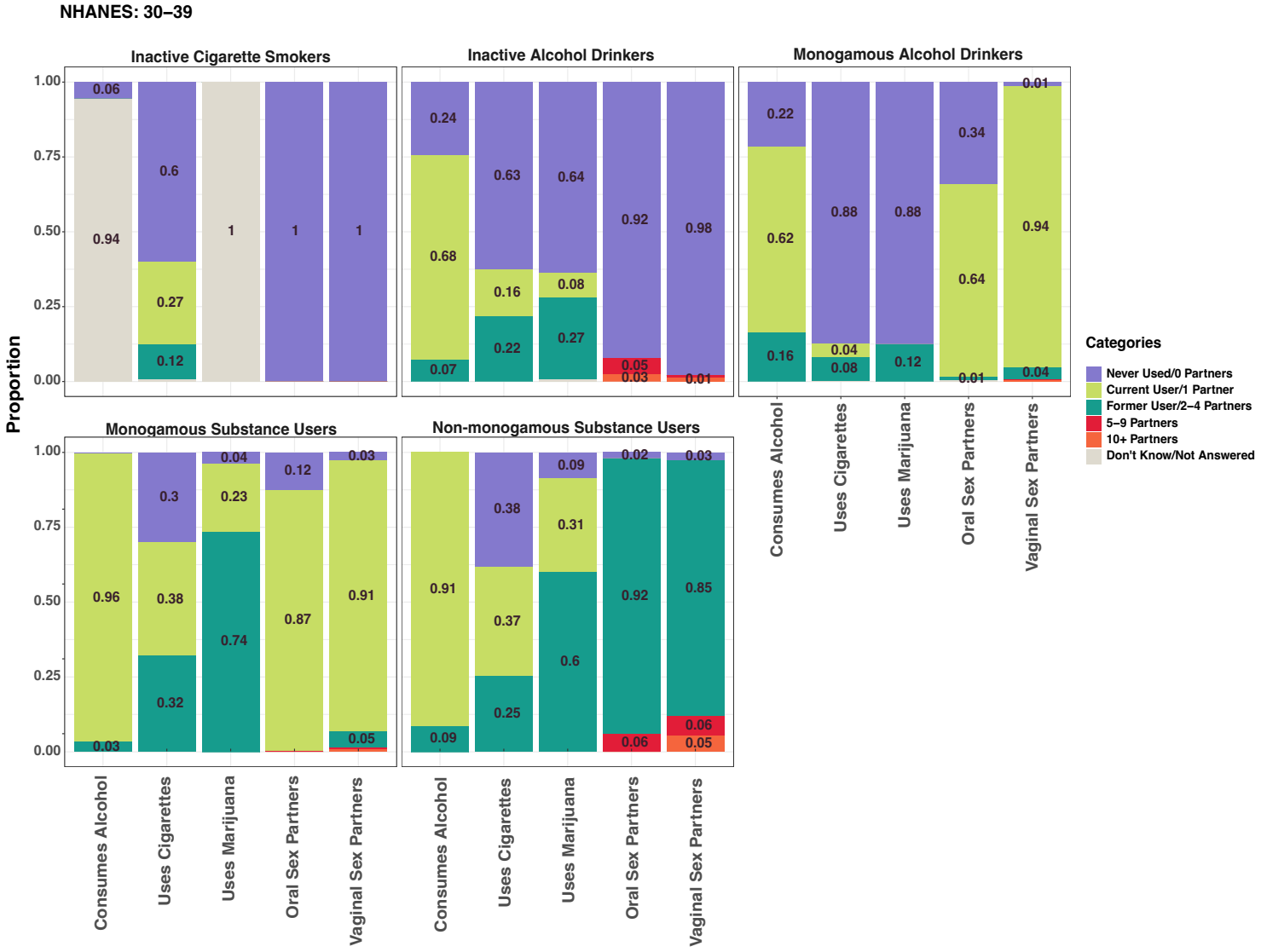


Figure 3-4 5 Class LCA Solution for 30-39 Year Olds

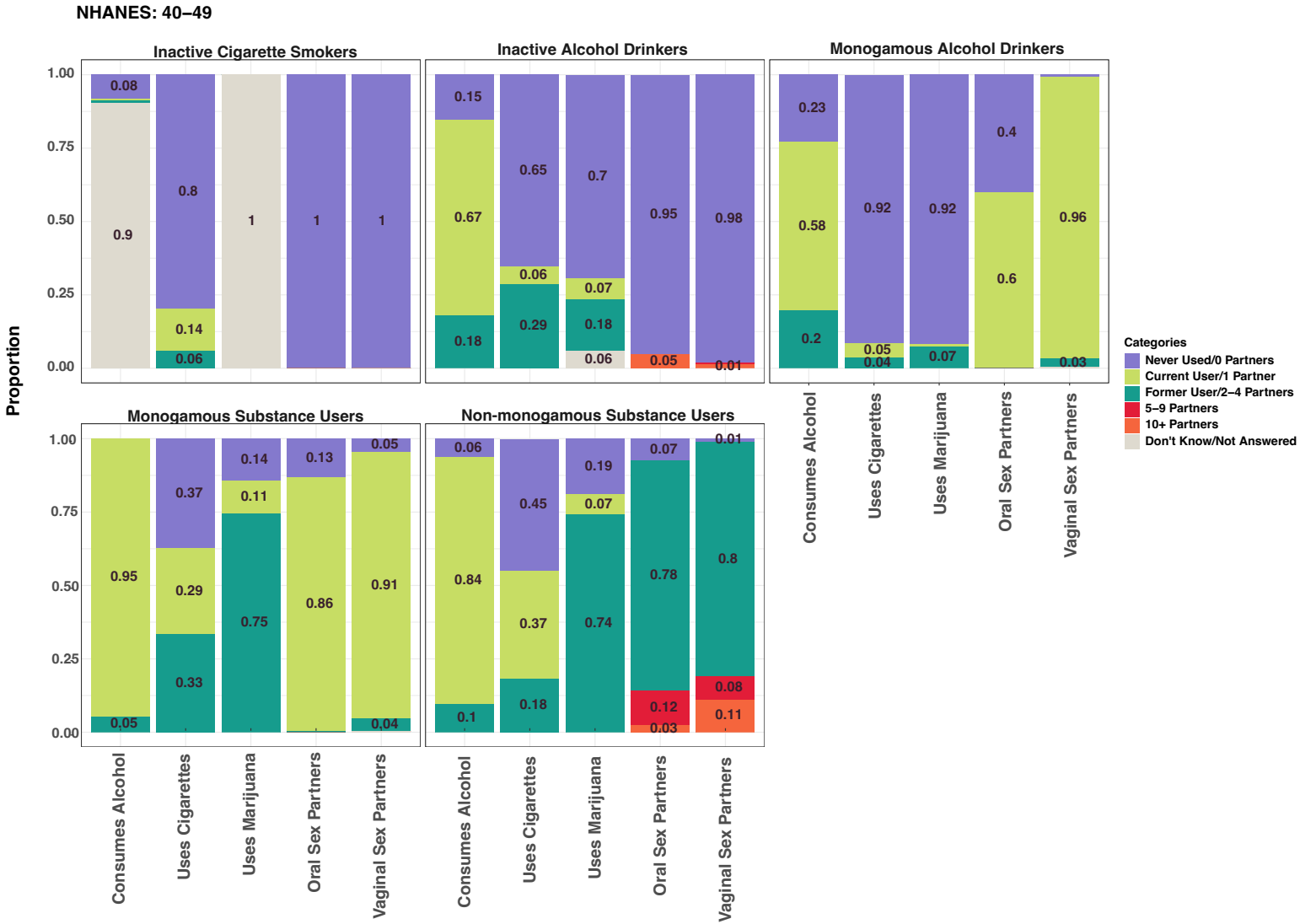


Figure 3-5 5 Class LCA Solution for 40-49 Year Olds

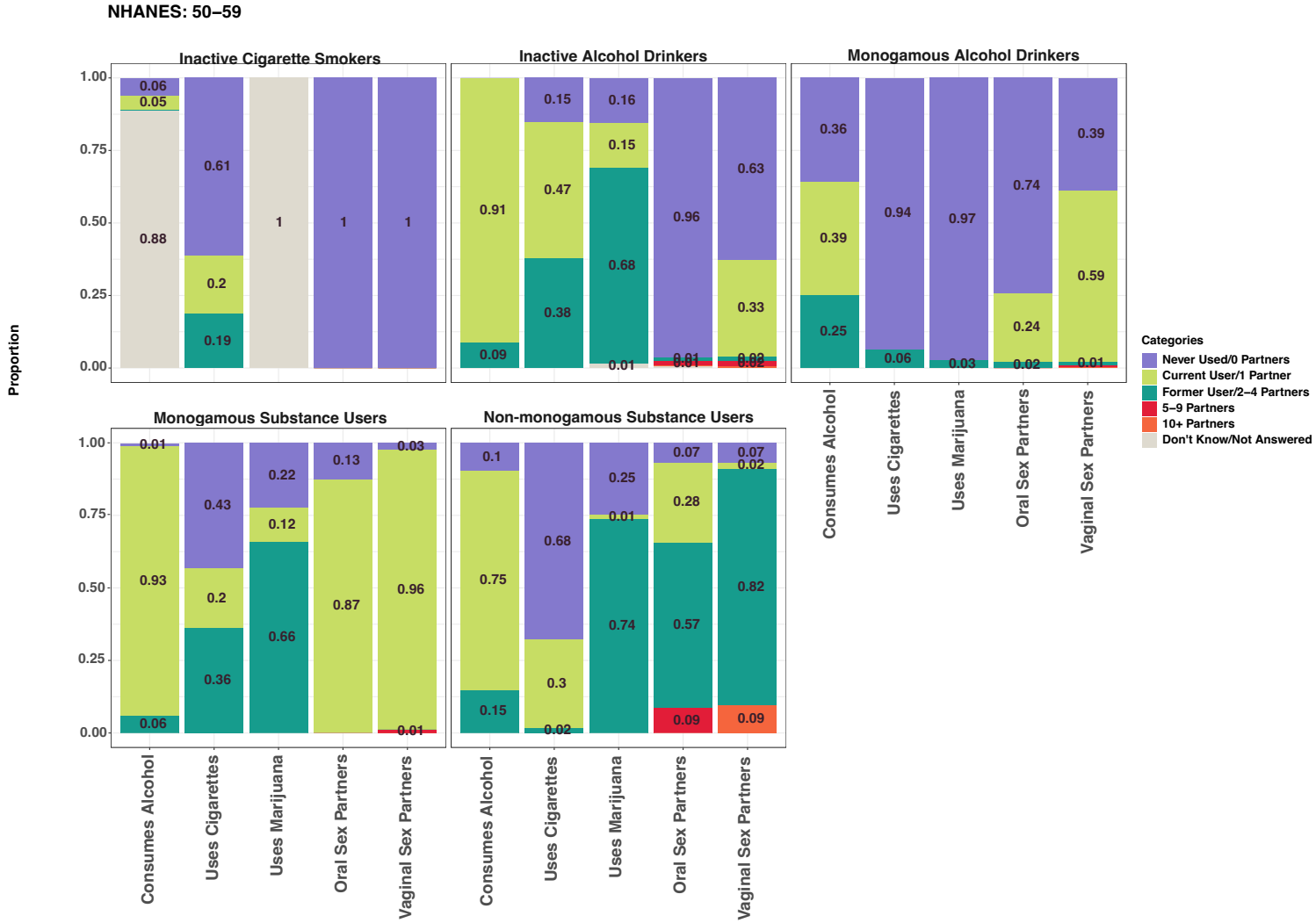


Figure 3-6 5 Class LCA Solution for 50-59 Year Old

s

3.3.2.1 Inactive Cigarette Smokers

The probabilities of endorsing sexual partnerships in the last year or endorsing marijuana consumption were largely the same across all 4 age groups. The endorsement of never, current and non-current cigarette consumption was largely similar across all age groups, and there doesn't appear to be a trend by age group. 18-29 year olds, and 50-59 year olds had a low probability of also endorsing current alcohol usage while 30-39 & 40-49 year olds did not.

3.3.2.2 Inactive Alcohol Drinkers

This class is characterized by a low probability of endorsing one or more oral or vaginal sex partnerships in the last year, and a higher probability of endorsing current alcohol consumption over any tobacco or marijuana use. If members did endorse tobacco or cannabis use it was mostly likely to be prior, non-current use. 50-59 year olds had the highest probability of endorsing non-current tobacco or marijuana use, across the 4 age groups. 18-29 and 50–59 year olds had a higher probability of endorsing at least 1 vaginal sex partnership within the last year, than 30-39 and 40-49 year old. The overall probability of endorsing one or more oral sex partnerships in the last year was similar across all age groups, with 30-39 year olds having the highest probability of 5 or more oral sex partners in the last year, and 40-49 year olds having the highest probability of endorsing 10 or more partnerships in the last year.

3.3.2.3 Monogamous Alcohol Drinkers

Similar to the prior class, this profile is defined by a greater probability of endorsing current alcohol consumption over any tobacco or marijuana use. Likewise, if members of this class did endorse any tobacco or marijuana use it is likely to be non-current use. 18-29 year olds

had the highest probability of endorsing any marijuana usage across the 4 age groups. (Figures 3.3)

The characteristic that differentiated this class from the previous one is the high probability that classes member had 1 oral or vaginal sex partnership in the last year. This probability was highest amongst 18-29 year olds, and lowest in 50-59 year olds. Interestingly, the 30-39 and 40-49 age categories had profiles that were nearly the same.

3.3.2.4 Monogamous Substance Users

This class of individuals is similar to the prior class in that the profiles indicate that members have the highest probability of endorsing 1 oral or vaginal sexual partnership within the last year. This being said, there is still a low probability that class members will endorse having 2 or more sex partners within the last year. This probability is highest in 18-29 year olds.

The profile differs from the prior group in that they have higher probabilities of endorsing tobacco or marijuana usage. Of the age groups 18-29 year olds in this class have the highest probabilities of endorsing current tobacco or marijuana use. Across the 30-39 and 40-49 year old age groups endorsement of non-current tobacco and marijuana use increases as endorsement of current marijuana and tobacco use decline. 50-59 year olds in this category have the highest probabilities of denying any tobacco or marijuana use.

3.3.2.5 Non-Monogamous Substance Users

This class has substance use profiles similar to the prior class. Current alcohol consumption is the most endorsed substance use, followed by tobacco use. Interestingly, compared to the other age groups. 50-59 year olds in this class have a higher probability of denying ever using tobacco than endorsing non-current use. 18-29 year old class members have

the highest probability of endorsing current marijuana use, while the other three classes are more likely to endorse non-current marijuana use.

This class has the highest proportion of people endorsing at least 2 sexual partnerships in the last year. Of these 18-29 year olds have the highest probability of endorsing 5 or more vaginal sex partnerships in the last year. Interestingly, 40-49 year old have the highest probability of endorsing 5 or more oral sex partnerships within the last year. 50-59 year olds in this class are unique relative to the other age groups in that they have a 0.28 probability of endorsing one oral sex partnership and a 0.02 probability of endorsing one vaginal sex partnership within the last year. (Figures 3.3-3.6)

3.3.2.6 Measurement Invariance

A multigroup analysis of the 4 decadal groups indicated that a class in one group does not measure the same construct as its corresponding class in a different group, which is to say that corresponding classes are not internally valid. In other words, the experience or state of being an 18-29 year old "Inactive Alcohol Drinker" is not the same as the experience or state of being a 30-39 year old "Inactive Alcohol Drinker". These two classes, although similar, are influenced by differing external forces, and dispose class members to potentially different risks.

This lack of exchangeability is further evidenced when a latent class analysis is run on the unstratified data (Figure 3.1). If matching classes across age groups measured the same construct, we would expect to see a five class solution in the unstratified data that was comparable to the solutions seen in the age group stratified datasets. Instead, our fit statistics found that the six class results was the best for the unstratified dataset. Further, when an individual's decadal class was mapped to their unstratified class, via Sankey diagrams, we see class crossover. (Figure 3.7) For all age strata those people who were in the lowest class in the

stratified analysis, remained in the class endorsing the lowest number of behaviors in the unstratified analysis. This potentially speaks to the continuity and robustness of this profile time across the life course. Likewise, we find that individuals who were in the highest class remained in the highest class when mapped to their unstratified class assignment. The other classes show flux, via the extent to which they cross between classes. This could be an indication that while profiles containing extreme behaviors remain constant, the profiles amongst 30-39, and 40-49 are more susceptible to the external pressures of one's environment.

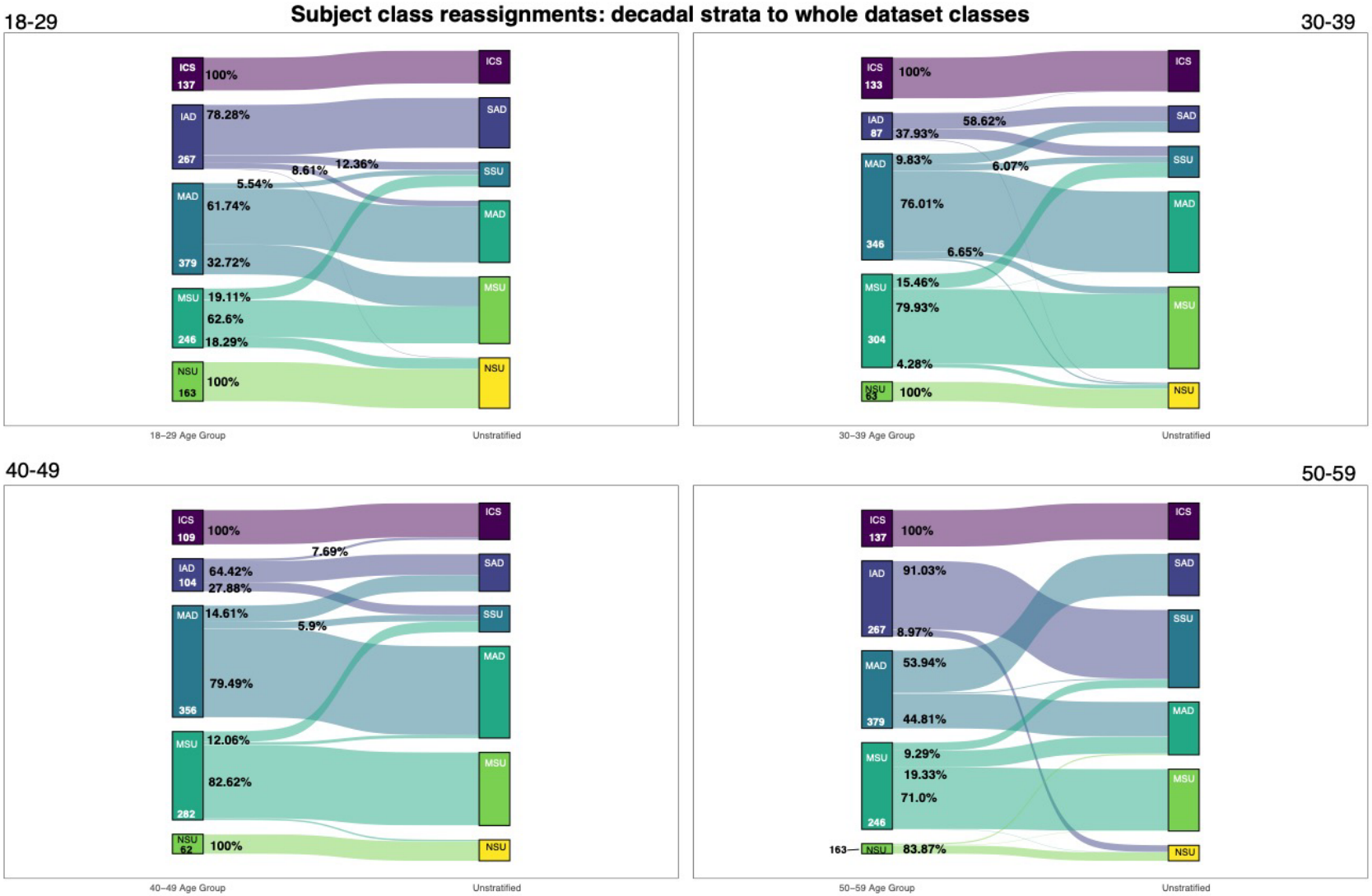


Figure 3-7 Sankey Diagram Showing Sorting of Class Assignments Between Stratified and Unstratified LC. Numbers in the columns reflect class size. Percentages refer to the proportion of individuals from the stratified class that map to the unstratified class at the end of the ribbon

.

3.3.2.7 Logistic Regression

In the unstratified dataset, the odds of infection from 'Any STI' were higher in the Semi-active Substance Users through Non-Monogamous Substance Users relative to Inactive Cigarette Smokers; however not all odds ratios were statistically significant. Monogamous Substance Users report an odds of infection that is 1.43 (95% CI 1.08, 1.90) times greater than that of Inactive Cigarette Users, while Non-monogamous Substance Users shows an 2.36 (95% CI 1.65, 3.37) times greater infection relative to Inactive Cigarette Users (Table 3.2).

Age Strata	Semi-Active Alcohol Drinkers	Semi-Active Substance Users	Monogamous Alcohol Drinkers	Monogamous Substance Users	Non-Monogamous Substance Users
18-59	0.90 (0.65,1.24)	1.02 (0.74,1.41)	1.27 (0.95,1.68)	**1.43 (1.08,1.90)**	**2.36 (1.65,3.37)**

Table 3-2 Unstratified Logistic Regression Results . Odds Ratio of "Any STI" for each class with "Inactive Cigarette Smoker" as reference Bold indicates statistically significant with alpha = 0.05, italics indicates statistically significant with alpha = 0.10

The age group stratified analyses indicate that the odds ratios comparing STI infection in Non-monogamous Substances Users to Inactive Cigarette Smokers were largest and statistically significant, across all 4 age strata (OR 2.64, 95% CI 1.39, 5.02), (OR 3.56, 95% CI 1.18, 10.77), (OR 6.82, 95% CI 1.37,33.96), and (OR 3.74, 95% CI 1.22, 11.54) respectively. In Inactive Alcohol Drinkers, the odd of "Any STI" positivity were 0.46 times as great at the reference amongst of the 18-29 age group (OR 0.46, 95% CI 0.27,0.77), while Inactive Alcohol Drinkers in the 50-59 age category had 2.55 times greater odds of STI positivity compared to the reference group OR 2.55, 95% CI 1.32,4.93). Monogamous Alcohol Drinkers were nominally protective and significant in the 30-39 age strata (OR 0.92, 95% CI 0.53, 1.59) while Monogamous Substance Users were nominally significant in both the 18-29 age group and the 40-49 age group (OR 1.60, 95% CI 0.92, 2.77) and (OR 1.83, 95% CI 0.91, 3.68).

Age Strata	Inactive Alcohol Drinkers	Monogamous Alcohol Drinkers	Monogamous Substance Users	Non-Monogamous Substance Users
18-29	**0.46 (0.27,0.77)**	0.98 (0.59,1.63)	*1.60 (0.92,2.77)*	**2.64 (1.39,5.02)**
30-39	1.81 (0.82,3.99)	*0.92 (0.53,1.59)*	**1.79 (1.01,3.18)**	**3.56 (1.18,10.77)**
40-49	1.18 (0.49,2.84)	1.44 (0.72,2.86)	*1.83 (0.91,3.68)*	**6.82 (1.37,33.96)**
50-59	**2.55 (1.32,4.93)**	1.18 (0.59,2.36)	1.60 (0.84,3.07)	**3.74 (1.22,11.45)**

Table 3-3 Stratified Logistic Regression Results. Odds Ratio of "Any STI" for each class with "Inactive Cigarette Smoker" as reference Bold indicates statistically significant with alpha = 0.05, italics indicates statistically significant with alpha = 0.10

3.4 Discussion

Sexually transmitted infections continue to be a public health problem for all age groups within the United States.[19,20,87,88] While it is understood that substance use and sexual activity are known risk factors for STI positivity, integrating the many different measures of sexual activity and substance use into a coherent set of profiles or patterns has been a challenge, and little is known about how these factors interact across generations and over the life course.[21,22,24,92,93,95] This is problematic because older individuals are growing more vulnerable to substance use disorders and sexually transmitted infections.[18,19,116,117] Identifying profiles composed of STI risk factors, and comparing their composition across age groups can help us understand the structure of the combined profile of sexual and substance use activity, and how it changes with age—enabling us to shed light on how one's risk for STIs changes with age. It also allows us identify differences caused by age and birth cohort. This study presents avenues for future investigation that may fill this literature gap, helps to fill this gap by using latent class analysis to understand how substance use and sexual partnership profiles compare between 4 decadal age groups.

