Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

ESCOLA NACIONAL DE SAÚDE PÚBLICA
SERGIO AROUCA
ENSP

Daniel Camargo Stokes

Social network analysis of body mass indices in a cohort of Brazilian professionals (ELSA-Brasil)

Rio de Janeiro
2016

Daniel Camargo Stokes

Social network analysis of body mass indices in a cohort of Brazilian professionals (ELSA-Brasil)

Dissertação apresentada ao Programa de Pós-graduação em Epidemiologia em saúde pública da Escola Nacional de Saúde Pública Sérgio Arouca, na Fundação Oswaldo Cruz, como requisito parcial para obtenção do título de Mestre em Ciências na área de Epidemiologia em saúde pública.

Orientador: Prof. Dr. Leonardo Soares Bastos
Coorientador: Prof.ª Dr.ª Marilia Sá Carvalho

Rio de Janeiro

2016

Catalogação na fonte
Instituto de Comunicação e Informação Científica e Tecnológica
Biblioteca de Saúde Pública

S874s Stokes, Daniel Camargo
Social network analysis of body mass indices in a cohort of Brazilian professionals (ELSA-Brasil). / Daniel Camargo Stokes. -- 2016.
93 f. : il. ; tab. ; graf.

Orientador: Leonardo Soares Bastos
Marilia Sá Carvalho
Dissertação (Mestrado) – Escola Nacional de Saúde Pública Sergio Arouca, Rio de Janeiro, 2016.
1. Rede Social. 2. Índice de Massa Corporal. 3. Obesidade. 4. Prevalência. 5. Saúde do Trabalhador. 6. Saúde do Adulto. 7. Estudos Longitudinais. 8. Estudos de Coortes. 9. Brasil. I. Título.

CDD – 22.ed. – 613.04340981

Daniel Camargo Stokes

Social network analysis of body mass indices in a cohort of Brazilian professionals (ELSA-Brasil)

Dissertação apresentada ao Programa de Pós-graduação em Epidemiologia em saúde pública da Escola Nacional de Saúde Pública Sérgio Arouca, na Fundação Oswaldo Cruz, como requisito parcial para obtenção do título de Mestre em Ciências na área de Epidemiologia em saúde pública.

Aprovada em: 3 de março de 2016

Banca examinadora

Prof[a]./Dr[a]. Claúdia Medina Coeli, UFRJ

Prof./Dr. Marcelo Ferreira da Costa Gomes, PROCC/Fiocruz

Prof./Dr. Leonardo Soares Bastos, PROCC/ENSP

Rio de Janeiro

2016

Table of Contents

Index of Figures

Index of Tables

Acronyms and Abbreviations

BMI	Body mass index*
ELSA	*Estudo Longitudinal da Saúde do Adulto* (Longitudinal Study of Adult Health)
ELSA-RioSC	Sociocentric graph of Wave 2 participants actively employed by Fiocruz at the time of data collection
ERGMs	Exponential random graph models
Fiocruz	*Fundação Oswaldo Cruz* (Oswaldo Cruz Foundation)
NCDs	Non-communicable Diseases
SES	Socioeconomic Status
Variable.1	Variable collected in ELSA Wave 1, e.g. bmi.1
Variable.2	Variable collected in ELSA Wave 2, e.g. bmi.2

*Calculated according to Equation 1:

$$BMI = \frac{Weight(kg)}{Height(meters)^2} \quad [1]$$

I. Abstract/Resumo

Resumo

Objetivo: Mais da metade da população brasileira encontra-se acima do peso normal, e as prevalências de obesidade e de sobrepeso continuam subindo, especialmente nas populações urbanas de baixo nível socioeconômico. Embora o governo tenha começado a montar uma resposta, incluindo programas especificamente destinados à melhoria da saúde dos funcionários nos ambientes de trabalho, o papel da dinâmica social entre funcionários na ganha e perda de peso ainda é pouco entendido. Através de uma análise da distribuição dos índices de massa corporal (IMCs) em uma rede social de funcionários públicos, este estudo tem como objetivo iniciar uma discussão sobre como as amizades podem afetar os pesos dos trabalhadores de meia idade no Brasil.