Latent class analysis of each age group yielded 5 profiles, whose overall compositions remained similar between the age groups. This finding is perhaps unexpected as one might anticipate each age group (or at least some age groups) would have their own unique set of profiles given the broad changes that occur over the life course. The presence of recurring profiles may be reflective of a consistent set of underlying culture specific archetypes that have their own unique trajectories and that can describe substance use and sexual partnerships amongst Americans. Given the similarities in composition of the profiles across age groups, it is somewhat surprising that the profiles were not exchangeable under a multigroup analysis. However, this is consistent with the larger number of classes/profiles observed in the overall unstratified analysis. The non-equivalence of the classes across age groups raises the possibility that the differences in these profiles may be reflective of Age-Period-Cohort based trends, wherein while the classes appear similar across age groups, there are distinct features that vary as a function of birth cohort or period (e.g. differences in marijuana use endorsement) reflecting changing social mores.

Indeed, we see elements of the variance in trajectories and possible Age-Period-Cohort trends when we compare similar classes across age groups.[118,119] For instance, when comparing "Inactive Cigarette Smokers" and "Inactive Alcohol Drinkers" in the 18-29 strata against those in the other age strata, we see a marked increase in substance usage. When comparing "Monogamous Alcohol Drinkers" the 18-29 group to those in other age groups, we see a decline in current substance use, and an increase in non-current cigarette use between 18-29 and 30-39, followed by a leveling off in the remaining age classes. Similarly, in this class we find that current alcohol consumption and the probability of having more than 1 oral sex steadily decreases across these age groups while the probability of being a non-current drinker increases.

We find a similar decline in the probability of having 2 or more sexual partners in the "Monogamous Substance Users" group. Likewise, a peak in current use followed by a decline in substance use can be found when comparing current alcohol consumption and tobacco use profiles in the "Non-monogamous Substance Users" class. In short, although the five archetypes appear in each of the age groups, what it means to belong to one of these classes changes across age.

These results are in accordance with other findings from other studies. Using data from the *Key Substance Use and Mental Health Indicators in the United States: Results from the 2015 National Survey on Drug Use and Health* report, Lee and Sher found that substance use peaks during early adulthood, then tapers off over the remaining life span.[99,120] Likewise, Windle found similar trajectories when assessing cigarette, alcohol and marijuana use data taken from a longitudinal study that followed participants from the ages of 18 to 33.[100]

Citing research by others, Sher and Lee posit that the emergence of parenthood & marriage, or the added responsibilities of adulthood may be mechanisms behind "maturing out" of substance use. [99,116,121–124] These things too may also explain the increase in zero or singleton sexual partnerships, and relative stability in profiles observed in the 30-39 and 40-49 age strata.

Finally, some of the characteristics of the 50-59 year old profiles may be signatures of this particular birth cohort. Yang and Andrade note that due to increased exposure during childhood, those born from 1946-1964, tend to have a more lenient view of alcohol, tobacco and illicit drug use, relative to younger generations. [117] They posit that this attitude coupled with longer lifespans, partner loss, access to pain medication, and other life stressors, maybe responsible for an ongoing increase in substance abuse in elders.[18,117] This may explain the elevated number of current and prior substance user amongst the "Inactive Alcohol Drinkers".

Likewise, Schick et. al. note that women 50 and older are twice as likely as men of the same age to be single, and that 5% of people 50 and older report dating one or more person.[116] Schick et al point to partner loss due to death, divorce, or hospitalization as the motivation behind new sexual partnerships for those 50 and older.[116] These changes may explain why we find an uptick in 1 or more sexual partnerships amongst "Inactive Alcohol Drinkers" , "Monogamous Alcohol Drinkers", "Monogamous Substance Users", and "Non-monogamous Substance Users".

That our regression results revealed high odds of infection amongst the classes with the highest substance use and number of sexual partnerships is not surprising. Nor is it surprising to see monotonically increasing odds ratios across the classes as behaviors and substance use escalate. These findings corroborate the utility and veracity of using latent classes as regression parameters in situations wherein collinearity makes it challenging to study variables individually. It was unexpected to find that the odds of "Any STI" infection in 50-59 year old "Inactive Alcohol Drinkers" was statistically significant while the middle two classes were not. It is possible that this finding is related to class size and power, although it may also be a marker of better partitioning between "Inactive Alcohol Drinkers" and the other classes.

Likewise, it was also surprising to find that "Inactive Alcohol Drinkers" in the 18-29 year old age group experienced a significant, protective effect relative to Inactive cigarette smokers, because 18-29 "Inactive Alcohol Drinkers" endorse low probabilities of sexual partnerships while 18-29 "Inactive Cigarette Smokers" do not endorse any sexual behavior. This difference may be explained by reporting or social desirability bias- i.e., those in the "Inactive Cigarette Smoker" class are more inclined to deny sexual activity for personal or cultural reasons, compared to those individuals in other groups. More investigation is needed in order to explain this finding.

While this study focused on understanding sexual and substance use behavior profiles to understand the changing dynamics of STI risk over the life course, these combined sexual and substance use behavior profiles may be useful for understand a range of health outcomes—from mental health and wellbeing to alcohol-related harms, among others. [125–130] LCA and similar types of clustering analysis methods can illuminate the underlying intertwined structure of sexual and substance use behavior, which may be particularly useful for understanding mental health and wellbeing outcomes, as these may be likely to draw more heavily from the overall risk factor patterns than from any single risk factor alone.

Our study was supported by the nationally representative data in NHANES. However, we were limited by some aspects of the data collection. First, most participants were not asked or didn't answer questions about anal sex partnerships, hindering our ability to incorporate an important aspect of sexual behavior. The omission of anal sex partnerships data could result in poor class separation, yielding profiles that do not fully articulate differences between latent classes. This may not be an issue in this analysis, however. Given that sexual behaviors are strongly collinear, vaginal sex and oral sex partnerships may potentially redundantly describe many of the same latent constructs anal sex partnerships address. Further studies using a different data set (ideally one designed to understand sexual behavior) to explore this question would be warranted to further address this question.

Next, current/non-current/never alcohol and tobacco consumption were operationalized differently than current/non-current/never marijuana usage. The first two establish a threshold of use and thus filter out one-off or experimental usage. The third variable doesn't offer such screening and instead accounts for any and all instances, including single, experimental instances.

However, given the ubiquity and affordability of alcohol and tobacco products relative to marijuana, the "current/non-current/never" operationalization of marijuana consumption may be sufficient to discern risk profiles. The greater number of restrictions placed on marijuana access may serve a similar function as the "thresholds" used to assess alcohol and tobacco consumption. Thus "current/non-current/never" marijuana usage as a manifest variable maybe sufficient to separate classes.

Despite these shortcomings, we were able to demonstrate the ability of latent class analysis to describe and differentiate trends in substance use and sexual partnership behaviors across age groups. Indeed, the results of our regression echo similar findings from the literature.

The public health significance of this analysis lies in its use of a person-oriented approach to identify commonly followed patterns in sexual and substance use behavior, which can then be explored over the life course using the nationally representative data available in NHANES. These patterns/profiles can then be used to understand risk factors for sexually transmitted infections. To the best of our knowledge, this work appears to be one of the first studies to use latent class analysis to compare STI risk profiles between generations, across the life course. Previous analyses of infection risk type have largely studied the effect single variables have on infection risk (potentially in multivariate models or with interaction terms, but still largely as individual variables included in a model). Our methodology on the other hand is more holistic, taking in the sum of personal attributes to generate a risk profile, instead of isolating single variables an assessing their effect. These profiles give us a picture of what an average high-risk person may look like, and allow us to capture, quantify, and describe the concomitant nature of sexual activity variables that is typically unattainable in more classical analyses. The resulting profiles may help to provide guidance to public health and policy officials in developing targeted

screening programs and will generate insight into which risk factors or factor groups appear to the most associated with infection.

4 Characterization of Network Structure and Patient Volume at Michigan Medicine - University Hospital

4.1 Introduction

Hospital Acquired Respiratory Viral Infections (HA-RVI) account for approximately 19,000 nosocomial infections per year. Of these nosocomial infections, approximately 20% of hospital acquired pneumonia cases are due to viral infections, although this figure is thought to be an underestimate.[131,132] The transmission characteristics of viral respiratory pathogens make them adept at spreading broadly. Respiratory viruses can be transmitted through a variety of modes, however bioaerosol transmission is of particular importance. Because of their fine size, bioaerosols allow virus to remain suspended in the air for long durations thereby facilitating their movement farther through the surrounding space.[133–137] The COVID-19 pandemic has drawn attention to the transmissibility of respiratory viruses, and has heightened the use of preventative measures, particularly in healthcare spaces, where, as of 2022 masking in hospitals continues to be broadly required. Indeed, as cases increased many hospitals saw some degree of infection spread and transmission events within hospital spaces.[138,139] In spite of this, that hospital transmission has been relatively limited during the pandemic speaks to the importance of the different measures that hospital settings have implemented to prevent nosocomial spread.[140,141]

However, to understand hospital transmission of respiratory viruses, in is important to consider how the structure of hospital units and patient flow may impact transmission patterns. One strategy used in the past to account for these types of distinct, ordered interactions has been to model the relationships as a network. Indeed, contact tracing and contact networks have long

been a tool used by epidemiologists to investigate outbreaks.[142,143] Likewise, synthetic networks have proven useful at understanding the transmission of infections transmitted through close-contact, like HIV, amongst homogeneous and heterogeneous populations.[144–147]

As it stands, there is an existing body of work investigating patient movement networks that use locations rather than individuals as the nodes.[148,149] However, these analyses considered patient movement between hospitals, leaving a paucity of literature on patient movement within a hospital.[150] Since it is amenable to network analysis, we will conceptualize within-hospital patient movement as a network in which hospital units are nodes and the mean daily number of patients transferred between units constitutes the edges. When represented this way, we are able to use network methodology to describe the volume of movement between units, identify clusters of units, and describe structural properties of the network like the interconnectivity of units.

In short, it is the intention of this analysis is to understand the network structure of hospital flow patterns, and in doing so to address this literature gap. We begin by characterizing the network structure of a Michigan hospital via centralization, degree distributions and community detection. Then we investigate the relationship between unit centrality and the volume of patient movement. The information gleaned from this study will be used as the basis of future analyses simulating disease outbreaks and interventions within a healthcare setting.

4.2 Methods

4.2.1 Hospital data and patient volume measures.

Deidentified patient transfer data and unit occupancy data from the University Hospital at the Michigan Medicine Health System were obtained for January 1, 2019 through December 31, 2019 via Michigan Medicine's Data Direct service. Patient transfer data consisted of a unique encounter number, the time and date the transfer took place, and the origin and destination of the

transfer. Transfers were aggregated by their respective in-patient floors or specialty units, and then the daily mean for each transfer dyad was calculated. Admissions and discharges were not included in this analysis, as we wanted to consider the dynamics of the hospital in isolation.

Similarly, unit occupancy data from Data Direct for January 1, 2019 through December 31, 2019 was aggregated to each unit, and the mean daily occupancy was calculated. This measure was not used to construct the network, but was used in conjunction with mean daily out transfers in the statistical analyses. Due to the range of values for mean daily out transfers and mean daily occupancy, these volume measures were log transformed for our analysis.

4.2.2 Network generation and descriptive analysis.

The Igraph package from R was used to create a weighted, directed graph from the between unit transfer data. [30,151] The floors and specialty units at the University Hospital functioned as the nodes of our network. The mean daily in-node and out-node transfers were used to weight the directed edges of our graph. From here, betweenness, undirected closeness, and undirected degree centrality were obtained via Igraph functions and were able to account for edge weights. For definitions of the metrics used in this analysis please refer to Chapter 1 Section 3.

Community detection was performed on the hospital network. Seven separate weighted clustering methods were performed on the network: Louvain, Edge-Betweenness, Walktrap, Spinglass, and Eigenvalue Clustering Algorithms and integer programming. [29,152] The results with the highest modularity score are presented here. Analyses were conducted in Igraph using specific functions for each algorithm.

Igraph was also used to obtain node and graph level descriptive statistics like diameter, a measure of the longest route that gets from one node to another efficiently; transitivity, or a

measure of how tightly interconnected a network is; and degree distributions, or the degree values (number of edges each node has), represented in distribution form (Table 4.1).[30] The Chorddiag Package in R was used to render images of patient transfers between units (Figure 4.1, Table 4.1).[153]

Because mean daily unit out-transfers and mean daily occupancy each varied by orders of magnitude, these measures were log transformed for analytic purposes. Ggplot2 was used to generate scatterplots comparing log transformed mean daily unit out-transfers (a proxy for between node movement) to our three centrality measures; and log transformed mean daily occupancy to our centrality measures.[154] The Stats package was used to calculate Pearson's Correlation Coefficient between log transformed mean daily unit out-transfers (LOT) and centrality, and log transformed mean daily unit occupancy (LOC) and centrality.[114]

Unit or Floor	Abbreviation	Mean Daily Out Transfers	Mean Daily Unit Occupancy	Degree Centrality	Closeness Centrality	Betweenness Centrality	Function
Adult Emergency Services	AES	73.53	41.24	16	0.88	15.69	Adult Emergency Medicine
Burn Acute Care	BURNS	3.79	6.56	17	0.71	0	Burn Unit
Electroconvulsive Therapy	ECT	1.73	1.48	8	0.58	38.5	Electroconvulsive Therapy
Medical Short Stay Unit Maize	MSSU MAIZE	9.01	12.52	14	0.71	0.13	Treats patients expected to stay less than 48 hours
In-patient Floor 4	FL 4 MAIN	26.39	23.47	22	0.79	3.25	Neurosurgery Intensive Care/Patient Rooms
In-patient floor 4 South Wing	FL 4 SOUTH	1.72	4.54	17	0.71	0	Palliative Care/ Acute Stroke Recovery
In-patient Floor 5	FL 5	33.46	26.34	22	0.79	3.25	Surgical Intensive Care/ Patient Rooms
In-patient Floor 6	FL 6	21.04	26	24	0.88	27.71	Critical Care Medical Unit/ Patient Rooms
In-patient Floor 7	FL 7	23.48	26.03	23	0.88	16.05	Cardiac Intensive Care/Dialysis/Patient Rooms
In-patient Floor 8	FL 8	25.36	27.06	23	0.88	16.05	Adult Intermediate Care Unit/ Patient Rooms
In-patient Floor 9	FL 9	4.09	21.67	3	0.41	14	Psychiatric Clinic/Psychiatric Patient Rooms
Medical Short Stay Unit Blue	MSSU BLUE	8.74	18.79	18	0.75	0.13	Treats patients expected to stay less than 48 hours
Operating Room	OPERRM	37.2	17.29	24	0.79	14.87	Operating Rooms
Psychiatric Emergency Services	PES	3.7	2.31	2	0.54	1.5	Psychiatric Emergency Services
Radiology	RAD	9.68	2.36	24	0.79	14.87	Radiology
Surgical Short Stay Unit	SSU	7.45	8.47	11	0.68	0	Treats surgical patients expected to stay less than 48 hours

Table 4-1 Description of Hospital Units, Their Function, Patient Volume, and Centrality

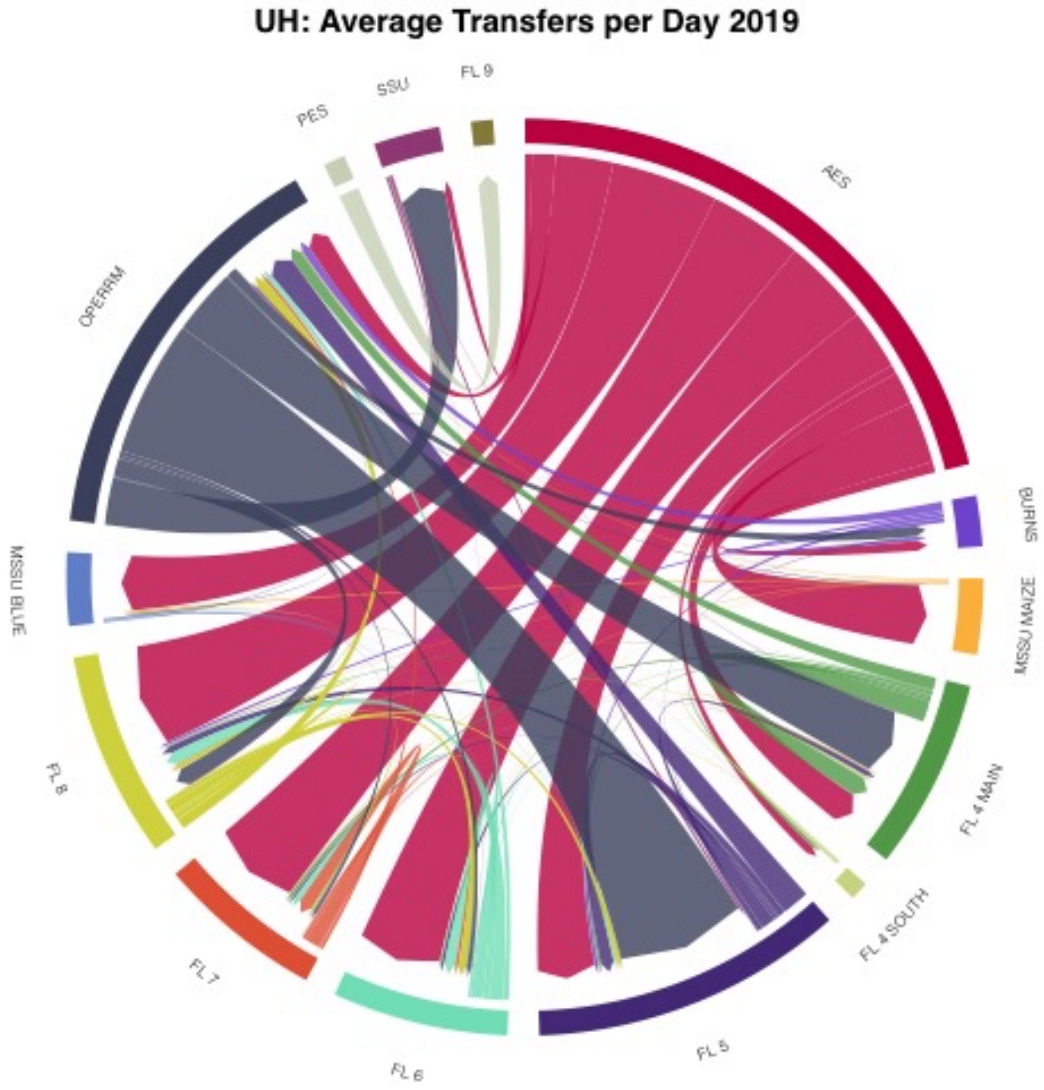


Figure 4-1 Chord Diagram Showing Mean Daily Internal Transfers

4.3 Results

4.3.1 Descriptive statistics

Centralization ranges from 0.14 to 0.32, depending on method used. These scores are considered "low" and thus indicate that the hospital isn't strongly centered around any particular unit. The unweighted and weighted diameters of the graph are 5 and 4.03 respectively. The units on the unweighted diameter path are Psychiatric Emergency Services (PES), FL 9, Electro-convulsive Therapy (ECT), FL 6, The Operating Room (OPERRM), and the Surgical Short Stay Unit (SSU). Interestingly, despite being shorter, the weighted diameter has more nodes along its path: PES, FL 9, ECT, FL 8, FL 4 SOUTH, OPERRM, Radiology (RAD), SSU. The network's mean unweighted degree is 16.75 and its mean weighted degree is 18.49. Perhaps counter to expectations given the pervasiveness of scale-free networks in the literature around human movement and contact patterns, the weighted and particularly the unweighted degree distributions do not follow a power-law trend (nonlinear on a log-log scale) and therefore indicate that the hospital is likely not a scale-free network (Figures 4.2,4.3). The weighted degree distributions were somewhat more similar to a scale free network, although for the most part they appeared visually to be more lognormal (or similar unimodal skewed distribution) than power-law distributed (although the network is likely too small to be able to fully distinguish an underlying distribution). Finally, the network's transitivity is 0.89. This metric suggests that the units within the hospital are highly connected with one another (Table 4.2).

4.3.2 Community detection

The weighted Louvain algorithm produced the highest modularity score, 0.28. 3 distinct communities were detected (Figure 4.2). Community one consisted of Burns, FL 4 Main, FL

5,OPERRM,SSU. Community two contained AES, FLs ,4 South, 6,7,8,MSSUs Blue Maize. Community three contained PES,ECT FL 9.

Unit or Floor	In Degree (Unweighted)	Out Degree (Unweighted)	Total Degree (Unweighted)	In Degree (Weighted)	Out Degree (Weighted)	Total Degree (Weighted)	Transitivity
Adult Emergency Services	3	13	16	0.08	68.4	68.48	0.78
Burn Acute Care	8	9	17	2.82	2.11	4.93	1
Electroconvulsive Therapy	4	4	8	1.21	1.21	2.42	0.5
Medical Short Stay Unit Maize	6	8	14	8.46	0.96	9.42	0.98
In-patient Floor 4	12	10	22	14.92	3.81	18.73	0.92
In-patient floor 4 South Wing	10	7	17	1.69	0.66	2.35	1
In-patient Floor 5	12	10	22	28.58	7.58	36.16	0.92
In-patient Floor 6	13	11	24	15.15	5.05	20.2	0.81
In-patient Floor 7	13	10	23	15.35	2.23	17.58	0.81
In-patient Floor 8	13	10	23	15.15	5.39	20.54	0.81
In-patient Floor 9	2	1	3	3.67	1.16	4.83	0
Medical Short Stay Unit Blue	10	8	18	8.29	0.7	8.99	0.96
Operating Room	12	12	24	12.01	35.88	47.89	0.92
Psychiatric Emergency Services	1	1	2	0	2.15	2.15	0
Radiology	12	12	24	7.73	9.12	16.85	0.92
Surgical Short Stay Unit	3	8	11	7.01	1.16	8.17	1
Average	8.38	8.38	16.75	8.88	9.22	18.11	0.77

Table 4-2 Weighted and Unweighted Degree, Node-level Transitivity (Local Clustering Coefficient)

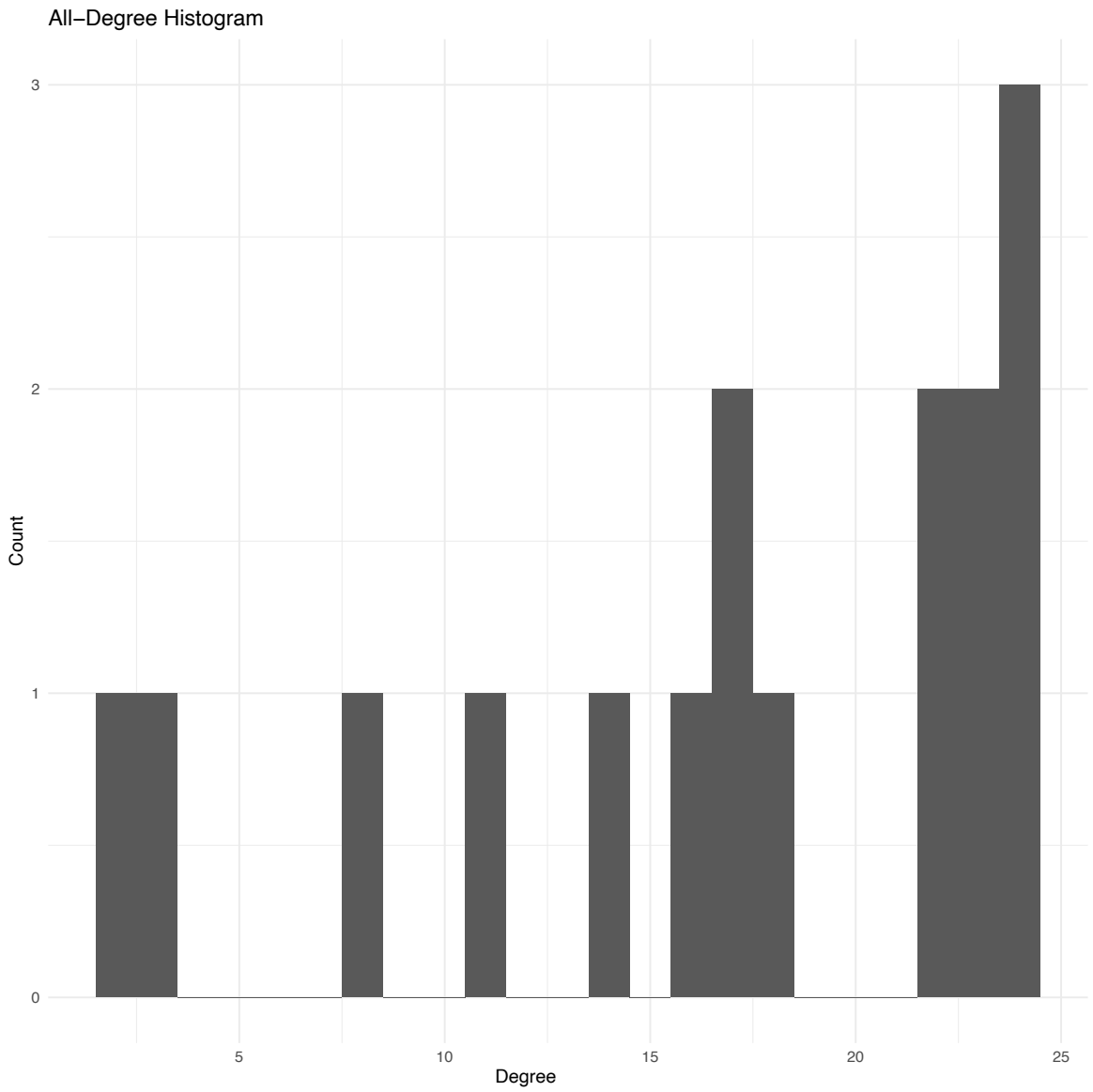


Figure 4-2 Unweighted Total Degree Histogram

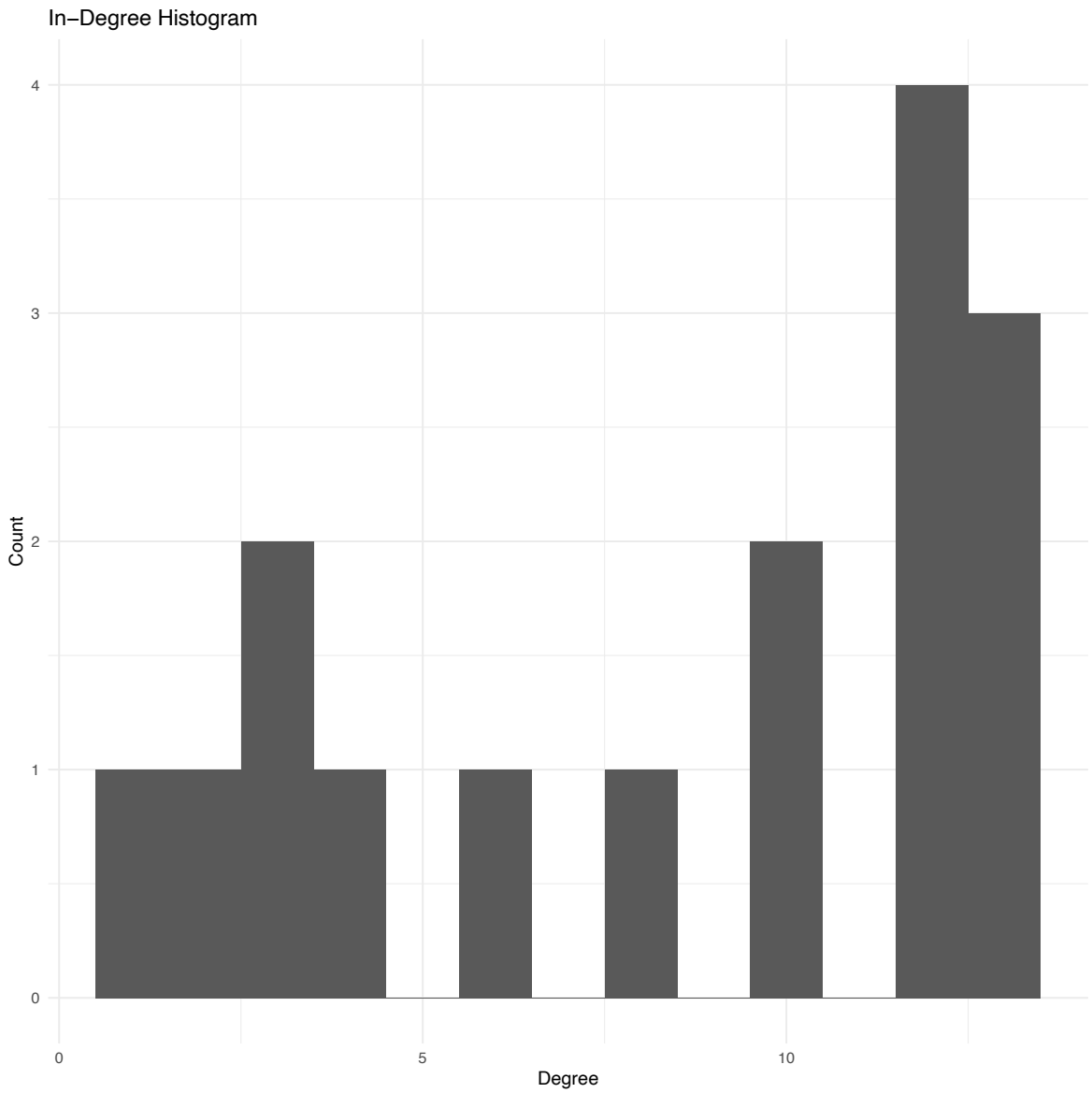


Figure 4-3 Unweighted In-Degree Histogram

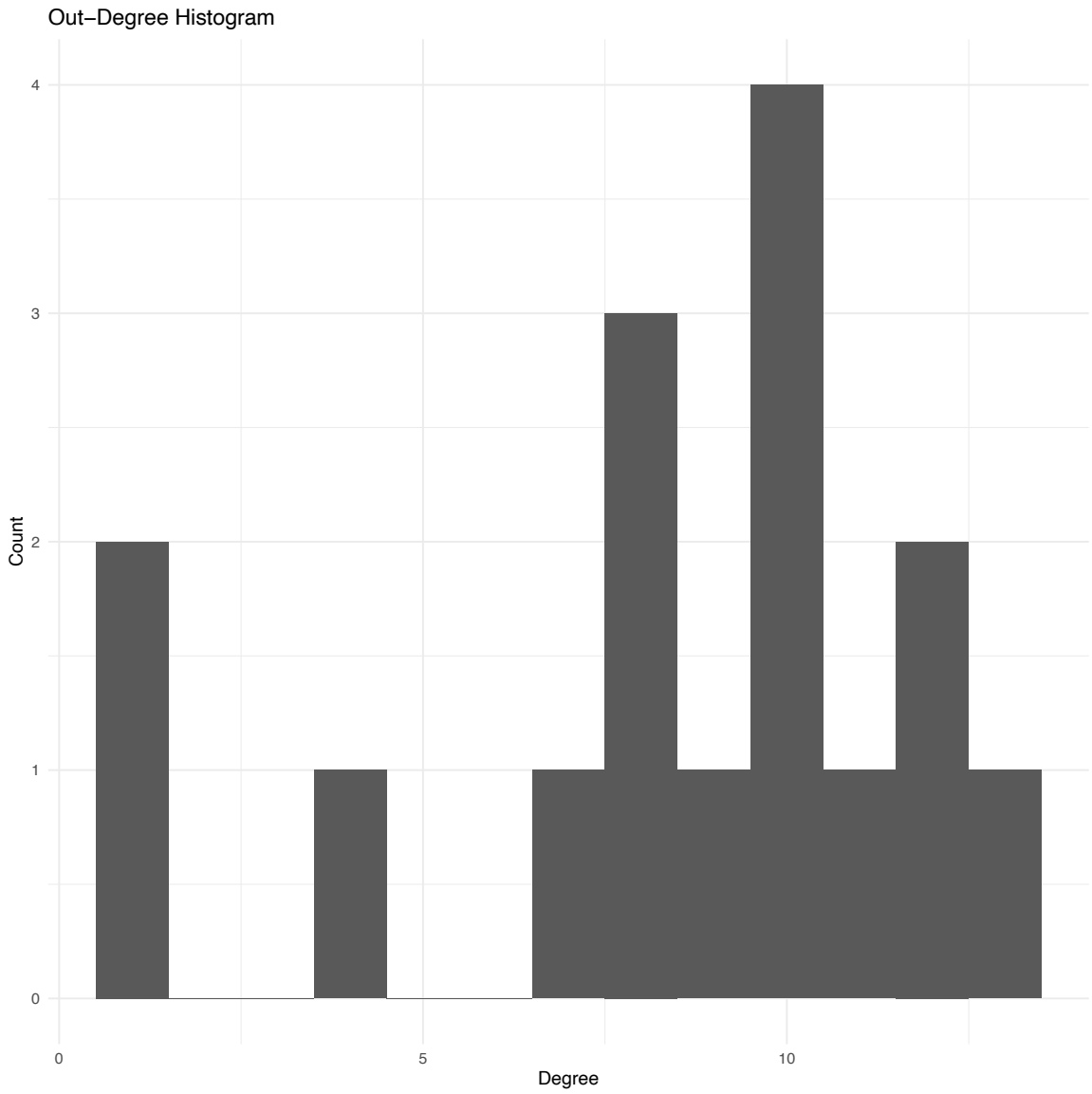


Figure 4-4 Unweighted Out-Degree Histogram

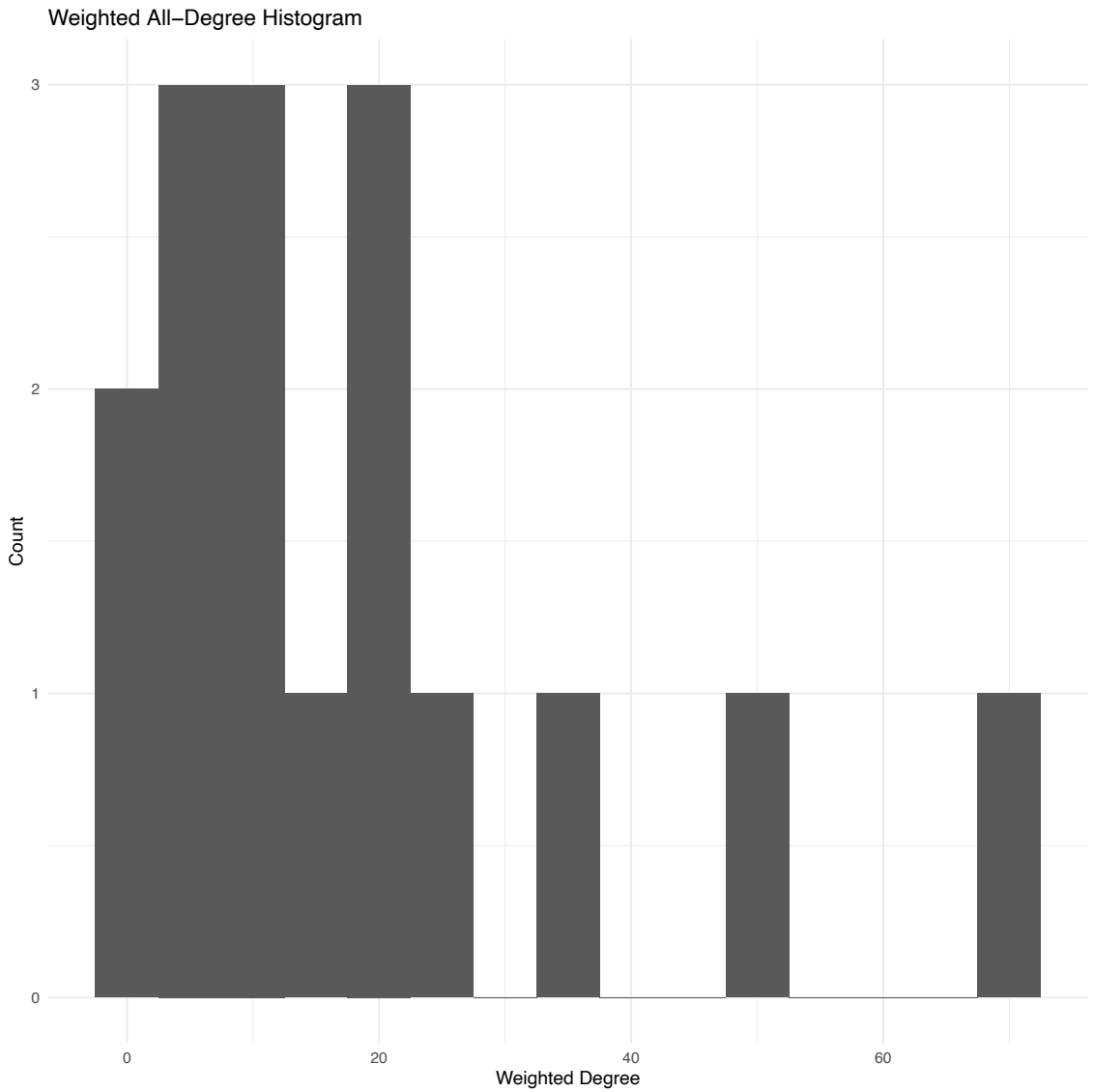


Figure 4-5 Weighted Total Degree Histogram

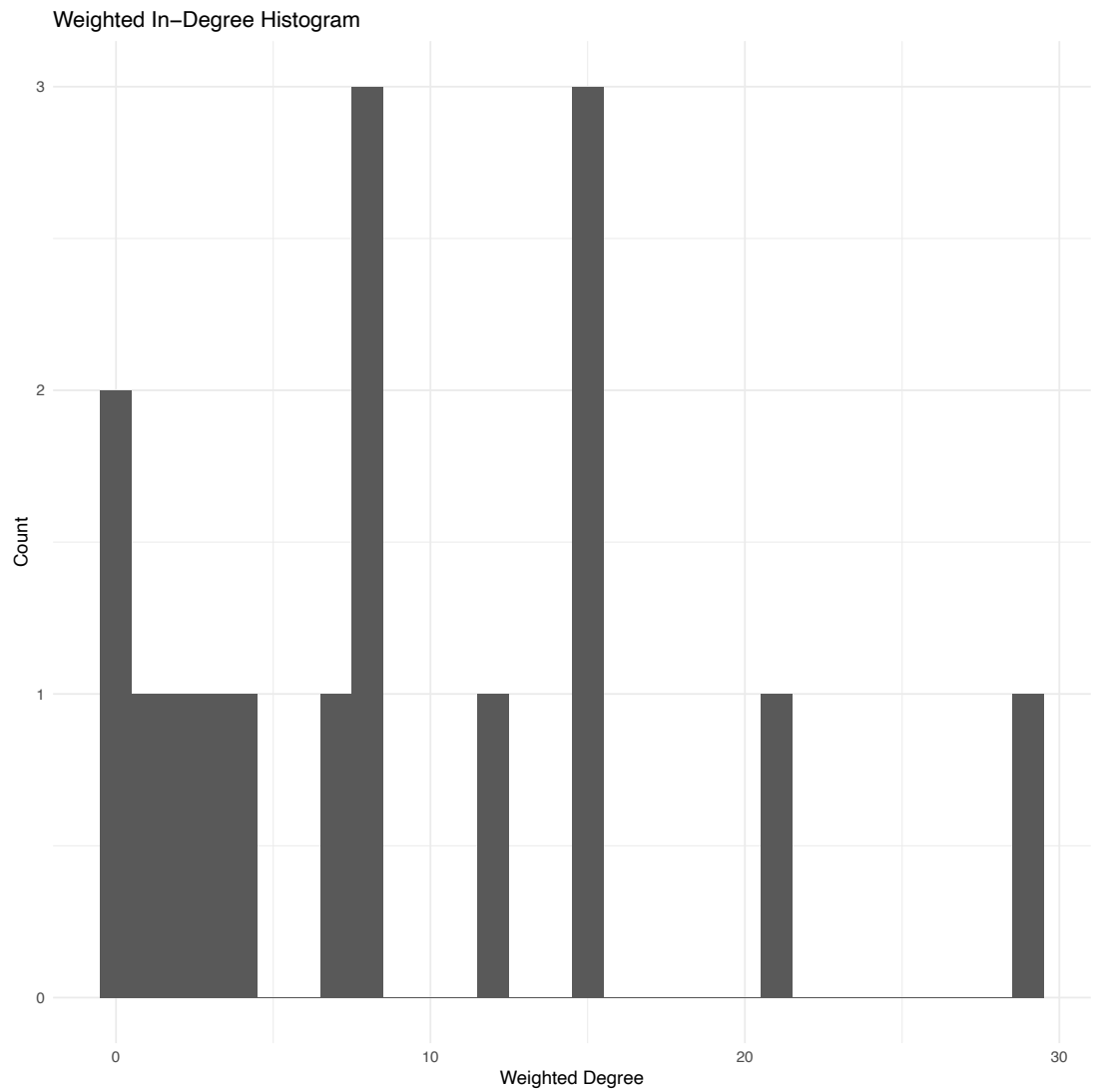


Figure 4-6 Weighted In-Degree Histogram

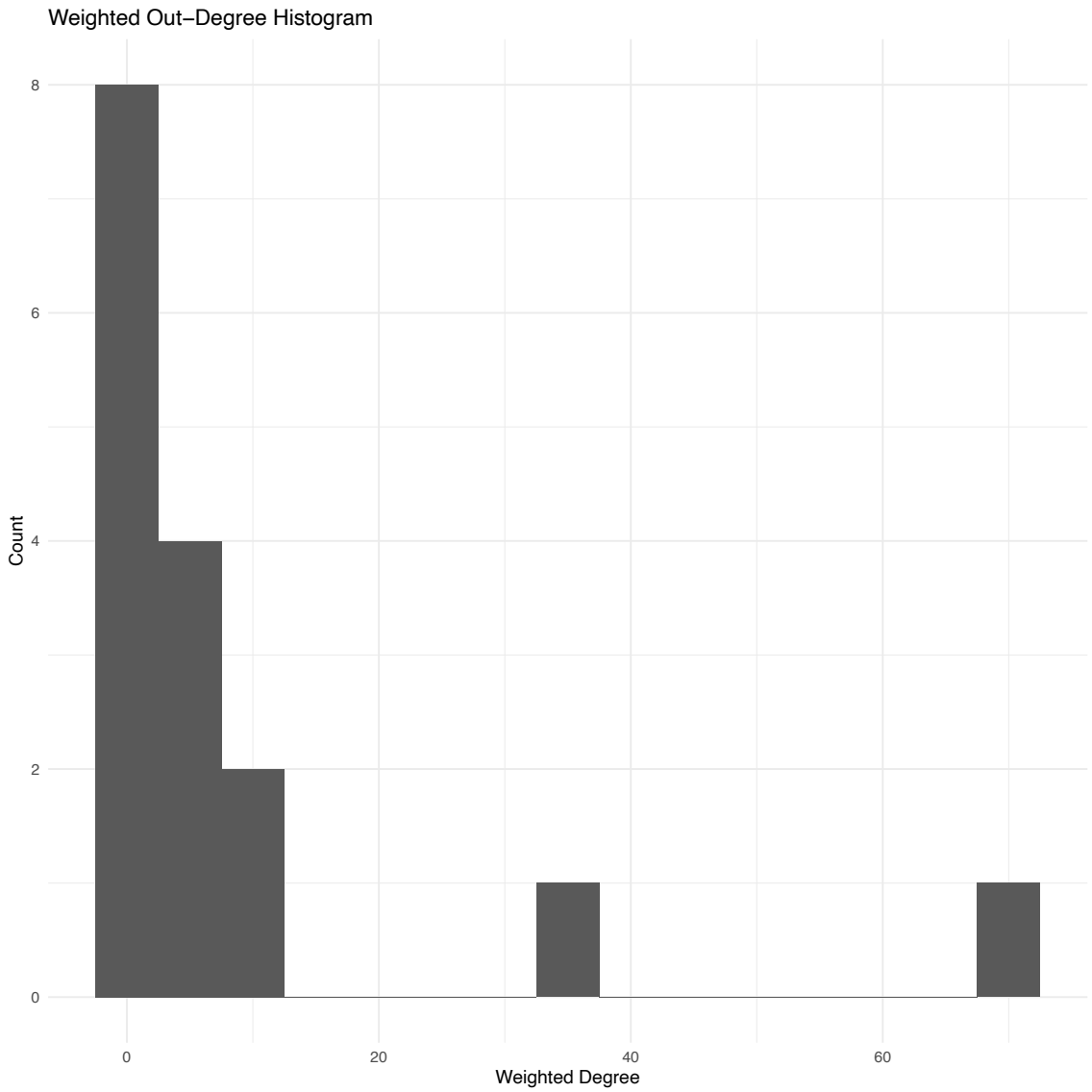


Figure 4-7 Weighted Out-Degree Histogram

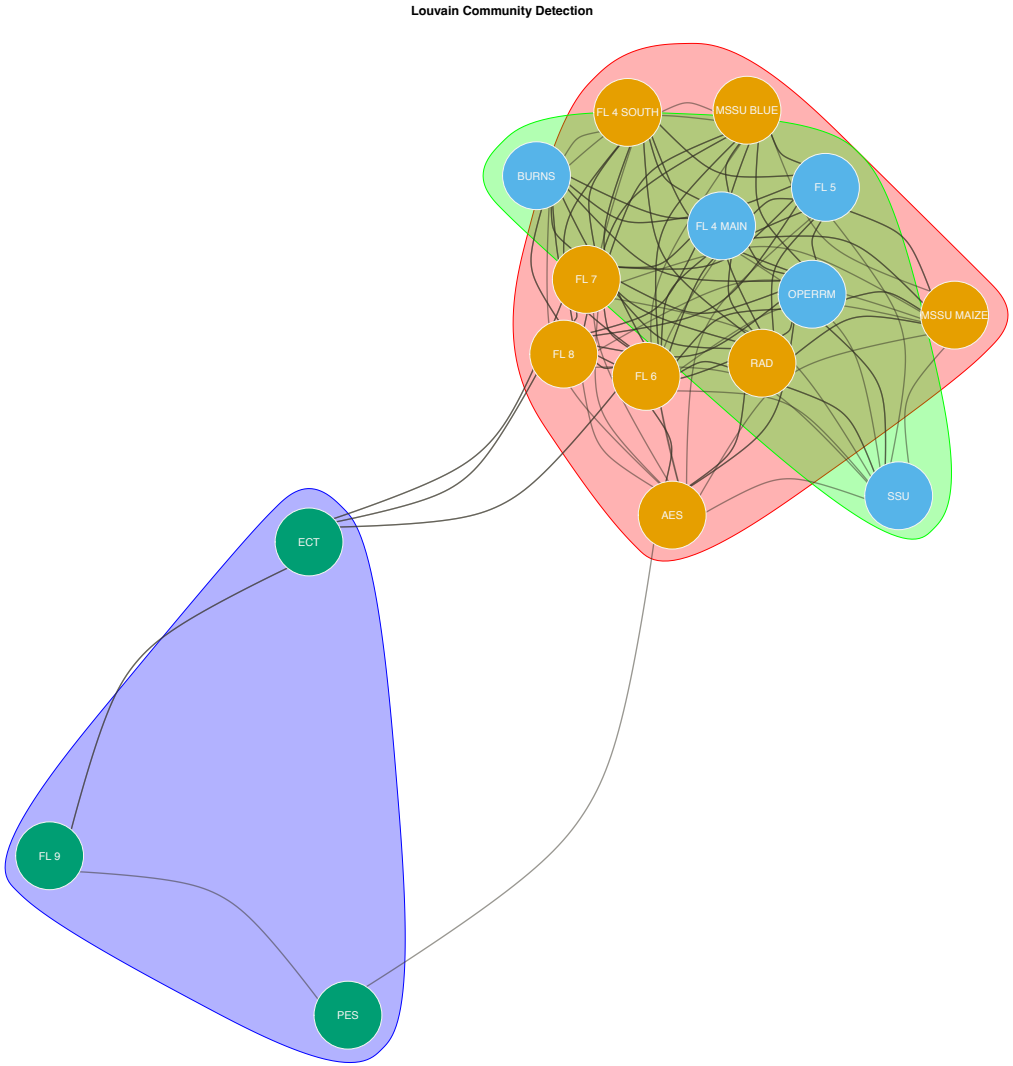


Figure 4-8 Communities Detected by Weighted Louvain Algorithm

4.3.3 Centrality

Of our three centrality measures, degree centrality and closeness centrality had the strongest associations with log transformed mean daily unit out-transfers (LOT) and log transformed mean daily unit occupancy (LOC) (Tables 4.3, Figures 4.10-11, 4.13-14). Of our two patient volume measures LOT had a stronger association with centrality compared to LOC. These four trends were positive, and with the exception of LOC and betweenness centrality, statistically significant to an alpha < 0.05. Betweenness centrality showed a u-shaped relationship with LOC, and no association with LOT (Figures 4.9, 4.12).