Métodos: Entre 2010 e 2012, 1.521 funcionários públicos, empregados pela Fundação Oswaldo Cruz (Fiocruz), com base no Rio de Janeiro, foram convidados a nomear, como parte do questionário da segunda onda do Estudo Longitudinal da Saúde do Adulto (ELSA-Brasil), os seus cinco amigos mais próximos no trabalho. Aonde possível, nomes citados foram ligados a nomes oficiais no diretório de funcionários da Fiocruz com a ajuda de um programa de relacionamento probabilístico de registros. As amizades entre os participantes do ELSA definiram as conexões no ELSA-RioSC, uma rede social sócio-cêntrica dos participantes da segunda onda do estudo de coorte. As relações entre o IMC de um indivíduo, a sua posição social no ELSA-RioSC, e os IMCs dos seus amigos foram avaliadas através da modelagem Gaussiana latente, controlando por idade, renda familiar mensal per capita, e nível de educação, com dados estratificados por sexo. Um modelo de grafo aleatório exponencial (ERGM, pelo termo em inglês, “exponential random graph model”) foi desenvolvido para testar se uma semelhança no IMC de dois indivíduos afeta a probabilidade destes indivíduos formarem uma amizade, controlando por fatores exógenos e endógenos.

Resultados:ELSA-RioSC foi composta por 1.973 amizades, e 332 dos 1.521 membros do grafo foram isolados (não conectados por uma amizade a nenhum outro indivíduo). Entre os membros, 69,4% estavam acima do peso normal (IMC$\geq$25kg/m^2) com 28,2% considerados obesos (IMC$\geq$30kg/m^2). Nos dois a seis anos entre a primeira e a segunda onda do ELSA, os IMCs dos participantes geralmente aumentaram (em média por 0,691% anualmente). Os modelos Gaussianos latentes mostraram que, para ambos homens e mulheres, indivíduos com nenhum amigo ou poucos amigos na média tinham maior IMC na Onda 2 e maior ganho de peso percentual entre Onda 1 e Onda 2. O IMC seccional das mulheres mostrou relação inversa com a educação e com a renda familiar per capita – educação mais básica e renda menor corresponderam a um IMC maior – enquanto o IMC seccional dos homens mostrou uma relação mais complicada com as variáveis de controle, mas uma relação direta com a média dos IMCs dos amigos. A mudança percentual no IMC dos homens entre as Ondas 1 e 2 também mostrou relação direta com a mudança percentual média no IMC dos seus amigos, com uma associação especialmente significativa para perdas de peso, mas esta relação não foi observada para as mulheres do estudo. Os resultados do ERGM indicaram que indivíduos com IMCs semelhantes não mostraram significativamente maior propensão a serem amigos do que indivíduos com uma diferença em IMC superior a 4 kg/m^2, mesmo antes de controlar por fatores exógenos e endógenos. Os fatores mais significativos na escolha de amigos foram: compartilhar uma unidade de trabalho e retribuir uma amizade unidirecional.

Conclusão: Os resultados mostram que os IMCs de homens, mas não aqueles de mulheres, estão associados aos IMCs dos seus amigso no trabalho. Isso indica que homens modelam os seus hábitos de comer e fazer exercício nos hábitos dos seus amigos no trabalho, selecionam amigos com hábitos similares aos seus, ou tendem a compartir ambientes de nutrição e exercício com seus amigos. Dados longitudinais de amizade e de comportamento de consumo e gasto de energia ajudariam a elucidar alguns dos mecanismos que levaram às associações observadas.

Palavras-chave: Rede Social; Índice de Massa Corporal; Obesidade; Saúde do Trabalhador

Abstract

Purpose: The prevalences of obesity and overweight in Brazil continue to rise, with more than half of the population now considered above normal weight, and the burden of the epidemic is increasingly shifting to the urban poor. While the government has begun to mount a response, including workplace-specific programs aimed at helping employees to live healthier lives, research regarding the role of workplace social dynamics in the processes of gaining and losing weight is sparse. Through analysis of the distribution of body mass indices (BMIs) in a Rio de Janeiro civil servant at-work friendship network, this study aims to begin a discussion regarding the roles friends may play in the changing weights of working adults in Brazil.