When plotted against each other, LOT and closeness centrality had a Pearson's correlation coefficient of 0.74 (95% CI 0.38,0.90, p-value = 0.001). We found that Adult Emergency Services (AES) and long-stay in-patient floors 6, 7, and 8 (FL6, FL7, FL8) had the highest LOT (21.04,23.48,25.36 respectively) and closeness centrality values (24,23,23. Given the liminal nature of emergency departments, this association is likely due to individuals being transferred to in-patient care floors after arriving to AES with medical emergencies. Within Floors 6,7, and 8, this association is likely due to patients visiting units like radiology (RAD) or the operating room (OPERRM) for procedures, or due to patients being transferred to step down units before being discharged.

Likewise, we found that the Electro-Convulsive Therapy Unit (ECT) and Psychiatric Emergency Services (PES) have low volumes of LOT (1.37 and 3.7 respectively), and low closeness centrality scores (0.58 and 0.54 respectively). These findings suggest that patients admitted to these units remain localized to a smaller group of units, a possible signature of the specialty care provide.

Table 4-3 Log Transformed Mean Daily Out-Transfers v Centrality

	r	95% CI	P-value
Degree	0.63	0.2,0.86	0.009
Closeness	0.74	0.38,0.90	0.001
Betweenness	0.07	-0.44,0.55	0.79

Table 4-4 Log Transformed Mean Daily Unit Occupancy v Centrality

	r	95% CI	P-value
Degree	0.43	-0.08,0.76	0.009
Closeness	0.5	0.001,0.8	0.001
Betweenness	-0.08	-0.44,0.55	0.786

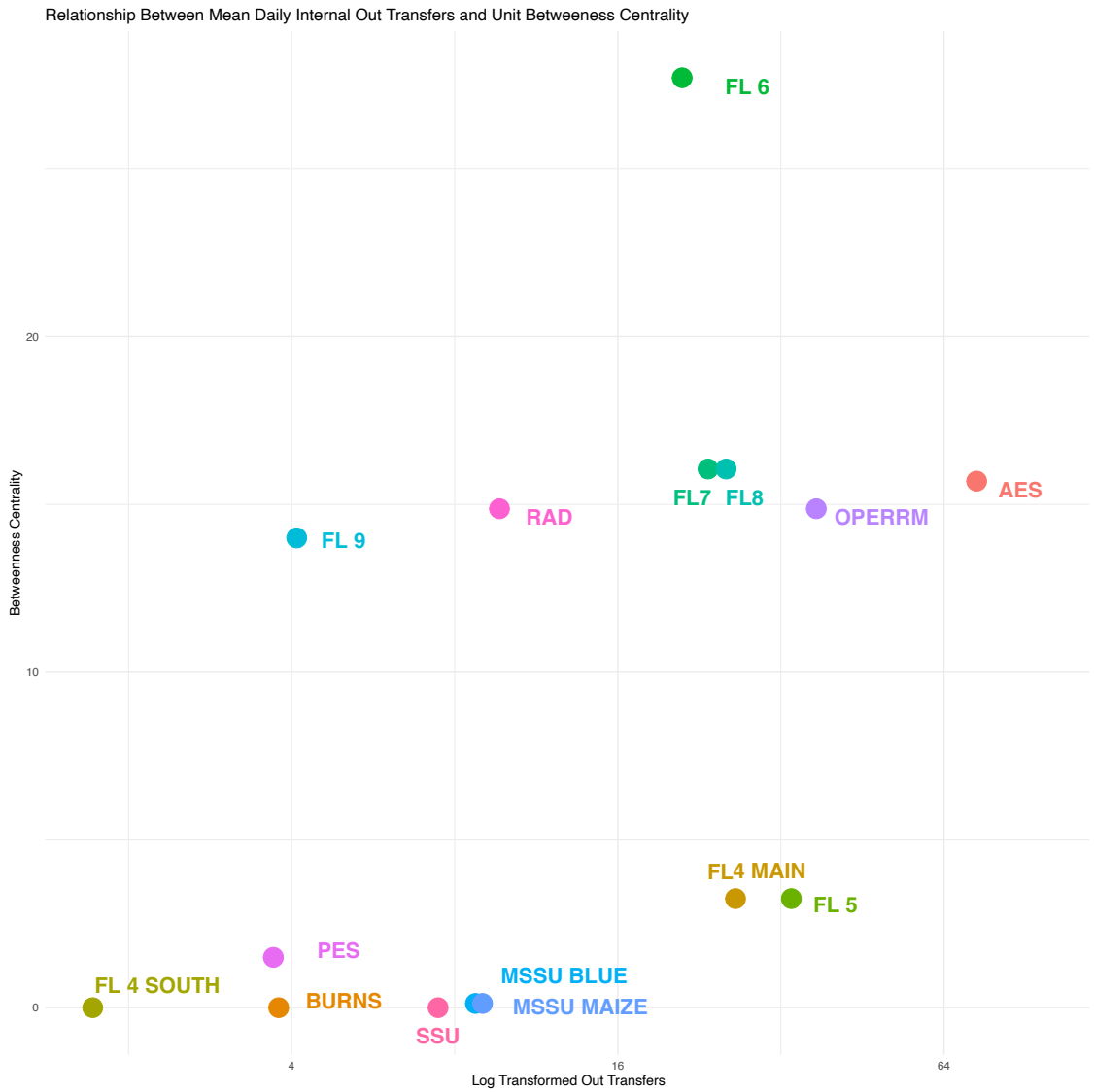


Figure 4-9 Relationship Between Mean Daily Internal Out Transfers and Unit Betweenness Centrality

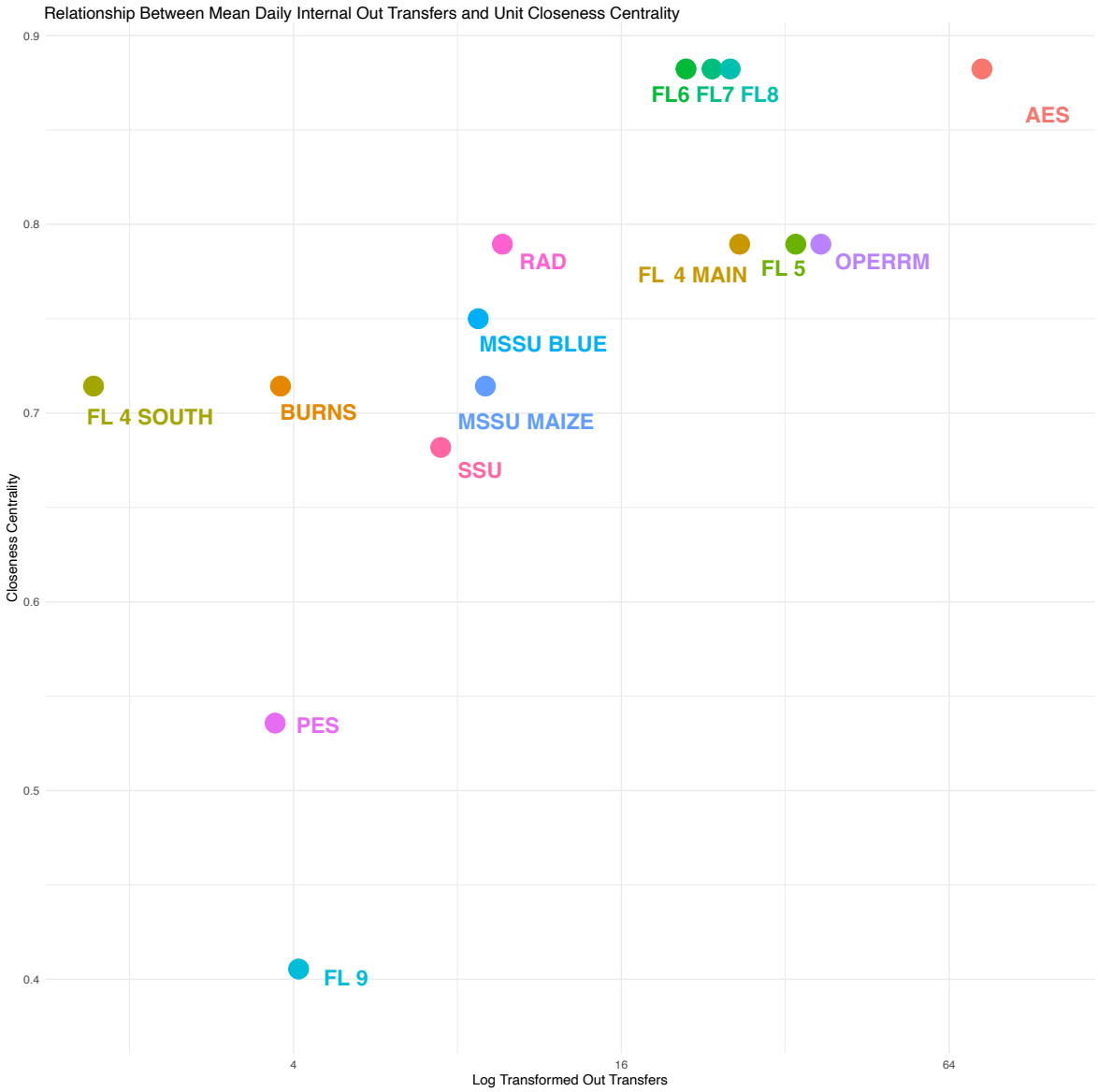


Figure 4-10 Relationship Between Mean Daily Internal Out Transfers and Unit Closeness Centrality

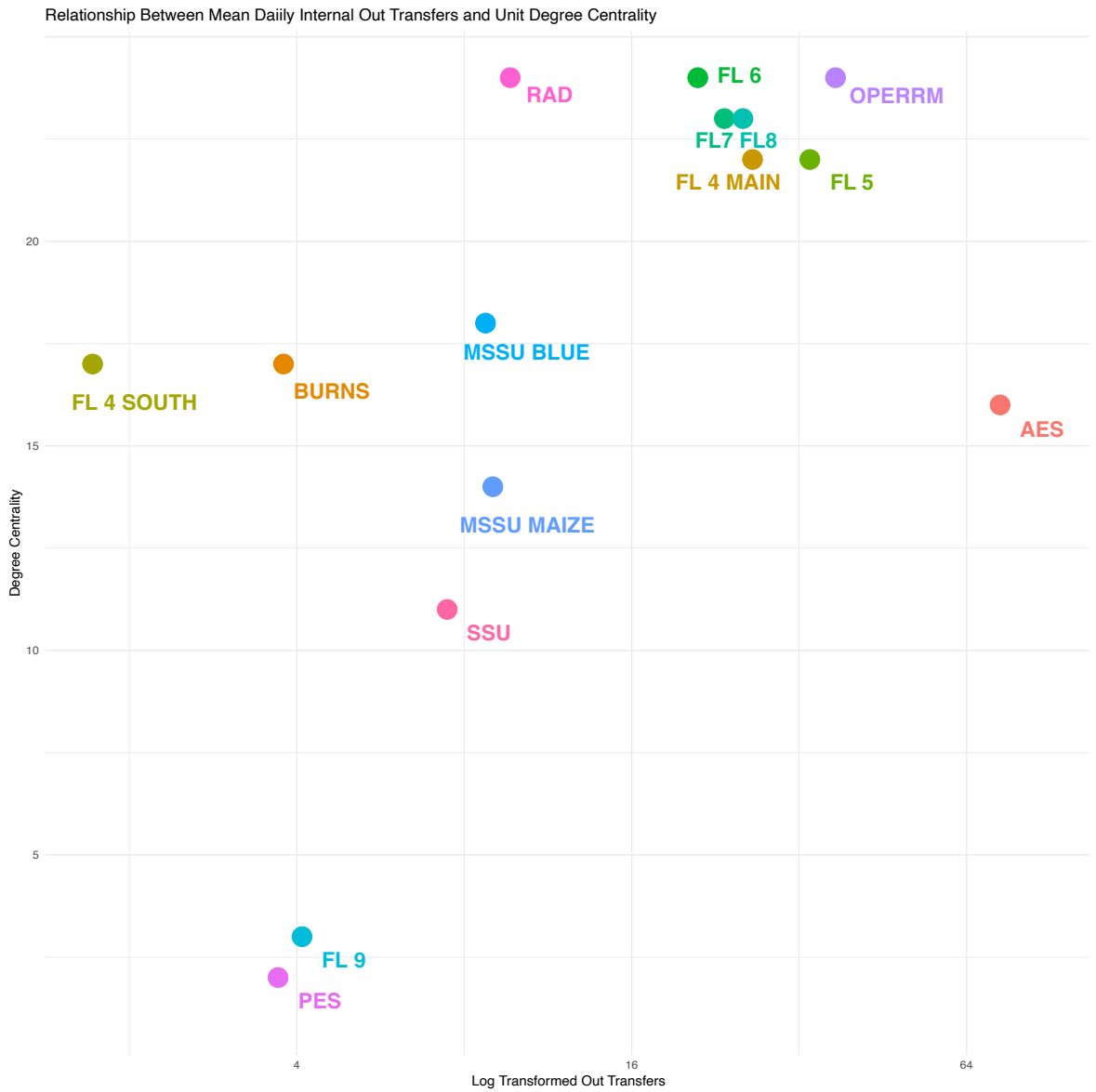


Figure 4-11 Relationship Between Mean Daily Internal Out Transfers and Unit Degree Centrality

Figure 4-12 Relationship Between Mean Daily Occupancy and Unit Betweenness Centrality

Figure 4-13 Relationship Between Mean Daily Occupancy and Unit Closeness Centrality

Figure 4-14 Relationship Between Mean Daily Occupancy and Unit Degree Centrality

Short term in-patient units like Floor 1 (FL 1), Short Stay Unit (SSU) and Medical Short Stay Unit- Blue (MSSU Blue) appear to fall in the middle, having lower closeness centralities, and between unit movement than the long-term in-patient floors, but more than ECT and PES.

In-patient Floor 9 (FL9) is a possible outlier, with a very low closeness centrality score (0.41) and a lower volume of daily out-patient transfers (4.09). As with PES and ECT, this indicates that FL 9 may have a very low volume of between unit transfers.

The relationship between LOT and degree centrality were similar to that of LOT and closeness centrality, and had a Pearson's correlation coefficient of 0.63 (95% CI 0.20,0.86, p-value = 0.009). In-patient Floors 6,7, and 8 had high degree centrality (24,23,23 respectively) and a high volume of out-transfers as described previously. ECT had the low degree centrality (8) and the lowest volume of transfers (1.73 per day), and the short stay units fell somewhere in between. These results reiterated the finding that units with higher patient volume tend to share the same patient base.

When plotted against each other LOC and closeness centrality had a Pearson's correlation coefficient of 0.5 (95% CI 0.01, 0.80, p-value = 0.048). We found that a high number of occupancies from in-patient floors 6,7, and 8 and Adult Emergency Services were positively associated with high closeness centrality. This may suggest that units with high closeness centrality may share the same patients. Similarly, ECT and PES had smaller patient populations and low closeness centrality. As with the prior results, this could be due to the highly specialized treatment they render.

We found two outliers, FL 9, which had a mean daily occupancy of 21.67 but only a closeness centrality score of 0.41 and Radiology (RAD), which had lower mean daily occupancy

2.36, but a higher closeness centrality of 0.79. FL 9's statistics suggest a possible specialty unit that sees a high volume of very specific patients. Conversely radiology's statistics suggest that this unit sees fewer patients but has stronger connections with a wider array of units.

LOC and degree centrality had a Pearson's correlation coefficient of 0.43 (95% CI -0.08,0.76, p-value = 0.095). As with LOC and closeness centrality, most in-patient floors had a greater number of direct connections to other units (0.88 for all three floors), while ECT and PES have fewer direct connections to other units. A unit with high LOC with high degree centrality suggests that this unit shares its patients with other units, while a unit with low LOC and low degree centrality could indicate a unit that treats a specific or rare condition.

Additionally, RAD and FL9 remain outliers. FL9's relative position remained unchanged, but RAD's relative position increased in tandem with OPERRM. From these graphs alone we cannot conclude if a relationship exists between OPERRM and RAD. We can conclude, however, that RAD also sees a lower volume of people that are from a broader range of units.

4.4 Discussion

Although there are studies analyzing between-hospital health care system networks, there is a paucity of research on within-hospitals networks of wards and units. Research filling this gap is important because network analysis gives us tools to analyze patient volume and movement, and to identify the locations that render hospital staff and patients most vulnerable to respiratory illness transmission. The objectives of this analysis were to describe the network structure of University Hospital at the University of Michigan and to characterize the relationship between unit centrality and patient volume, with the greater goal of understand how a respiratory infection may travel throughout a hospital.

Our descriptive statistics reveal a non-scale free decentralized yet interconnected structure. Any given node shares a directed edge with approximately half of the remaining nodes, and the longest unweighted path consisted of 5 or 31% of the existing nodes. Despite the high transitivity, the Louvain community detection algorithm yielded a modularity score 0.28 and was able to identify 3 distinct communities on the weighted network. This finding indicates the presence of distinct unit groups on the network and potentially indicates the importance of edge direction and weight on the structure of this graph. Our network is decentralized with high interconnectivity and thus unlikely to follow a power law distribution characteristic of scale free graphs. Indeed, it is commonly assumed that real-world networks are scale free, however, Broido and Clauset have argued that scale free networks are rarer than expected, and thus this finding is unremarkable.[155]

We considered three centrality measures: betweenness, undirected closeness, and undirected degree centrality; and two patient volume measures: the log mean daily transfers out of a unit, and the log mean daily unit occupancy. We assessed the strength of these relationships via Pearson's correlation coefficient. Of our centrality measures, undirected closeness and undirected degree centrality had the strongest correlation with patient volume, and when taken together may provide insight into the function or behavior of a particular unit.

There were commonalities between the closeness and degree centrality plots for both log mean daily out-transfers and log mean daily occupancy. In-Patient Floors 6,7, and 8 had among the highest closeness and degree centrality, as well as high log transformed mean daily out-transfers and mean daily occupancy. Electroconvulsive Therapy, on the other hand, had the among the lowest of all four. A combination of high closeness and degree centrality suggests that the destinations patients from FL6, FL7, FL8 travel to overlap with each other, while a high-

volume of patient transfers and occupancy speaks to the probability of patients from these units crossing paths.

Likewise, low closeness, degree centrality, patient transfers and unit occupancy, as seen in ECT, suggest that nodes of this configuration remain largely self-contained. This configuration may be characteristic of a highly specialized unit like ECT, although the inverse cannot be said about high centrality high patient volume units.

Further the discordance marked by the change in relative position of AES, PES, RAD and OPERRM, between the closeness centrality and degree centrality plots can tell us about the character of these unit. The drop in relative degree centrality compared to relative closeness centrality suggests that AES and PES have fewer direct connections with other units. Similarly, the increase in relative degree centrality versus closeness centrality implies that these units are connected with a wider array of units.

AES and OPERRM maintain a high volume of transfers and occupancies which could suggest that patients from these units are more likely to move throughout the hospital and increase the risk that a pathogen spreads. Conversely, the lower volume of transfers and occupancies seen in PES suggests that this unit is at lower risk of being a source of an outbreak.

This analysis had shortcomings. First, we were not able to track patients' entire movement during their admission. This inhibited us identifying common paths, and high traffic loops that could serve as infection hotspots. Next, this analysis features data from only one hospital. We cannot make claims about the universality of our findings, and more investigation is needed. Finally, we did not model the movement of health care workers. Despite these limitations, these analyses provided contributions to the literature: It is, to our knowledge, one of the first to use network analysis on units within a hospital.

This analysis has laid the foundation for future outbreak simulations and containment strategy testing at the University Hospital which will prove essential for identifying locations where risks of infection are highest. These analyses may also be beneficial during the planning, construction, or reorganization of a hospital building. They also offer insight into the network characteristics of a large hospital that may prove useful to others desiring to conduct similar analyses.

We found that University Hospital is largely decentralized and interconnected (transitive), and has degree distributions that do not follow a power-law distribution and are not scale free in nature. Additionally, we found a strong positive association between closeness centrality and patient movement, a potential indication that patient volume and centrality measures speak to a unit's function and the diversity of patients it sees. This positive correlation also speaks to increased potential for infection risk for a unit, as the network thus contains units which have both high numbers of patients who could potentially transmit/become infected (occupancy) and high levels of connection within the network (centrality). While this analysis provides one example of a hospital network structure, more work is needed to understand how (if at all), the network structure of hospitals varies.

5 Simulation of Respiratory Virus Outbreak on the University Hospital Network

5.1 Introduction

Hospital Acquired Respiratory Viral Infections (HA-RVI) account for approximately 19,000 nosocomial infections per year.[1,2] HA-RVI can be transmitted through multiple modes, including bioaerosols, fomites and direct contact. It is estimated that HA-RVI is responsible for approximately 20% hospital acquired pneumonia, although this may be an underestimate.[1,2]

The COVID-19 pandemic has reified existing vulnerabilities to HA-RVI.[138,156,157] At the beginning of the pandemic, hospitals had to rapidly decide on how to room patients, assign infection control precautions, and ration personal protective equipment, while still grappling with large uncertainty in the scientific literature about the transmission pathways and needed protective procedures for SARS-CoV-2 (e.g. the likelihood of aerosol vs. droplet transmission, potential for fomite transmission). To assist in decision-making in the presence of vulnerability and uncertainty in the future, identifying the units most susceptible or integral to the spread of infection is necessary to inform decision making.

Compartmental transmission models are a commonly used approach to evaluate potential for transmission and alternative containment and prevention strategies, in part because they are effective at recording the changes at the population level as individuals pass through the phases of an illness.[34] However, given the discrete units and separated populations within a hospital, a single compartmental model (such as one Susceptible-Infectious-Recovered or SIR model for the entire hospital population) may not be ideal for simulating transmission, as it assumes equal and random mixing of individuals.

One way to improve the accuracy of the compartmental model then, is to account for the distinct unit structure and patient flow within the hospital. This can be achieved by treating a hospital as a network of wards or units with distinct, homogenous populations in which random mixing does occur. Doing so allows one to capture the isolation between unit populations and the movement of patients between units, while still accounting for disease transmission at the unit level.

In this study, we simulated and tracked the trajectory of an HA-RVI throughout a hospital by conceptualizing the hospital as a network of units interconnected by patient transfers. We then nested a compartmental transmission model within each node of the network to simulate HA-RVI transmission and connected these nodes with rates based on the patient transfer data. Using this model, our goal is to ascertain each unit's burden of infection of a range of model parameters, and to then to ascertain characteristics that increase a unit's risk of a disease outbreak.

5.2 Methods

5.2.1 Conceptual Model

To simulate infection transmission between units within a hospital, we constructed a nested model. The first level represented the hospital structure rendered as a directed network. Individual units or hospital floors were treated as nodes, and the inverse of the mean time a patient spent in a unit before being transferred to a second unit, was used to create the edges of the network. This level allowed us to simulate the contained environments of hospital units, and capture the discrete between-unit patient movement.

The second level consisted of separate Susceptible-Exposed-Infected-Recovered compartmental models for each unit, each structured as:

$$dS/dt = -\beta SI$$
$$dE/dt = \beta SI - \delta E$$
$$dI/dt = \delta E - \gamma R$$
$$dR/dt = \gamma R$$

These models allowed us to simulate within-unit infection transmission, wherein it was assumed that mixing was random and homogenous. The overall final differential equations combined the above base SEIR model with linear movement rates representing patient transfers, where the model moved patients between units using the rates described above, assuming no difference in movement rate between the S, E, I, and R classes (perhaps not realistic for COVID-19 currently, but we assumed the simulated outbreak was early on before infection control procedures were in place).

5.2.2 Model Parameters: Between Unit Movement Rates

The between unit patient movement rates were derived from deidentified, patient occupancy and transfer data. These data were collected from January 1, 2019 through December 31, 2019 for University Hospital, via Michigan Medicine's Data Direct service. For these analyses patient movement was defined as one over the median time in hours a patient spent in a unit before being transferred to a separate unit.

To obtain these values, patient transfer data was first aggregated into directed, "Start"-"End" unit dyads. Then, the median occupancy duration for the patients in the "Start" node of the dyad was calculated. The rate was obtained by finding the reciprocal of this value. Admissions and discharges were denoted in the patient movement data as originating from or terminating "outside". [TABLE1]

Start	End	Rate (hr^{-1})	Start	End	Rate (hr^{-1})	Start	End	Rate (hr^{-1})	Start	End	Rate (hr^{-1})
aes	burns	0.2464	fl4main	burns	0.1705	fl7	fl4south	0.0048	operrm	fl4south	0.1754
aes	fl4main	0.0988	fl4main	fl4south	0.0257	fl7	burns	0.0080	operrm	aes	0.4082
aes	fl4south	0.1519	fl4main	fl5	0.0320	fl7	ect	0.0160	operrm	burns	0.2542
aes	fl5	0.1011	fl4main	fl6	0.0175	fl7	fl4main	0.0143	operrm	fl4main	0.1085
aes	fl6	0.0656	fl4main	fl7	0.0185	fl7	fl5	0.0215	operrm	fl5	0.1185
aes	fl7	0.0925	fl4main	fl8	0.0204	fl7	fl6	0.0198	operrm	fl6	0.2346
aes	fl8	0.0723	fl4main	mssublue	0.0102	fl7	fl8	0.0152	operrm	fl7	0.2400
aes	mssublue	0.1268	fl4main	mssumaize	0.4839	fl7	mssublue	0.0047	operrm	fl8	0.1576
aes	mssumaize	0.1407	fl4main	operrm	0.0220	fl7	operrm	0.0144	operrm	mssublue	0.1796
aes	operrm	0.1441	fl4main	out	0.0172	fl7	out	0.0147	operrm	mssumaize	0.2260
aes	out	0.0525	fl4main	rad	0.0121	fl7	rad	0.0112	operrm	out	0.1667
aes	pes	0.0861	fl4south	fl4main	0.0157	fl8	fl4south	0.0164	operrm	rad	1.3187
aes	rad	0.0921	fl4south	fl5	0.0201	fl8	burns	0.0351	operrm	ssu	0.1307
aes	ssu	0.0968	fl4south	fl6	0.0267	fl8	ect	0.0151	pes	fl9	0.1188
burns	fl4main	0.0218	fl4south	fl7	0.0130	fl8	fl4main	0.0330	pes	out	0.1345
burns	fl4south	0.0200	fl4south	fl8	0.0222	fl8	fl5	0.0236	rad	fl4south	0.2752
burns	fl5	0.0213	fl4south	operrm	0.0474	fl8	fl6	0.0187	rad	aes	0.2214
burns	fl6	0.0109	fl4south	out	0.0255	fl8	fl7	0.0171	rad	burns	0.5139
burns	fl7	0.0299	fl4south	rad	0.0172	fl8	mssublue	0.0142	rad	fl4main	0.4589
burns	fl8	0.0242	fl5	burns	0.0233	fl8	operrm	0.0194	rad	fl5	0.4428
burns	mssublue	0.0125	fl5	fl4main	0.0199	fl8	out	0.0144	rad	fl6	0.4152
burns	operrm	0.0224	fl5	fl4south	0.0250	fl8	rad	0.0151	rad	fl7	0.4054
burns	out	0.0163	fl5	fl6	0.0215	fl9	ect	0.0208	rad	fl8	0.4174
burns	rad	0.0153	fl5	fl7	0.0175	fl9	out	0.0079	rad	mssublue	0.3604
ect	fl6	0.4839	fl5	fl8	0.0190	mssublue	fl4south	0.0259	rad	mssumaize	0.1917
ect	fl7	0.4348	fl5	mssublue	0.0127	mssublue	fl4main	0.0182	rad	operrm	0.3301
ect	fl8	0.3399	fl5	mssumaize	0.0367	mssublue	fl5	0.0237	rad	out	0.5240
ect	fl9	0.4959	fl5	operrm	0.0378	mssublue	fl6	0.0271	rad	ssu	0.4286
ect	out	0.4412	fl5	out	0.0162	mssublue	fl7	0.0236	ssu	fl4main	0.0415
entry	aes	3.0571	fl5	rad	0.0174	mssublue	fl8	0.0138	ssu	fl5	0.0435
entry	burns	0.0122	fl6	aes	3.2432	mssublue	operrm	0.0185	ssu	fl6	0.0300
entry	fl4main	0.0873	fl6	burns	0.0120	mssublue	out	0.0263	ssu	fl7	0.0216
entry	fl4south	0.0006	fl6	ect	0.0172	mssublue	rad	0.0203	ssu	fl8	0.0541
entry	fl5	0.1057	fl6	fl4main	0.0178	mssumaize	fl4main	0.0454	ssu	mssumaize	0.4412
entry	fl6	0.1587	fl6	fl4south	0.0419	mssumaize	fl5	0.0247	ssu	operrm	0.0762
entry	fl7	0.1108	fl6	fl5	0.0227	mssumaize	fl6	0.0315	ssu	out	0.0494
entry	fl8	0.1265	fl6	fl7	0.0230	mssumaize	fl7	0.0263	ssu	rad	0.0659
entry	fl9	0.0175	fl6	fl8	0.0217	mssumaize	fl8	0.0172			
entry	mssublue	0.0064	fl6	mssublue	0.0120	mssumaize	mssublue	0.0230			
entry	operrm	1.0700	fl6	operrm	0.0183	mssumaize	operrm	0.0245			
entry	pes	0.1539	fl6	out	0.0141	mssumaize	out	0.0416			
entry	rad	0.0611	fl6	rad	0.0146	mssumaize	rad	0.0522			
entry	ssu	0.0001									

Table 5-1 Network Level Between Unit Patient Movement Rates

Parameter	Value
Beta	(0.01,0.02,0.04,0.08,0.1)
Gamma	0.0069
Delta	0.0139
N	22

Table 5-2 SEIR Parameters

Unit	Model Initial Conditions
ssu	3
aes	42
pes	2
fl 4 south	7
fl 5	28
fl 6	28
fl 7	29
fl 8	29
fl 9	22
burns	7
operrm	16
mssublue	19
mssumaize	14
rad	2
ect	1
fl 4 main	24

Table 5-3 SEIR ODE Model Initital Conditions for Units When Not Used as Outbreak Source

5.2.3 Model Parameters: SEIR Compartmental Rates

Beta, the per capita effective contact rate was swept across a range of values: 0.01, 0.02,0.04,0.08,0.1 (roughly equivalent to an R0 of 3.69 to 37 using a single compartmental model without unit structure). Since we do not overtly capture staff patient interactions within the model, and because unit size varies, a range was chosen over a single value to account for a range of rates that could be influenced by outside factors. All other compartmental rates were based on SARS-CoV-2 Omicron transmission dynamics (although the model is not specifically intended to model SARS-CoV-2 Omicron. Delta, the transition rate between exposed and infected states, was set to 1/72 hours, using the estimate of 3 days provided by the CDC.[158] Gamma, the recovery rate, was set to 1/144 hours, using the estimate of 6-day duration of infectiousness published by the CDC.[158]

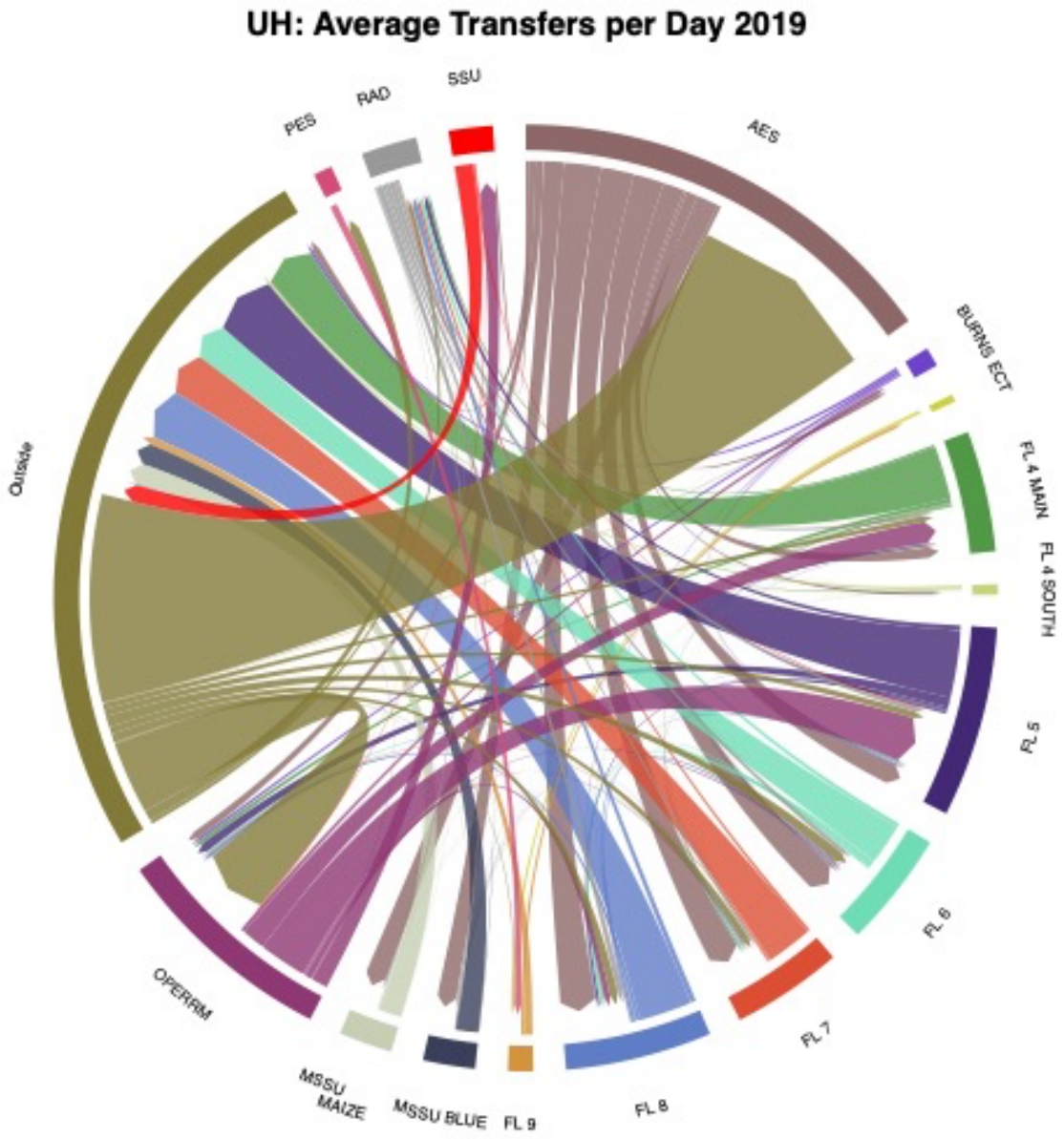
5.2.4 Initial conditions for occupancies

The median occupancy for each unit was obtained from the January 1, 2019 through December 31, 2019 patient movement data. These values were used as the initial conditions for the susceptible compartments in our model. [TABLE 5.3]. The model was initialized with no exposed or recovered individuals as we wanted to assess the outcome in a strictly naïve population, to explore how the network structure affects how a newly introduced infection might spread.

Outbreaks of 5 patients were initialized in one of three compartments, Adult Emergency Services, the Operating Room, and Electroconvulsive Therapy. Adult Emergency Services and the Operating Room were chosen because most patients enter the hospital through these two units. Electroconvulsive therapy was chosen as it serves as a bridge between Floor 9 and

Psychiatric Emergency Services and the rest of the hospital and is the primary entry route into these two units, and a higher volume of people enter Floor 9 through this unit rather than Psychiatric Emergency Services. (Figure 5.1)

Figure 5-1 Chord Diagram Showing Mean Daily Patient Movement



5.2.5 Measures:

Maximum number of infected, proportion of the unit population that is infected and mean stay of infected patients in person time were used in conjunction with movement rates and centrality to understand the scope and speed of infection transmission.

Unit population at the time of first infection represents the total number of people within a unit at the timestep when the first infection occurs. Maximum number of infected participants refers to the time point with the largest population of infected individuals. This measure was used along with the population at the time point of maximum infections to obtain the proportion of maximum infecteds.

5.2.6 Software:

All analyses were conducted in R.[114] The compartmental model was run using the deSolve package.[159] Igraph was used to visualize the network. Stats was used to calculate correlation and run the Pearson product-moment correlation test.[114]

5.3 Results

5.3.1 Infected as counts and proportions

Floors 4 South, 5 ,8, 7 and 9, Burns, MSSU Blue, MSSU Maize consistently had one or more infected patients for the range of beta values, regardless of where the outbreak was initialized. Floors 7 and 9 consistently the highest number of infected patients at each beta value, and the lowest out flow rates of the units. Floor 4 main had one or more infected patients at beta values 0.4 and above. All other units did not acquire one or more infected patients. (Figure 5.2, Supplement D for remaining catalogue of figures).

The proportion of infected patients in each unit was not impacted by the location of the initial outbreak. At a beta of 0.02 and above, the proportion of infected patients in each unit remained stable for betas 0.02 to 0.1 (Figure 5.3). At a beta of 0.01, the proportion of infected patients in a unit was approximately 5 points lower than at beta 0.02 and above. Person-Time did not appear to be strongly correlated with measures of centrality. (Figures 5.4-5.6). Even still, some crude trends can be captured. There does appear to be an inverse relationship between degree and closeness centrality and stay length among some units. However, stay in person-time did positively correspond to a unit's disease burden.

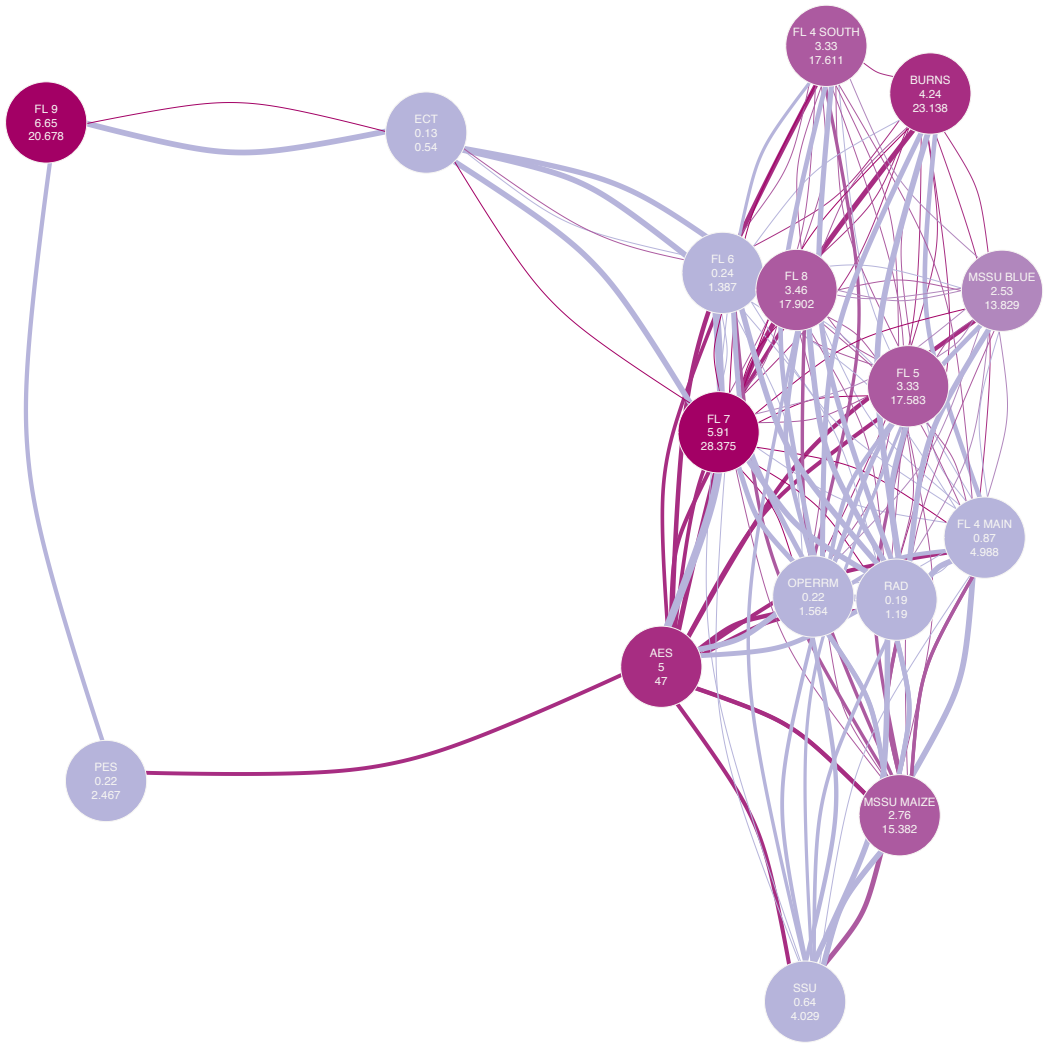


Figure 5-2 Network Nodes Colored by Maximum number of Infected. Nodes Show Unit Name, Max Count of Infected People, Total Unit Population

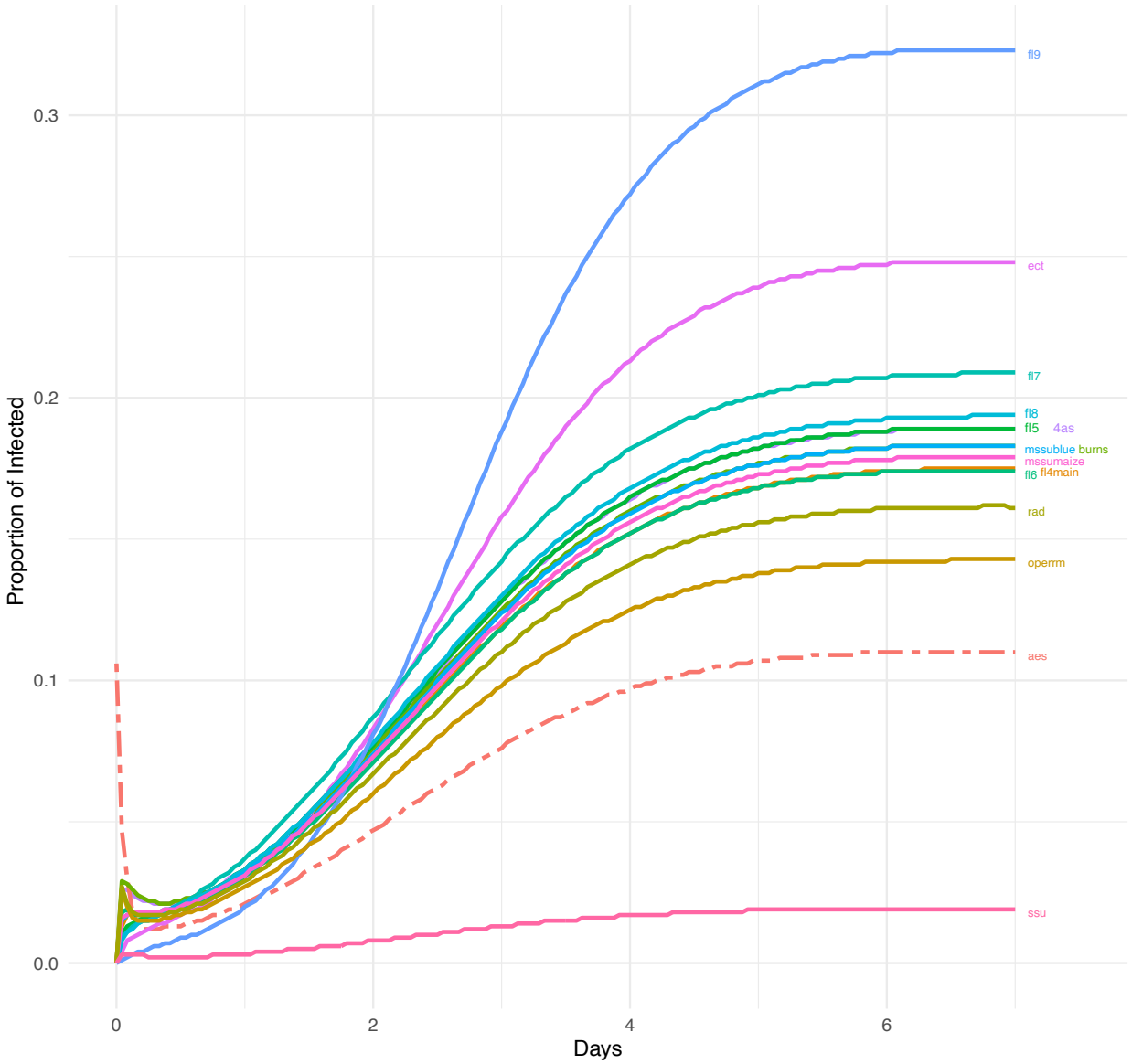


Figure 5-3 Number of Infected Over Time with Infection Starting in AES, Beta=0.02 Dashed Line Demarcates Unit were Infection Began

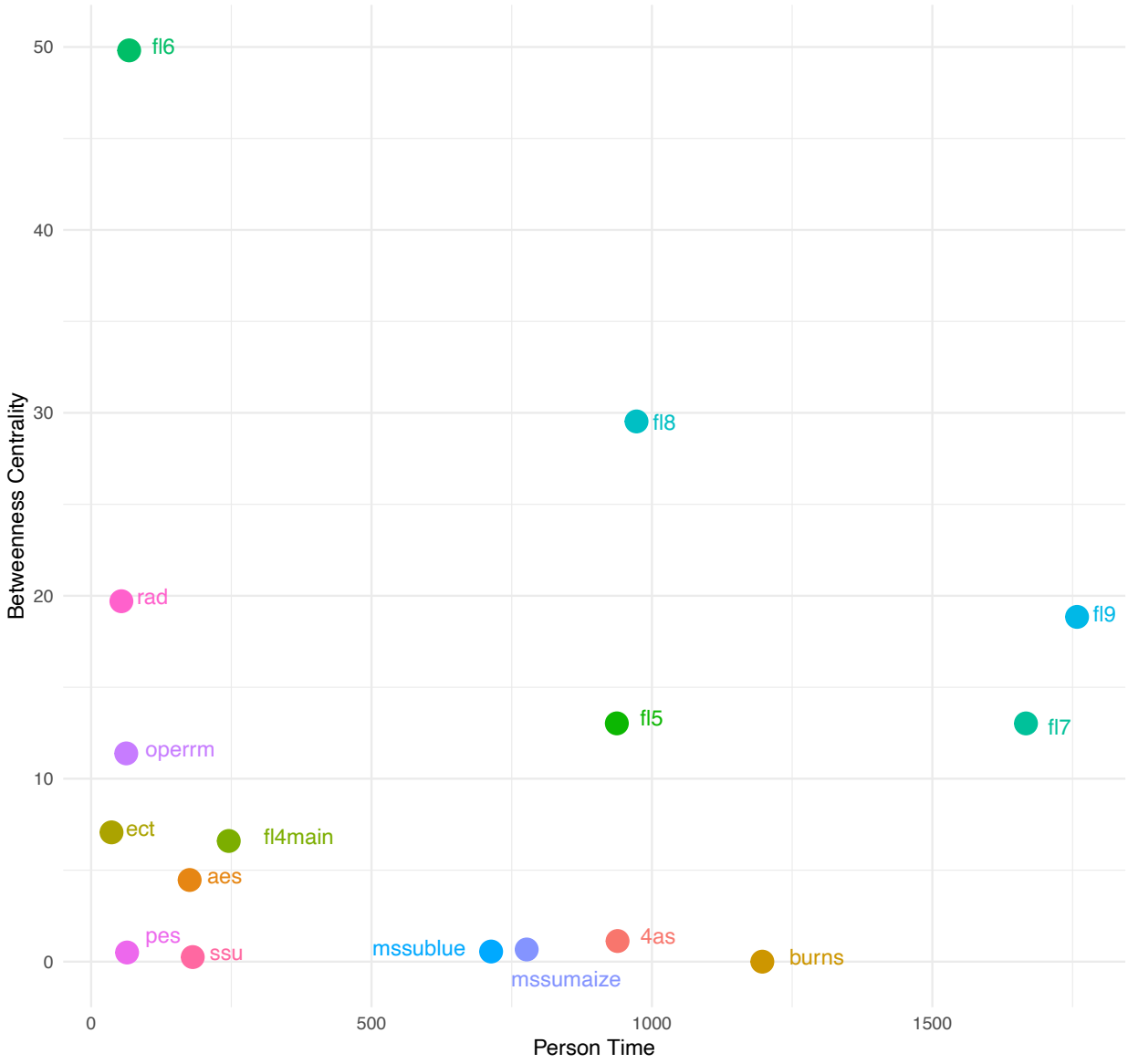


Figure 5-4 Relationship Between Mean Unit Stay in Person-Time of Infected People and Unit Betweenness Centrality

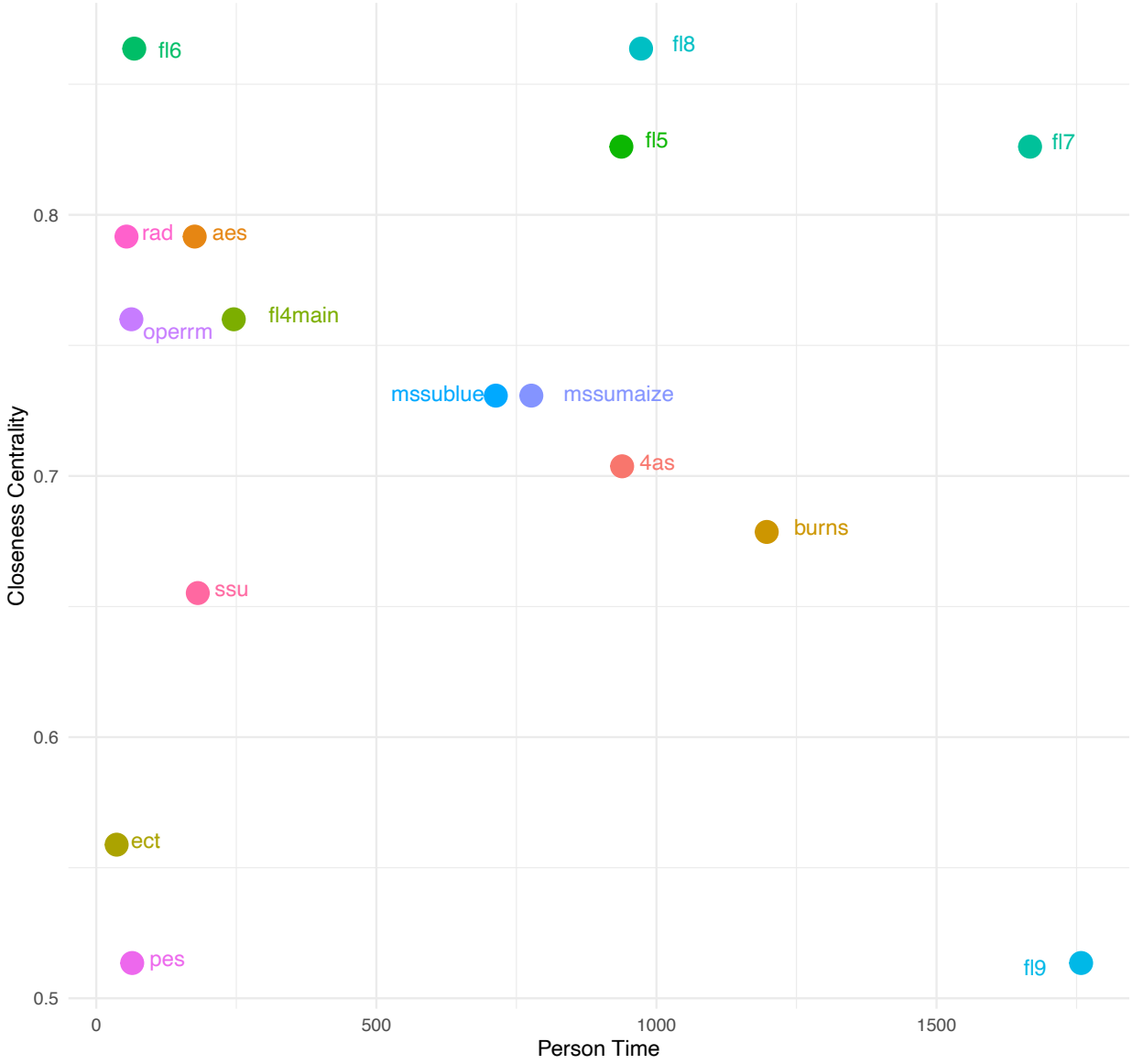


Figure 5-5 Relationship Between Unit Stay in Person-Time of Infected People and Unit Closeness Centrality

Figure 5-6 Relationship Between Unit Stay in Person-Time and Unit Degree Centrality

5.3.2 Correlates of Infection

Centrality does not appear to be significantly correlated with a unit's maximum number of infected. Likewise, with the exception of an infection beginning in AES at a beta of 0.01 the rate of patient movement into a unit does not show such a trend. Instead, it tends to be weakly correlated with a unit's maximum number of infected. (Table 5.4).

	0.01		0.02		0.04		0.08		0.1	
UNIT	**IN RATE**	**OUT RATE**	**IN RATE**	**OUT RATE**	**IN RATE**	**OUT RATE**	**IN RATE**	**OUT RATE**	**IN RATE**	**OUT RATE**
AES	**r:** 0.55 **p-val**: 0.026	**-0.59** **0.016**	0.34 0.203	**-0.69** **0.003**	0.24 0.376	**-0.73** **0.001**	0.19 0.481	**-0.74** **0.001**	0.18 0.50	**-0.74** **0.001**
ECT	-0.22 0.408	-0.15 0.584	-0.23 0.385	-0.36 0.168	-0.23 0.385	*-0.45* *0.083*	-0.23 0.385	**-0.49** **0.056**	-0.233 0.385	**-0.49** **0.052**
OPERRM	-0.16 0.564	**-0.36** **0.172**	-0.19 0.487	**-0.52** **0.037**	-0.19 0.47	**-0.59** **0.017**	-0.198 0.463	**-0.61** **0.011**	-0.20 0.461	**-0.62** **0.011**

Table 5-4 Correlation Coefficient and P-values for the Relationship between Unit In or Out-Rate and Infected Patients, By Unit of Outbreak Initiation, Across a Range of Betas

The rate at which people exit a unit on the other hand, does appear to be moderately to strongly inversely correlated with a unit's maximum number of infected, for infections beginning in AES and OPERRM. Interestingly, at betas less than 0.02, outbreaks beginning in ECT appear to be weakly inversely correlated with a unit's maximum number of infected individuals. At betas of 0.04 and 0.08 this correlation strengthens somewhat. For outbreaks that began in ECT, the relationship between unit-outflow and a unit's maximum number of infect patients becomes strongly inversely correlated. The correlation coefficients and their p-values can be found in Table 5-4.

5.3.3 Rate Threshold

There is evidence of a threshold for the unit-out rates. At beta values of 0.01 and 0.02, units with an out-rate of 0.032, (equivalent to a stay of 31.25 hours) or lower acquire one or more infected patients. At betas 0.04 and above units with an out-rate of 0.075 (equivalent to a stay of 13.3 hours) or below acquire one or more infected patients.

5.3.4 Discussion

The COVID-19 pandemic has illuminated existing weaknesses in the outbreak preparedness protocols of many hospitals. At the start of the pandemic hospitals were required to make rapid decisions on resource allocation, staffing, and patient isolation procedures with little guidance from existing literature. In an effort to assist such decisions in the future, data are needed to identify the characteristics that make hospital units most vulnerable to an outbreak. In this analysis we used a combination of SEIR compartmental models, and network analysis to identify aspects of patient movement that increase a unit's susceptibility to an outbreak.

We found that the rate at which patients leave a unit had a strong inverse correlation with the maximum number of infected patients in a unit. This finding was bolstered by the observation of a positive association between unit stay in person-time and maximum number of infections. Furthermore, rate seemed to elicit a modest threshold effect depending upon the per capita rate of effective contacts occurring in the population. Specifically, we found that for our model parameters, a patient stay of at least 13.3 hour was needed in order for at least one infected person to appear in a unit. This finding counters our expectation that a susceptible unit's centrality would drive an outbreak, or that susceptible units with high patient turnover would be most vulnerable to infection outbreaks.

Likewise, we found that outbreaks originating in ECT had the weakest correlation between a unit's outrate and their maximum number infected. This observation maybe due to the lower population size of ECT relative to AES or OPERRM, or it may be due to ECT's lower overall centrality compared to AES and OPERRM. More research is needed to understand what factors contribute to this finding.

This analysis has several shortcomings. First, staff were not included in our simulations, as we were unable to obtain information on staff numbers or movement patterns for each unit. This is problematic because in an in-patient hospital setting, patients spend most of their time stationary, having limited interactions with other patients. Staff on the other hand do move around more and interact with other patients. We were able to mitigate some of the impact of this by tuning beta through a range of values, which can implicitly capture some of the impact of staff and patient interactions on transmission rates. Likewise, staff are likely to be more compliant with infection prevention measures, and so the role they may play in infection transmission could be much lower than that of patients. Furthermore, future analysis should also

include model simulations wherein the mean wait time between units (rather than median) is used to define the rates, or where the full distribution of wait times is sampled from within the simulation. Additionally, given the relatively small population size of each unit, future work should use a stochastic rather than deterministic compartmental model, which would enable us to capture the distribution of outcomes rather than just the mean trajectory examined here. Similarly, this analysis assessed a narrow range of model parameters. While we did tune beta, we left the rate at which people become infectious, and the rate at which the recover constant. Consequently, these findings may only be generalizable to diseases with a similar set of transmission parameters. However, these finding do still indicate that the nested compartmental-network model used here is a useful tool for modeling outbreak hotspots in a hospital setting. Finally, to evaluate the role the network structure plays in the infection dynamics, future work should include a comparison of the infection spread when the network structure is included in the model, versus when the model (for example) collapses the entire population into a single randomly mixing group.

Despite these shortcomings, this analysis contributes to the literature in several ways. It demonstrates that a nested compartmental-network model is a useful tool for simulating outbreaks in highly separated spaces. It identifies factors that influence a hospital unit's susceptibility to an outbreak; and it uses a network framework to characterize patient movement. More investigation is needed to understand how changes in model parameters influence how outbreaks spread throughout a health care setting. Lastly, a major advantage and new contribution of this study is that we consider patient movement and transmission using an explicit, data-driven patient flow network—this type of network data is not commonly available

at such small spatial scales (e.g. unlike population flow networks using mobility or air travel data), and allows for network-generated insight into transmission patterns.

6 Conclusion

Sometimes traditional applications of our usual kit of analytic tools are not enough to analyze complex or large data. Canonical methods may not be able to detect important effects or interrelationships which in turn may influence data interpretation, and the application of findings. This dissertation explored three contexts across four aims in which such hidden relationships existed, and it tested methodologies that uncovered these relationships.

Aim 1 re-examined multiyear and multi-country survey data, and used a novel application of a standard tool. Motivated by recent findings that have challenged long held assumptions, we used meta-regression and 43 surveys worth of data to re-evaluate the association between wealth and HIV in Sub-Saharan Africa. Our findings reveal a complicated country specific relationship between country-level wealth, individual wealth, built environment, and HIV positivity; findings that are contrary to the long held belief that HIV burden in Sub-Saharan Africa is uniformly, inversely, and monotonically associated with country and individual wealth.

Beyond illustrating the utility of meta-regression, and supporting newer findings, this aim encourages introspection. Meta-regression is not a new tool, and earlier analyses were not limited because it had not been developed. As such, we are left to reflect on the hold of power assumptions and how seemingly minor analytic decisions can influence policy, and shape

the way investigators approach future analyses. Beyond highlighting the flexibility of meta-regression, it may be that this aim serves as a cautionary tale about the powers of assumptions.

The second analytic context featured data with variables that are strongly, and dynamically interrelated. Aim 2 illustrated how latent relationships can be utilized to both circumvent analytic issues caused by temporal and collinear data, and to tease apart associations that have become muddled by external factors. This aim begins by reflecting on the complexity temporal trends and socio-cultural factors can bring to an epidemiologic analysis. Next, it offers an analytic solution that exploits this complexity. Using emergent characteristics of correlated covariates, we were able to define and enumerate a set of latent variables that were used to intergenerational changes in behaviors associated with increased risk of a contracting sexually transmitted infection; specifically number of oral and vaginal sex partners in the last year, and usage of marijuana, alcohol, and tobacco.

The final analytic context investigated nested, enmeshed systems, and reflected on the importance of scale and external influence in mathematical models. Aims 3 & 4 framed a hospital in terms of a network of units connected by patient traffic. They investigated how this structure, which serves as a proxy the physical space patients and staff move though, exerted an influence on the volume of patient movement and the dissemination of an infectious pathogen after an outbreak. Like Aim 1, these aims illustrated the importance of considering if and how macro-scale factors influence mezzo and micro-scale phenomena. Unlike Aim 1 these aims considered how these external influences impacted the functioning of dynamical systems. While model parsimony is often sought by modelers, these aims asked us to reflect on our assumptions carefully and to consider what we stand to lose for the sake of model simplicity.

At its heart this dissertation is introspective and philosophical in nature. On its face it offers practical solutions for problems and complexity encountered in data analysis. However, in the broader scheme it asks readers to reflect on and challenge analytic assumptions. In doing so it encourages readers to evaluate alternative perspectives that may change the how data are used, and to consider how existing tools can be extended to solve novel problems.

References

1. Faust, L., Yaya, S. & Ekholuenetale, M. Wealth inequality as a predictor of HIV-related knowledge in Nigeria. *BMJ Glob Health* (2017) doi:10.1136/bmjgh-2017-000461.
2. Fox, A. M. The Social Determinants of HIV Serostatus in Sub-Saharan Africa: An Inverse Relationship between Poverty and HIV? *Public Health Reports* (2010) doi:10.1177/00333549101250s405.
3. Fox, A. M. The HIV-poverty thesis RE-examined: Poverty, wealth or inequality as a social determinant of hiv infection in sub-Saharan Africa? *J Biosoc Sci* (2012) doi:10.1017/S0021932011000745.
4. Richardson, E. T. *et al.* Gender inequality and HIV transmission: A global analysis. *J Int AIDS Soc* (2014) doi:10.7448/IAS.17.1.19035.
5. Hajizadeh, M., Sia, D., Heymann, S. J. & Nandi, A. Socioeconomic inequalities in HIV/AIDS prevalence in sub-Saharan African countries: Evidence from the Demographic Health Surveys. *Int J Equity Health* (2014) doi:10.1186/1475-9276-13-18.
6. Barankanira, E., Molinari, N., Niyongabo, T. & Laurent, C. Spatial analysis of HIV infection and associated individual characteristics in Burundi: Indications for effective prevention. *BMC Public Health* (2016) doi:10.1186/s12889-016-2760-3.
7. Lakew, Y., Benedict, S. & Haile, D. Social determinants of HIV infection, hotspot areas and subpopulation groups in Ethiopia: evidence from the National Demographic and Health Survey in 2011. *BMJ Open* (2015) doi:10.1136/bmjopen-2015-008669.

8. Magadi, M. A. The disproportionate high risk of HIV infection among the urban poor in sub-Saharan Africa. *AIDS Behav* (2013) doi:10.1007/s10461-012-0217-y.

9. Mishra, V. *et al.* HIV infection does not disproportionately affect the poorer in sub-Saharan Africa. in *AIDS* (2007). doi:10.1097/01.aids.0000300532.51860.2a.

10. Higgins, J. *et al.* Chapter 10: Analysing data and undertaking meta-analyses | Cochrane Training. *Cochrane Handbook for Systematic Reviews of Interventions version 6.2* (2021).

11. Columbia University Mailman School of Public Health. Population Health Methods. *https://www.publichealth.columbia.edu/research/population-health-methods/meta-regression* (2022).

12. Thompson, S. G. & Higgins, J. P. T. How should meta-regression analyses be undertaken and interpreted? *Stat Med* **21**, (2002).

13. The World Bank. DataBank Metadata Glossary. *https://databank.worldbank.org/metadataglossary/jobs/series/SI.POV.GINI* (2022).

14. Veron, J. *et al.* Statisticians of the Centuries. *Population (French Edition)* **57**, (2002).

15. Lorenz, M. O. Methods of measuring the concentration of wealth. *Publications of the American Statistical Association* **9**, (1905).

16. Gastwirth, J. L. The Estimation of the Lorenz Curve and Gini Index. *Rev Econ Stat* **54**, (1972).

17. United Nations Development Program. Human Development Index. *https://hdr.undp.org/data-center/human-development-index#/indicies/HDI* (2022).

18. Pilowsky, D. & Wu, L.-T. Sexual risk behaviors and HIV risk among Americans aged 50 years or older: a review. *Subst Abuse Rehabil* (2015) doi:10.2147/sar.s78808.

19. Minichiello, V., Rahman, S., Hawkes, G. & Pitts, M. STI epidemiology in the global older population: Emerging challenges. *Perspectives in Public Health* vol. 132 Preprint at https://doi.org/10.1177/1757913912445688 (2012).

20. Centers for Disease Control and Prevention. CDC Atlas Plus. *https://gis.cdc.gov/grasp/nchhstpatlas/charts.html* (2022).

21. Espinoza, L. *et al.* Adolescent Substance Use and Sensation-Seeking on Sexual Behaviors Among Young Adults from Continuation High Schools. *Subst Use Misuse* **54**, 373–383 (2019).

22. Doran, K. A. & Waldron, M. Timing of First Alcohol Use and First Sex in Male and Female Adolescents. *J Adolesc Health* **61**, 606–611 (2017).

23. Vasilenko, S. A. & Lanza, S. T. Predictors of multiple sexual partners from adolescence through young adulthood. *J Adolesc Health* **55**, 491–497 (2014).

24. Chambers, R. S. *et al.* Predictors of Sexually Transmitted Infection Positivity among Substance-Using Native American Adults. *Sex Transm Dis* **47**, (2020).

25. Berlin, K. S., Williams, N. A. & Parra, G. R. An Introduction to Latent Variable Mixture Modeling (Part 1): Overview and Cross-Sectional Latent Class and Latent Profile Analyses. *J Pediatr Psychol* **39**, 174–187 (2014).

26. Petersen, K. J., Qualter, P. & Humphrey, N. The Application of Latent Class Analysis for Investigating Population Child Mental Health: A Systematic Review. *Front Psychol* **10**, (2019).

27. Kempf-Leonard, K. *Encyclopedia of Social Measurement*. *Encyclopedia of Social Measurement* (2004). doi:10.5860/choice.42-5629.

28. Collins, L. M. & Lanza, S. T. *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences* (2010). doi:10.1002/9780470567333.

29. Newman, M. Networks: An introduction. *Underst Complex Syst* 163–186 (2014) doi:10.1007/978-3-319-03518-5-8.

30. Csardi, G. & Nepusz, T. The igraph Software Package for Complex Network Research. *InterJournal* **Complex Sy**, 1695 (2006).

31. Kolaczyk, E. D. & Csárdi, G. *Statistical Analysis of Network Data with R*. (Springer International Publishing, 2020). doi:10.1007/978-3-030-44129-6.

32. Krnc, M. & Škrekovski, R. Group degree centrality and centralization in networks. *Mathematics* **8**, (2020).

33. Freeman, L. C. Centrality in social networks conceptual clarification. *Soc Networks* **1**, (1978).

34. Vynnycky, E. & White, R. *An Introduction to Infectious Disease Modeling*. (Oxford University Press, 2010).

35. Gillespie, S., Kadiyala, S. & Greener, R. Is poverty or wealth driving HIV transmission? in *AIDS* (2007). doi:10.1097/01.aids.0000300531.74730.72.

36. Phelan, J. C., Link, B. G. & Tehranifar, P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. *J Health Soc Behav* (2010) doi:10.1177/0022146510383498.

37. Smith, J. P. Socioeconomic Status and Health. *Am Econ Rev* **88**, 192–196 (1998).

38. Robert, S. A. Community-Level Socioeconomic Status Effects on Adult Health. *J Health Soc Behav* **39**, 18–37 (1998).

39. Phelan, J. C. & Link, B. G. Fundamental Cause Theory BT - Medical Sociology on the Move: New Directions in Theory. in (ed. Cockerham, W. C.) 105–125 (Springer Netherlands, 2013). doi:10.1007/978-94-007-6193-3_6.

40. Clouston, S. A. P., Rubin, M. S., Phelan, J. C. & Link, B. G. A Social History of Disease: Contextualizing the Rise and Fall of Social Inequalities in Cause-Specific Mortality. *Demography* (2016) doi:10.1007/s13524-016-0495-5.

41. The World Bank. World Bank Open Data. https://data.worldbank.org (2019).

42. Johnson, K. & Way, A. Risk factors for HIV infection in a national adult population: Evidence from the 2003 Kenya demographic and health survey. *J Acquir Immune Defic Syndr (1988)* (2006) doi:10.1097/01.qai.0000225870.87456.ae.

43. Mojola, S. A. *Love, money, and HIV: Becoming a modern African woman in the age of AIDS. Love, Money, and HIV: Becoming a Modern African Woman in the Age of AIDS* (2014). doi:10.1177/0094306116641407ee.

44. Hadley, C., Maxfield, A. & Hruschka, D. Different forms of household wealth are associated with opposing risks for HIV infection in East Africa. *World Dev* (2019) doi:10.1016/j.worlddev.2018.09.015.

45. Abimiku, A. & Gallo, R. C. HIV: Basic Virology and Pathophysiology. in *HIV in US Women* (eds. Minkoff, H., DeHovitz, J. A. & Duerr, A.) 13–32 (Raven Press, 1995).

46. Bolan, G., Ehrhardt, A. A. & Wasserheit, J. N. Gendered Perspectives and STDs. in *Sexually Transmitted Diseases* (eds. Holmes, K. K. et al.) 117–128 (McGraw Hill Professional, 1999).

47. Glynn, J. R. *et al.* Why do young women have a much higher prevalence of HIV than young men? A study in Kisumu, Kenya and Ndola, Zambia. *AIDS* (2001) doi:10.1097/00002030-200108004-00006.

48. Bunyasi, E. W. & Coetzee, D. J. Relationship between socioeconomic status and HIV infection: Findings from a survey in the Free State and Western Cape Provinces of South Africa. *BMJ Open* Preprint at https://doi.org/10.1136/bmjopen-2017-016232 (2017).

49. Ishida, K., Arnold, M., Stupp, P., Kizito, P. & Ichwara, J. Exploring the connections between HIV serostatus and individual, household, and community socioeconomic resources: Evidence from two population-based surveys in Kenya. *Soc Sci Med* (2012) doi:10.1016/j.socscimed.2011.10.019.

50. Parkhurst, J. O. Understanding the correlations between wealth, poverty and human immunodeficiency virus infection in African countries. *Bull World Health Organ* (2010) doi:10.2471/BLT.09.070185.

51. UNAIDS. *Access to Antiretroviral Therapy in Africa, 2013*. *Status report on progress towards the 2015 targets* (2013).

52. Burger, C., Burger, R. & van Doorslaer, E. The Health Impact of Free Access to Antiretroviral Therapy in South Africa. *SSRN Electronic Journal* (2020) doi:10.2139/ssrn.3506209.

53. Murphy, E. M., Greene, M. E., Mihailovic, A. & Olupot-Olupot, P. Was the 'ABC' approach (abstinence, being faithful, using condoms) responsible for Uganda's decline in HIV? *PLoS Medicine* Preprint at https://doi.org/10.1371/journal.pmed.0030379 (2006).

54. Okware, S., Kinsman, J., Onyango, S., Opio, A. & Kaggwa, P. Revisiting the ABC strategy: HIV prevention in Uganda in the era of antiretroviral therapy. *Postgraduate Medical Journal* Preprint at https://doi.org/10.1136/pgmj.2005.032425 (2005).

55. Giguère, K. *et al.* Trends in knowledge of HIV status and efficiency of HIV testing services in sub-Saharan Africa, 2000–20: a modelling study using survey and HIV testing programme data. *Lancet HIV* (2021) doi:10.1016/s2352-3018(20)30315-5.

56. Hlongwa, M., Mashamba-Thompson, T., Makhunga, S. & Hlongwana, K. Mapping evidence of intervention strategies to improving men's uptake to HIV testing services in sub-Saharan Africa: A systematic scoping review. *BMC Infectious Diseases* Preprint at https://doi.org/10.1186/s12879-019-4124-y (2019).

57. Singh, S., Darroch, J. E. & Bankole, A. A, B and C in Uganda: the roles of abstinence, monogamy and condom use in HIV decline. *Reprod Health Matters* (2004) doi:10.1016/S0968-8080(04)23118-4.

58. Cohen, S. Beyond slogans: lessons from Uganda's experience with ABC and HIV/AIDS. *Reprod Health Matters* (2004).

59. Murphy, E. M. Being Born Female is Dangerous for Your Health. *American Psychologist* Preprint at https://doi.org/10.1037/0003-066X.58.3.205 (2003).

60. Alsan, M. M. & Cutler, D. M. Girls' education and HIV risk: Evidence from Uganda. *J Health Econ* (2013) doi:10.1016/j.jhealeco.2013.06.002.

61. Sia, D. *et al.* The effect of gender inequality on HIV incidence in Sub-Saharan Africa. *Public Health* (2020) doi:10.1016/j.puhe.2020.01.014.

62. Adamczyk, A. & Greif, M. Education and risky sex in Africa: Unraveling the link between women's education and reproductive health behaviors in Kenya. *Soc Sci Res* (2011) doi:10.1016/j.ssresearch.2010.12.003.

63. Pulerwitz, J., Amaro, H., De Jong, W., Gortmaker, S. L. & Rudd, R. Relationship power, condom use and HIV risk among women in the USA. *AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV* (2002) doi:10.1080/0954012021000031868.

64. Jewkes, R. Intimate partner violence: Causes and prevention. *Lancet* Preprint at https://doi.org/10.1016/S0140-6736(02)08357-5 (2002).

65. Croft, T. N., Marshall, A. M. J. & Allen, C. K. Guide to DHS Statistics. *Rockville, Maryland, USA: ICF* (2018).

66. The Demographic and Health Survey Program. Protecting the Privacy of DHS Respondents. *The Demographic and Health Survey Program* (2021).

67. Rutstein, S. O. & Johnson, K. *The DHS Wealth Index, DHS Comparative Report 6*. *Measure DHS+, ORC Macro* (2004).

68. Magadi, M. A. Understanding the urban-rural disparity in HIV and poverty nexus: The case of Kenya. *Journal of Public Health (United Kingdom)* (2017) doi:10.1093/pubmed/fdw065.

69. United Nations Development Programme. *Technical note 1. Human Development Index*. *Human Development Report* (2020).

70. United Nations Development Programme. Human Development Index (HDI). *United Nations Development Reports* http://hdr.undp.org/en/content/human-development-index-hdi (2021).

71. Elizabeth Heger, B., King, M. & Sobek, M. PUMS-Demographic and Health Surveys: Version 8. *Minneapolis: University of Minnesota* https://www.idhsdata.org/idhs-action/variables/URBAN#description_section (2020).

72. R Core Team. R: A Language and Environment for Statistical Computing. Preprint at (2018).

73. Brilleman, S. L., Crowther, M. J., Moreno-Betancur, M., Buros Novik, J. & Wolfe, R. Joint longitudinal and time-to-event models via {Stan}.

74. Higgins, J. P. & Green, S. *Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series*. *Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series* (2008). doi:10.1002/9780470712184.

75. Viechtbauer, W. Metafor: Meta-Regression Package in R. 1--48 Preprint at (2010).

76. Kabudula, C. W. *et al.* Socioeconomic differences in mortality in the antiretroviral therapy era in Agincourt, rural South Africa, 2001–13: a population surveillance analysis. *Lancet Glob Health* (2017) doi:10.1016/S2214-109X(17)30297-8.

77. Bovbjerg, M. Differences between Confounding and Effect Modification. in *Foundations of Epidemiology* 97–108 (Oregon State University, 2020).

78. UNAIDS. UNAIDS 2019 Data. *Joint United Nations Programme on HIV/AIDS (UNAIDS)* (2019) doi:978-92-9173-945-5.

79. Riley, E. D., Gandhi, M., Hare, C. B., Cohen, J. & Hwang, S. W. Poverty, unstable housing, and HIV infection among women living in the United States. *Curr HIV/AIDS Rep* (2007) doi:10.1007/s11904-007-0026-5.

80. Rubin, M. S., Colen, C. G. & Link, B. G. Examination of inequalities in HIV/AIDS mortality in the United States from a fundamental cause perspective. *American Journal of Public Health* Preprint at https://doi.org/10.2105/AJPH.2009.170241 (2010).

81. Merten, S. *et al.* Patient-reported barriers and drivers of adherence to antiretrovirals in sub-Saharan Africa: a meta-ethnography. *Tropical Medicine & International Health* **15**, 16–33 (2010).

82. Kumarasamy, N. *et al.* Barriers and facilitators to antiretroviral medication adherence among patients with HIV in Chennai, India: A qualitative study. *AIDS Patient Care STDS* (2005) doi:10.1089/apc.2005.19.526.

83. Tuller, D. M. *et al.* Transportation costs impede sustained adherence and access to HAART in a clinic population in Southwestern Uganda: A qualitative study. *AIDS Behav* (2010) doi:10.1007/s10461-009-9533-2.

84. Price, A. J. *et al.* Sustained 10-year gain in adult life expectancy following antiretroviral therapy roll-out in rural Malawi: July 2005 to June 2014. *Int J Epidemiol* (2017) doi:10.1093/ije/dyw208.

85. Floyd, S. *et al.* Population-level reduction in adult mortality after extension of free anti-retroviral therapy provision into rural areas in Northern Malawi. *PLoS One* (2010) doi:10.1371/journal.pone.0013499.

86. Bor, J., Herbst, A. J., Newell, M. L. & Bärnighausen, T. Increases in adult life expectancy in rural South Africa: Valuing the scale-up of HIV treatment. *Science (1979)* (2013) doi:10.1126/science.1230413.

87. National Center forfor HIV Viral Hepatitis STD and TB Prevention. *1 in 5 People in the US have a sexually transmitted infection*.

https://www.cdc.gov/nchhstp/newsroom/2021/2018-STI-incidence-prevalence-estimates.html (2021).

88. National Center for HIV Viral Hepatitis STD and TB Prevention - Division of STD Prevention. Congenital Syphilis. *STI Treatment Guidelines* https://www.cdc.gov/std/treatment-guidelines/congenital-syphilis.htm (2021).

89. Belgrave, F. Z. *et al.* "I Don't Know and I Don't Want to Know": A Qualitative Examination of Older African American Women's Knowledge and Experiences With HIV. *Journal of Black Psychology* **44**, (2018).

90. Smith, M. L., Bergeron, C. D., Goltz, H. H., Coffey, T. & Boolani, A. Sexually transmitted infection knowledge among older adults: Psychometrics and test–retest reliability. *Int J Environ Res Public Health* **17**, (2020).

91. Leatherdale, S. T., Hammond, D. & Ahmed, R. Alcohol, marijuana, and tobacco use patterns among youth in Canada. *Cancer Causes and Control* **19**, (2008).

92. Crosby, G. M., Stall, R. D., Paul, J. P. & Barrett, D. C. Alcohol and drug use patterns have declined between generations of younger gay-bisexual men in San Francisco. *Drug Alcohol Depend* **52**, (1998).

93. Maksut, J. L., Eaton, L. A., Siembida, E. J., Driffin, D. D. & Baldwin, R. An evaluation of factors associated with sexual risk taking among Black men who have sex with men: a comparison of younger and older populations. *J Behav Med* **39**, (2016).

94. Lee, A. S. D. & Cody, S. L. The Stigma of Sexually Transmitted Infections. *Nursing Clinics of North America* vol. 55 Preprint at https://doi.org/10.1016/j.cnur.2020.05.002 (2020).

95. Chung, T., Hipwell, A. E., Stepp, S. D., Miller, E. & Sartor, C. E. Profiles of young women's alcohol and cannabis use linked to risk for sexually transmitted infection highlight the importance of multi-level targeted interventions: Findings from the Pittsburgh girls study. *Subst Abus* **43**, (2022).

96. Cavazos-Rehg, P. A. *et al.* Number of sexual partners and associations with initiation and intensity of substance use. *AIDS Behav* **15**, 869–874 (2011).

97. Harris, A. L. 'I got caught up in the game': Generational influences on contraceptive decision making in African-American women. *J Am Acad Nurse Pract* **25**, (2013).

98. Hill, M., Sternberg, A., Suk, H. W., Meier, M. H. & Chassin, L. The intergenerational transmission of cannabis use: Associations between parental history of cannabis use and cannabis use disorder, low positive parenting, and offspring cannabis use. in *Psychology of Addictive Behaviors* vol. 32 (2018).

99. Lee, M. R. & Sher, K. J. 'Maturing Out' of Binge and Problem Drinking. *Alcohol Res* **39**, 31–42 (2018).

100. Windle, M. Maturing Out of Alcohol Use in Young Adulthood: Latent Class Growth Trajectories and Concurrent Young Adult Correlates. *Alcohol Clin Exp Res* **44**, 532–540 (2020).

101. Trudel, G., Turgeon, L. & Piche, L. Marital and sexual aspects of old age. *Sexual and Relationship Therapy* vol. 15 Preprint at https://doi.org/10.1080/713697433 (2000).

102. Gillison, M. L. *et al.* Human papillomavirus and diseases of the upper airway: Head and neck cancer and respiratory papillomatosis. *Vaccine* Preprint at https://doi.org/10.1016/j.vaccine.2012.05.070 (2012).

103. Berlin, K. S., Parra, G. R. & Williams, N. A. An Introduction to Latent Variable Mixture Modeling (Part 2): Longitudinal Latent Class Growth Analysis and Growth Mixture Models. *J Pediatr Psychol* **39**, 188–203 (2014).

104. Muthén, L. K. & Muthén, B. O. Mplus user's guide (8 ed). *Los Angeles, CA: Muthén & Muthén* Preprint at (2017).

105. Centers for Disease Control and Prevention (CDC) & National Center for Health Statistics. National Health and Nutrition Examination Survey Data. https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2013 (2013).

106. Nylund, K. L., Asparouhov, T. & Muthén, B. O. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling* (2007) doi:10.1080/10705510701575396.

107. RStudio Team. RStudio: Integrated Development Environment for R. Preprint at (2015).

108. Visser, I. & Speekenbrink, M. depmixS4: An R Package for Hidden Markov Models. *J Stat Softw* **36**, 1–21 (2010).

109. Nylund-Gibson, K. & Choi, A. Y. Ten frequently asked questions about latent class analysis. *Transl Issues Psychol Sci* **4**, 440–461 (2018).

110. Finer, L. B., Darroch, J. E. & Singh, S. Sexual partnership patterns as a behavioral risk factor for sexually transmitted diseases. *Fam Plann Perspect* **31**, 228–236 (1999).

111. Garnett, G. P. & Anderson, R. M. Sexually Transmitted Diseases And Sexual Behavior: Insights From Mathematical Models. *J Infect Dis* **174**, S150–S161 (1996).

112. Joffe, G. P. *et al.* Multiple partners and partner choice as risk factors for sexually transmitted disease among female college students. *Sex Transm Dis* **19**, 272–278 (1992).

113. Jessor, R. Risk behavior in adolescence: a psychosocial framework for understanding and action. *J Adolesc Health* **12**, 597–605 (1991).

114. R Core Team. R: A Language and Environment for Statistical Computing. Preprint at https://www.r-project.org/ (2020).

115. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models Usinglme4. *J. Stat. Softw.* **67**, (2015).

116. Schick, V. *et al.* Sexual behaviors, condom use, and sexual health of americans over 50: Implications for sexual health promotion for older adults. *Journal of Sexual Medicine* **7**, (2010).

117. Wang, Y. P. & Andrade, L. H. Epidemiology of alcohol and drug use in the elderly. *Current Opinion in Psychiatry* vol. 26 Preprint at https://doi.org/10.1097/YCO.0b013e328360eafd (2013).

118. Brouwer, A. F., Eisenberg, M. C., Carey, T. E. & Meza, R. Multisite HPV infections in the United States (NHANES 2003–2014): An overview and synthesis. *Prev Med (Baltim)* **123**, (2019).

119. Brouwer, A. F., Eisenberg, M. C., Carey, T. E. & Meza, R. Trends in HPV cervical and seroprevalence and associations between oral and genital infection and serum antibodies in NHANES 2003-2012. *BMC Infect Dis* **15**, 575 (2015).

120. Bose J, Hedden SL, Lipari RN, P.-L. E. Key Substance Use and Mental Health Indicators in the United States: Results from the 2015 National Survey on Drug Use and Health. *Substance Abuse and Mental Health Services Administration* **75**, (2016).

121. Dawson, D. A., Grant, B. F., Stinson, F. S. & Chou, P. S. Maturing out of alcohol dependence: The impact of transitional life events. *J Stud Alcohol* **67**, (2006).

122. Lee, M. R., Chassin, L. & Mackinnon, D. P. Role Transitions and Young Adult Maturing Out of Heavy Drinking: Evidence for Larger Effects of Marriage Among More Severe Premarriage Problem Drinkers. *Alcohol Clin Exp Res* **39**, (2015).

123. Flora, D. B. & Chassin, L. Changes in drug use during young adulthood: The effects of parent alcoholism and transition into marriage. *Psychology of Addictive Behaviors* **19**, (2005).

124. Hoffmann, J. P., Dufur, M. & Huang, L. Drug use and job quits: A longitudinal analysis. *J Drug Issues* **37**, (2007).

125. Waller, M. W. *et al.* Gender differences in associations between depressive symptoms and patterns of substance use and risky sexual behavior among a nationally representative sample of U.S. adolescents. *Arch Womens Ment Health* **9**, (2006).

126. Firkey, M. K., Sheinfil, A. Z. & Woolf-King, S. E. Substance use, sexual behavior, and general well-being of U.S. college students during the COVID-19 pandemic: A brief report. *Journal of American College Health* (2020) doi:10.1080/07448481.2020.1869750.

127. Shrier, L. A., Harris, S. K., Sternberg, M. & Beardslee, W. R. Associations of depression, self-esteem, and substance use with sexual risk among adolescents. *Prev Med (Baltim)* **33**, (2001).

128. Chassin, L., Sher, K. J., Hussong, A. & Curran, P. The developmental psychopathology of alcohol use and alcohol disorders: Research achievements and future directions. *Dev Psychopathol* **25**, (2013).

129. Talley, A. E. *et al.* Addressing gaps on risk and resilience factors for alcohol use outcomes in sexual and gender minority populations. *Drug and Alcohol Review* vol. 35 Preprint at https://doi.org/10.1111/dar.12387 (2016).

130. White, A. *et al.* Converging Patterns of Alcohol Use and Related Outcomes Among Females and Males in the United States, 2002 to 2012. *Alcohol Clin Exp Res* **39**, (2015).

131. Chow, E. J. & Mermel, L. A. Hospital-Acquired Respiratory Viral Infections: Incidence, Morbidity, and Mortality in Pediatric and Adult Patients. *Open Forum Infect Dis* **4**, ofx006 (2017).

132. Tablan, O. C., Anderson, L. J., Besser, R., Bridges, C. & Hajjeh, R. Healthcare Infection Control Practices Advisory Committee. (2004).

133. Leung, N. H. L. Transmissibility and transmission of respiratory viruses. *Nat Rev Microbiol* **19**, 528–545 (2021).

134. Leung, N. H. L. *et al.* Quantification of Influenza Virus RNA in Aerosols in Patient Rooms. *PLoS One* **11**, e0148669 (2016).

135. Yu, I. T. S. *et al.* Evidence of Airborne Transmission of the Severe Acute Respiratory Syndrome Virus. *New England Journal of Medicine* **350**, 1731–1739 (2004).

136. Wang, C. C. *et al.* Airborne transmission of respiratory viruses. *Science* **373**, (2021).

137. Bischoff, W. E., Swett, K., Leng, I. & Peters, T. R. Exposure to influenza virus aerosols during routine patient care. *J Infect Dis* **207**, 1037–1046 (2013).

138. Hafner, C. M. The spread of the Covid-19 pandemic in time and space. *Int J Environ Res Public Health* **17**, (2020).

139. Callaway, E. Fast-spreading COVID variant can elude immune responses. *Nature* vol. 589 Preprint at https://doi.org/10.1038/d41586-021-00121-z (2021).

140. Howard, J. *et al.* An evidence review of face masks against COVID-19. *Proceedings of the National Academy of Sciences of the United States of America* vol. 118 Preprint at https://doi.org/10.1073/pnas.2014564118 (2021).

141. Tian, Z. *et al.* Personal protective equipment (PPE) and infection among healthcare workers – What is the evidence? *Int J Clin Pract* **74**, (2020).

142. Brandt, A. M. The History of Contact Tracing and the Future of Public Health. *Am J Public Health* **112**, 1097–1099 (2022).

143. El-Sadr, W. M., Platt, J., Bernitz, M. & Reyes, M. Contact Tracing: Barriers and Facilitators. *Am J Public Health* **112**, 1025–1033 (2022).

144. Goodreau, S. M. A decade of modelling research yields considerable evidence for the importance of concurrency: a response to Sawers and Stillwaggon. *J Int AIDS Soc* **14**, 12 (2011).

145. Morris, M. Data Driven Network Models for the Spread of Infectious Disease. in *Epidemic Models* (1995).

146. Kretzschmar, M. & Morris, M. Measures of concurrency in networks and the spread of infectious disease. *Math Biosci* **133**, (1996).

147. Jenness, S. M., Willebrand, K. S., Malik, A. A., Lopman, B. A. & Omer, S. B. Dynamic network strategies for SARS-CoV-2 control on a cruise ship. *Epidemics* **37**, (2021).

148. Donker, T., Wallinga, J. & Grundmann, H. Patient referral patterns and the spread of hospital-acquired infections through national health care networks. *PLoS Comput Biol* **6**, (2010).

149. Donker, T., Wallinga, J., Slack, R. & Grundmann, H. Hospital networks and the dispersal of hospital-acquired pathogens by patient transfer. *PLoS One* **7**, (2012).

150. English, K. M. *et al.* Contact among healthcare workers in the hospital setting: Developing the evidence base for innovative approaches to infection control. *BMC Infect Dis* **18**, (2018).

151. Luke, D. A. *A User's Guide to Network Analysis in R*. (2015). doi:10.1007/978-3-319-23883-8.

152. Newman, M. E. J. Modularity and community structure in networks. *Proc Natl Acad Sci U S A* **103**, (2006).

153. Flor, M. R interface to D3 chord diagrams. Chorddiag. *https://github.com/mattflor/chorddiag.*

154. Wickham, H. ggplot2 Elegant Graphics for Data Analysis.

155. Broido, A. D. & Clauset, A. Scale-free networks are rare. *Nat Commun* **10**, (2019).

156. LIM, R. H. F. *et al.* Fending off Delta – Hospital measures to reduce nosocomial transmission of COVID-19. *International Journal of Infectious Diseases* **117**, (2022).

157. McMichael, T. M. *et al.* COVID-19 in a Long-Term Care Facility - King County, Washington, February 27-March 9, 2020. *MMWR Morb Mortal Wkly Rep* (2020) doi:10.15585/mmwr.mm6912e1.

158. Centers for Disease Control and Prevention. COVID-19 Pandemic Planning Scenarios. *https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html* (2022).

159. Soetaert, K., Petzoldt, T. & Setzer, R. W. Package deSolve : Solving Initial Value Differential Equations in R. *J Stat Softw* **33**, (2010).

Appendices

Appendix A Chapter 2 Sensitivity Analysis

We conducted a sensitivity analysis to ensure that our qualitative findings were not an artifact of survivor bias related to increasing AIDS mortality positively associated with age. Specifically, we compared the odds ratio estimates for the wealth tertiles and gender obtained from the total population to those obtained from a sub-population of individuals 25 years of age and younger. Due to low cell counts in some country/year datasets resulting from restricting our analysis to only the youngest individuals, we were unable to stratify these data on urbanicity; as such these results were compared to those from an unstratified version of the complete dataset. If survivor bias was present, we expected to find lower odds of HIV in the total population compared to those 25 years of age and younger, as older individuals with HIV would likely have passed away by the time the survey was administered and thus would not be counted among those with HIV infection.

Results

Wealth

The results of our sensitivity analyses suggested that survivor bias was unlikely to impact our qualitative conclusions. We found that in many surveys, those 25 years old and under had lower or approximately equal odds of HIV infection compared to the total population. After controlling for gender, we found that in 23 out of 30 countries the odds in the total population of HIV infection in the middle wealth tertile were equal to or greater than that of the younger population.

Within the upper wealth tertile we found that in 26 out of 30 countries the total population were equal to or greater than those in the younger population.

Gender

Interestingly, in our sensitivity analyses, we did not find that gender had the same relationship to HIV trends as wealth did. Whereas over three-quarters of the countries surveyed showed a higher odds of HIV in the total population compared to those 25 and under, only 5 countries indicated higher odds of HIV in the total population compared to the younger population: Burkina Faso, Côte D'Ivoire, Liberia, Mali, and Senegal. These conflicting results may suggest an interaction between wealth, gender, and age, which warrants further investigation.

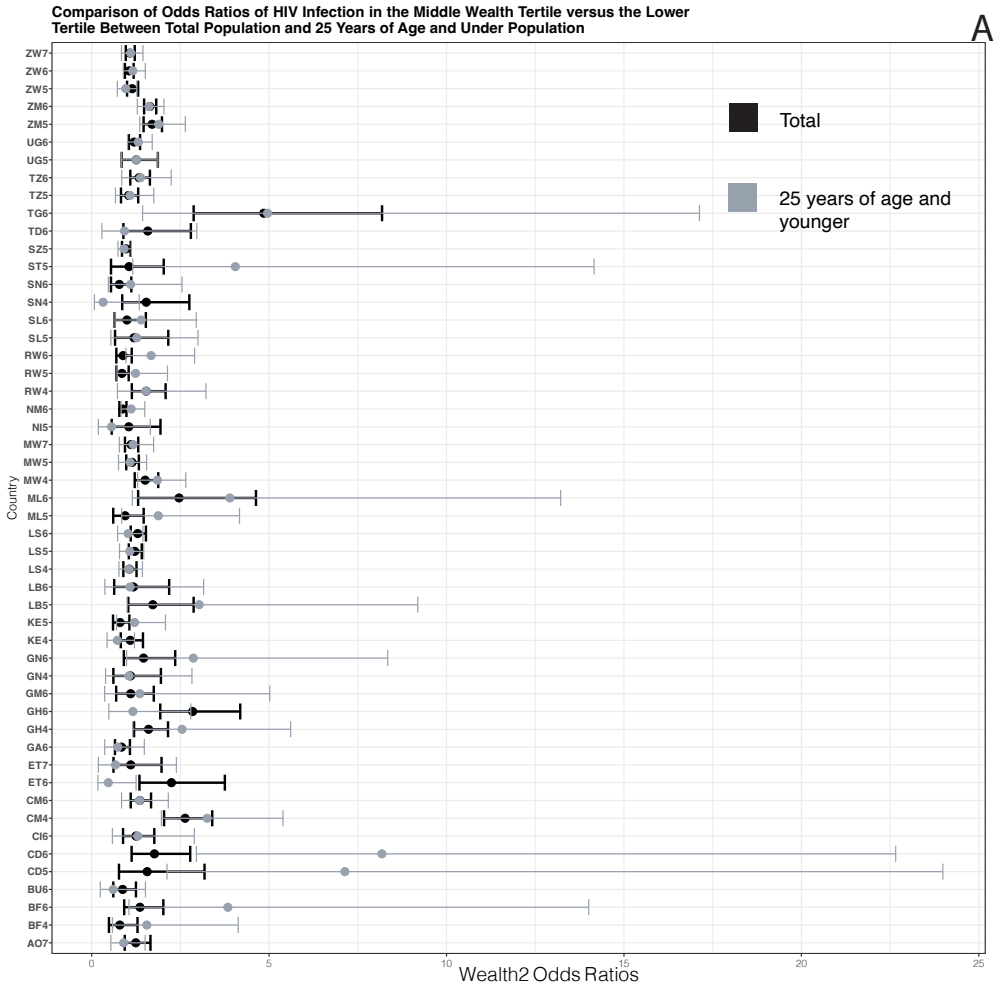


Figure A-1 Sensitivity Analysis Second Wealth Tertile, Urban v Rural. Black confidence intervals reflect the Odds Ratios and confidence Intervals for the unstratified population. Grey represent the Odds Ratios and Confidence Intervals for data solely from those 25 years old or younger.

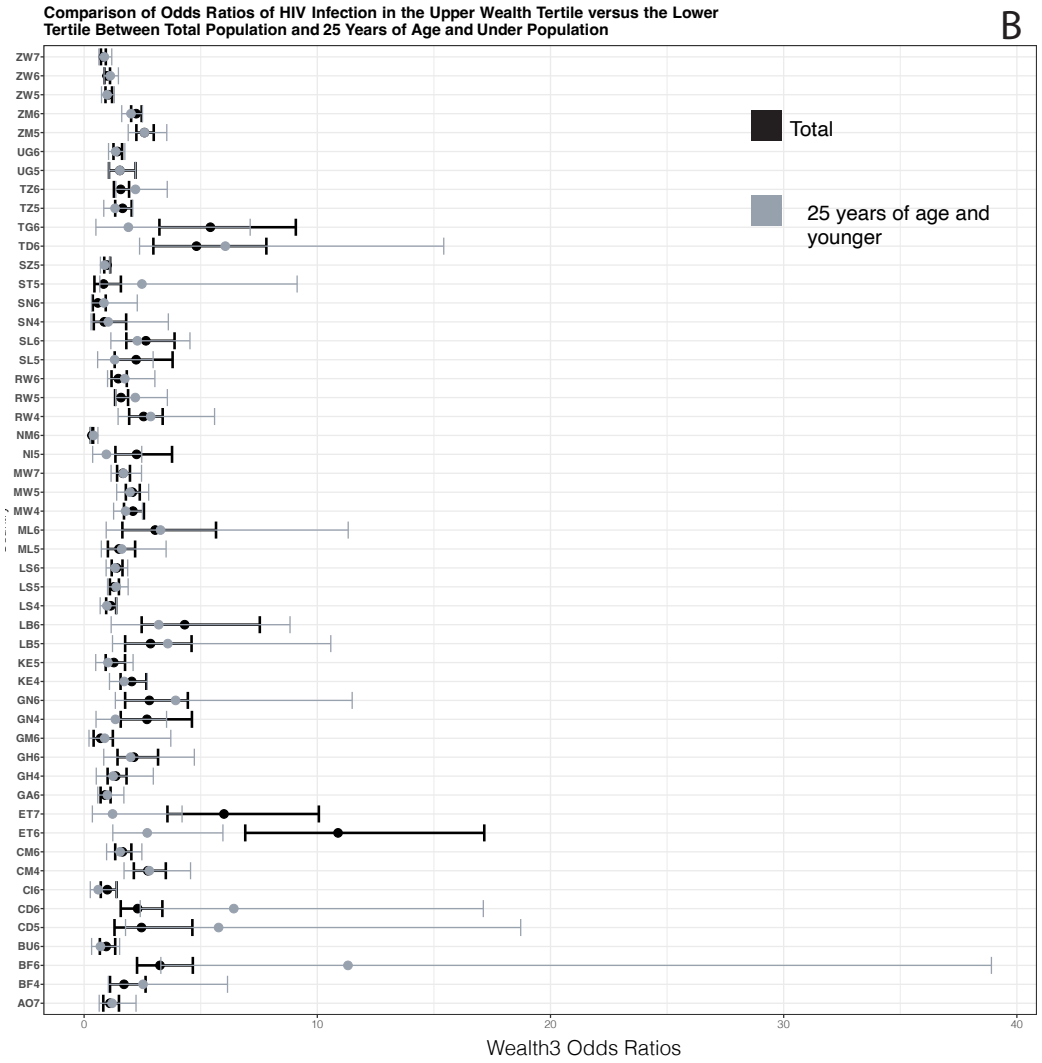


Figure A-2 Sensitivity Analysis Second Wealth Tertile, Urban v Rural. Black confidence intervals reflect the Odds Ratios and confidence Intervals for the unstratified population. Grey represent the Odds Ratios and Confidence Intervals for data solely from those 25 years old or younger.

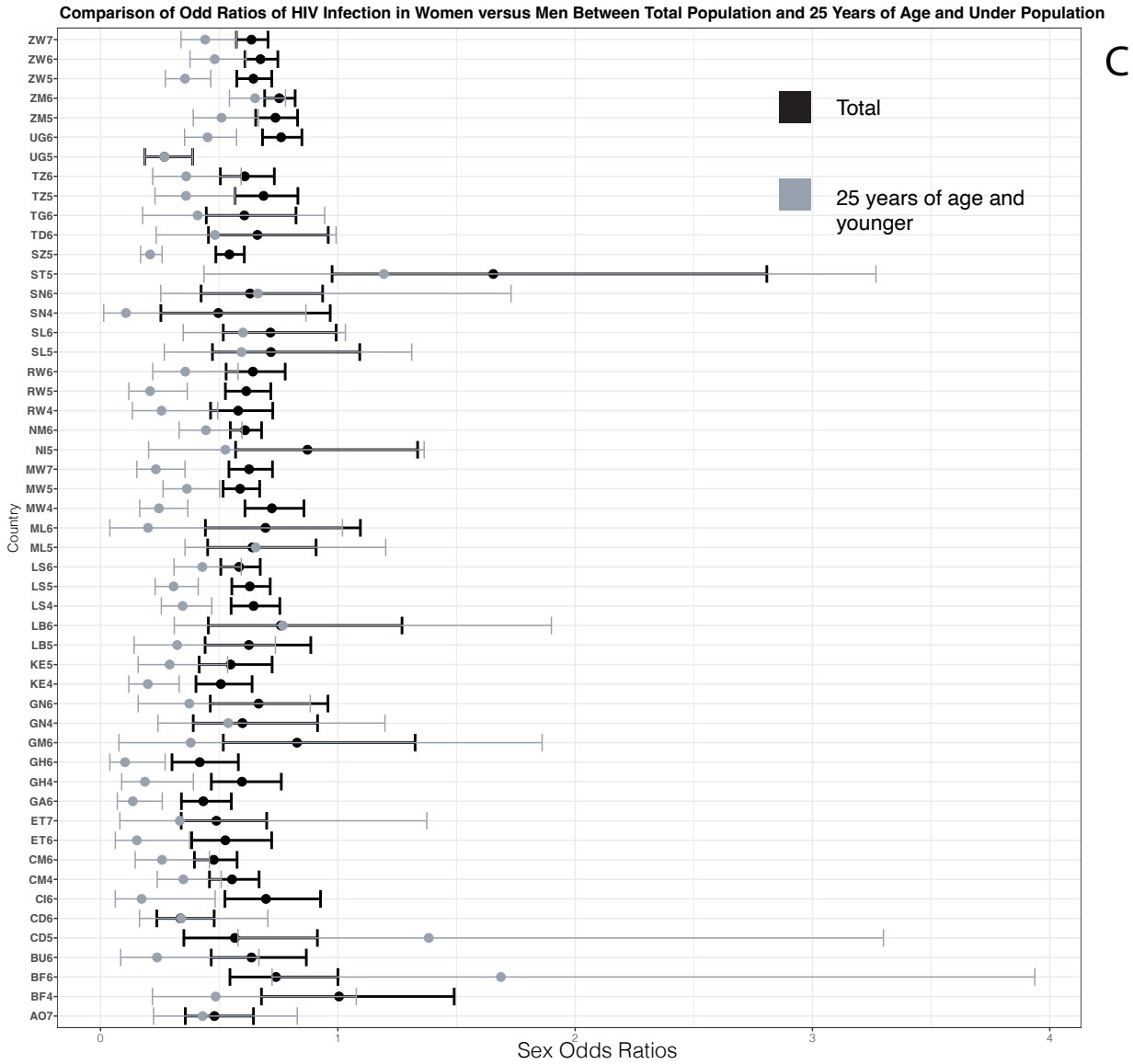


Figure A-3 Sensitivity Analysis for Gender, Urban v Rural. Black confidence intervals reflect the Odds Ratios and Confidence Intervals for the unstratified population. Grey represents the Odds Ratios and Confidence Intervals for data solely from those 25 years old or younger.

Appendix B Chapter 3 Test for Measurement Invariance Calculation

Log-likelihood value from unconstrained (simplified) model: -21106.791,

Unconstrained model free parameters: 387

Log-likelihood value from constrained (adjusted) model: -21048.227

Constrained model free parameters: 399

Log-likelihood ratio test: 2*(-21048.227 --21106.791)= 117.128

Degrees of Freedom = 399-387 = 12

Probability Chi-Squared (117.128,12)

P-Value from Chi-Squared =<0.005

Appendix C Chapter 3 Weighted Prevalence of LCA Variables by Age Group

Table C-1 Chapter 3 Weighted Prevalence of LCA Variables by Age Group

	ALL		18-29		30-39		40-49		50-59	
	N	%	N	%	N	%	N	%	N	%
Alcohol Consumption										
Never	19931487	0.12	7948514.02	0.17	4161938.39	0.11	4105972.47	0.1	3715062.43	0.09
Current	117144586	0.7	31580849.8	0.67	27286906.3	0.69	28443858.4	0.71	29832971.2	0.72
Non-Current	15674983	0.09	3542994.96	0.07	3320457.45	0.08	4524276.18	0.11	4287254.56	0.1
Did Not Respond	15403276	0.09	4231596.15	0.09	4741364.87	0.12	2744034.46	0.07	3686280.63	0.09
Cigarette Use										
Never	101587861	0.6	33169440.7	0.7	22874854.2	0.58	24791856.3	0.62	20751709.3	0.5
Current	34412802.3	0.2	8814157.92	0.19	8919332.13	0.23	7179673	0.18	9499639.29	0.23
Non-Current	32079657.9	0.19	5298092.22	0.11	7664733.23	0.19	7846612.25	0.2	11270220.2	0.27
Did Not Respond	74011.52	0	22264.07	0	51747.45	0	0	0	0	0
Marijuana Use										
Never	64258170	0.38	16781039	0.35	15283410.1	0.39	17980836.1	0.45	14212884.6	0.34
Current	19706896	0.12	9385459.69	0.2	4361032.28	0.11	2425468.48	0.06	3534935.8	0.09
Non-Current	66474119	0.4	16012213.6	0.34	14824979.2	0.38	16152609.4	0.41	19484316.7	0.47
Did Not Respond	17715147	0.11	5125242.75	0.11	5041245.44	0.13	3259227.52	0.08	4289431.72	0.1
Oral Sex Partners Past Year										
0	68324383.7	0.41	17154541	0.36	14453981.9	0.37	14508126.6	0.36	22207734.2	0.53
1	84933157	0.51	22786314.3	0.48	21947922.6	0.56	22481148.5	0.56	17717771.6	0.43
2 to 4	12160399.6	0.07	6022288.74	0.13	2649921.6	0.07	2261666.56	0.06	1226522.64	0.03
5 to 9	1850797.61	0.01	910481.08	0.02	338954.91	0.01	319213.75	0.01	282147.87	0.01
10 +	797880.54	0	430329.81	0.01	72022.98	0	247986.07	0.01	47541.69	0
Did Not Respond	87713.83	0	0	0	47862.98	0	0	0	39850.84	0
Vaginal Sex Partners Past Year										
0	43758363.5	0.26	13474223.1	0.28	8489176.61	0.21	7659676.99	0.19	14135286.8	0.34
1	104128481	0.62	24127452	0.51	26697605.7	0.68	28182821.9	0.71	25120601.1	0.61
2 to 4	15571621	0.09	7056292.96	0.15	3619499.65	0.09	3303156.08	0.08	1592672.33	0.04
5 to 9	2876797.4	0.02	1831818.78	0.04	377240.86	0.01	239699.7	0.01	428038.08	0.01
10 +	1631995.1	0.01	705288.01	0.01	327144.19	0.01	354592.5	0.01	244970.43	0.01
Did Not Respond	187074.4	0	108880.04	0	0	0	78194.36	0	0	0

Appendix D Maximum Infected in Counts, by Unit, Across Betas

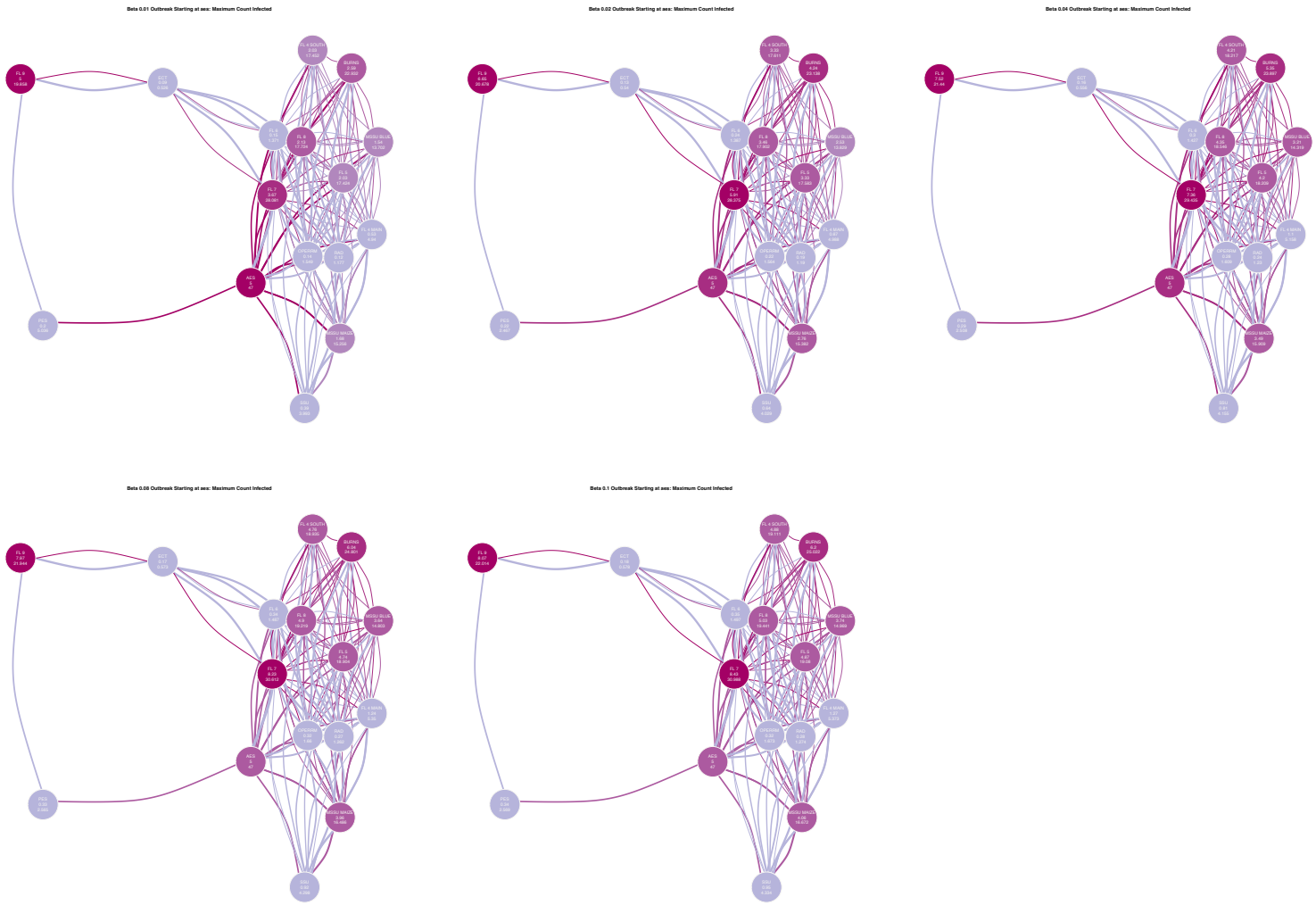


Figure D-1 Maximum Infected in Counts, by Unit, Across Betas, with Infection Starting in AES

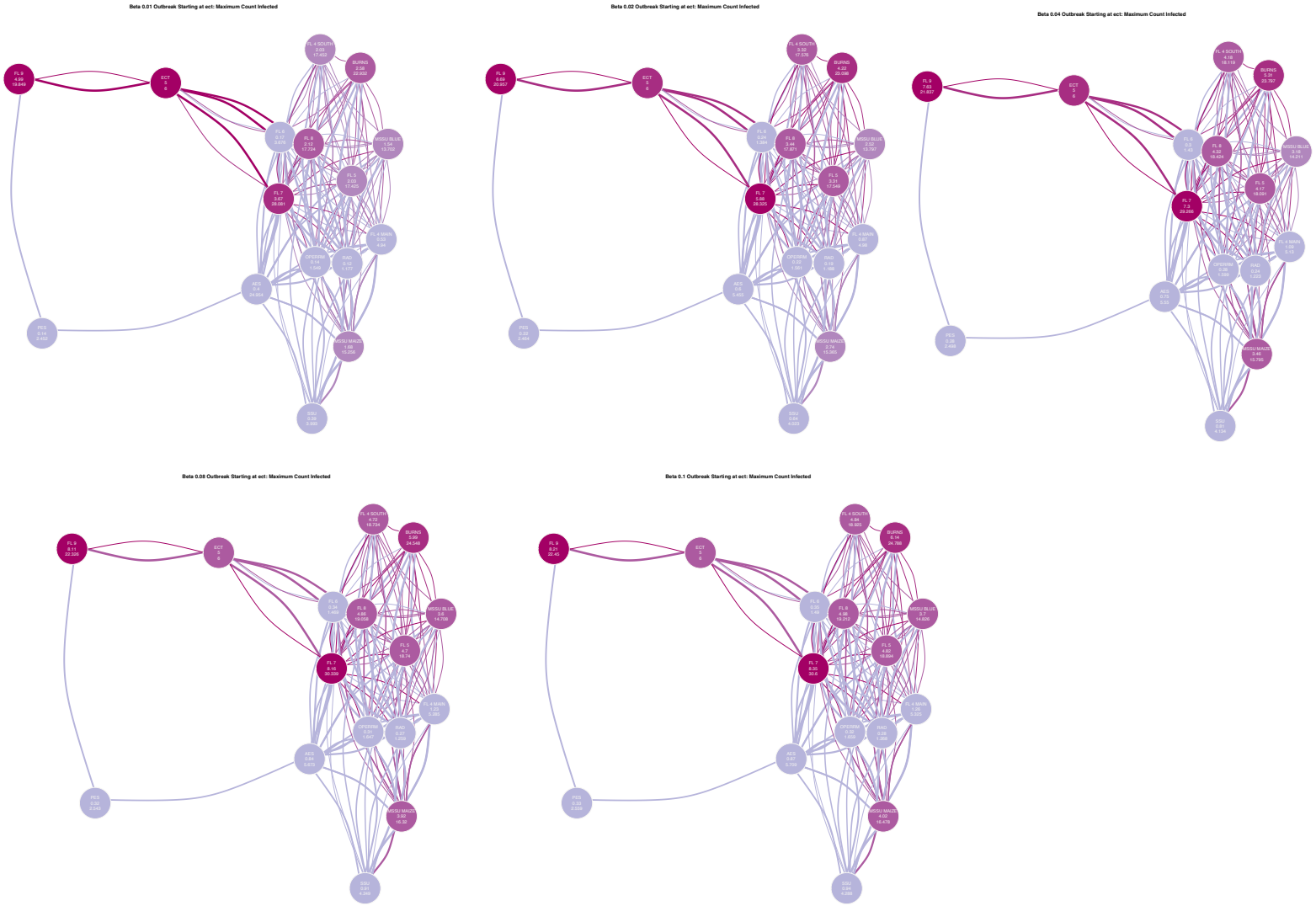


Figure D-2 Maximum Infected in Counts, by Unit, Across Betas, with Infection Starting in ECT

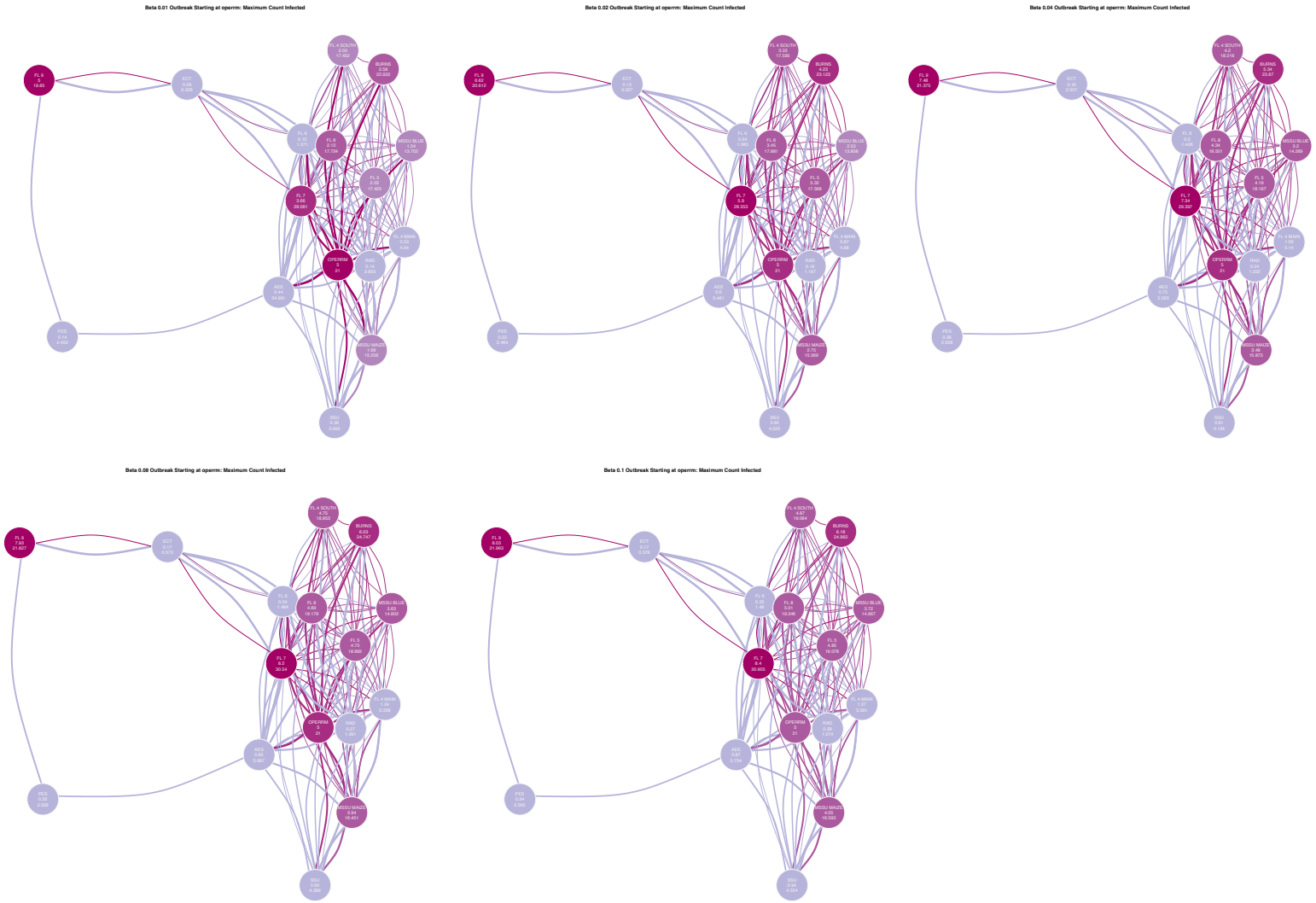


Figure D-3 Maximum Infected in Counts, by Unit, Across Betas, with Infection Starting in OPERRM