Methods: Between 2010 and 2012, 1,521 civil servants, employed by the Oswaldo Cruz Foundation in Rio de Janeiro, were asked to name their five closest friends at work as part of the questionnaire for the second wave of the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). Cited names were probabilistically linked to official employee names, and the friendships between ELSA-Brasil participants became the edge-list for a sociocentric network graph, here-in referred to as ELSA-RioSC. The relationships between an individual's BMI, his or her position in ELSA-RioSC, and the BMIs of his or her friends were assessed through latent Gaussian modeling, controlling for age, monthly per capita family income, and education level, and with data stratified by sex. An exponential random graph model (ERGM) was developed to test whether or not friends tend to share similar BMIs, and if they do, whether or not that relationship holds after controlling for exogenous and endogenous factors.

Results: ELSA-RioSC was comprised of 1,973 edges, and 332 of the 1,521 vertices were isolates. 69.4% of ELSA-RioSC members were above normal weight (BMI≥25kg/m^2), with 28.2% falling within the obese category (BMI≥30 kg/m^2), and in the 2-6 years between the first and second waves of data collection, individuals' BMIs generally increased, on average by 0.691% per year. Latent Gaussian models for both sectional Wave 2 BMI and annual percent change in BMI between Waves 1 and 2 indicated that, for both men and women, individuals with no or few friends in the network were likely to have higher BMIs and to gain more weight than individuals with several friends in the network. Women's sectional BMIs showed an inverse relationship with education and income level, with lower levels of education and income corresponding to greater BMIs, while men's sectional BMIs showed a less clear dependence on control variables but a direct relationship with friends' BMIs. Men's average annual percent changes in BMI were also directly associated with friends' average annual percent changes, especially for loss of weight, but the same relationship did not hold for women. ERGM results indicated that individuals with similar BMIs were not significantly more likely to be friends than individuals with a BMI difference greater than 4 kg/m^2, even before controlling for exogenous and endogenous variables, and the most significant predictors of friendship nomination were a shared work department and reciprocity.

Conclusions: The results indicate that the BMIs of men, but not those of women, are associated with the BMIs of their friends at work. This implies that men model their exercise and eating on friends, select friends based on shared exercise and eating habits, or tend to share similar nutrition and exercise environments with friends. The collection of longitudinal friendship data and information regarding specific energy consumption and expenditure behaviors would help to elucidate some of the mechanisms underlying the observed relationships.

Keywords: Social Networking; Body Mass Index; Obesity; Occupational Health

II. Objectives

General objective:

Through analyzing the associations between the body mass indices of Brazilian professionals, their social positions in an at-work friendship network, and the body mass indices of their friends at work, this study aims to identify some of the social aspects of the obesity epidemic in Brazil. The results of this study will contribute to an understanding of how friendship and social life interact with obesogenic environments and behaviors in affecting individuals' and communities' BMIs, and these results have the potential to inform future institutional efforts to address obesity in Brazil and elsewhere.

Specific objectives:

(1) To construct a sociocentric social network graph based on free-recall data collected through the ELSA-Brasil cohort study at the Fiocruz Institute in Rio de Janeiro and to develop a protocol for the construction of similar graphs from the data collected at other institutions participating in ELSA-Brasil;

(2) To compare the constructed graph to random graph models and to obtain descriptive statistics of the graph, which will allow for a better understanding of the social dynamics at Fiocruz and will serve as points of comparison for future social network analyses in Brazil;

(3) To investigate associations between an individual's BMI and his or her social position – as indicated by graph vertex characteristics – and his or her friends' BMIs, taking friendship directionality into consideration;

(4) To assess which factors are most predictive of friendship formation in ELSA-RioSC using Exponential Random Graph Modeling (ERGM);

(5) To consider the implications of the results of these analyses for future workplace-based efforts to control the obesity epidemic in Brazil.

III. Introduction

i. The global obesity epidemic

The universally growing prevalence of obesity constitutes a troubling modern epidemic. Since 1980, global prevalence has nearly doubled and risen in almost every country.[1] Defining overweight by a body mass index (BMI, Equation [1]) greater than or equal to 25 kg/m^2 and obesity by a BMI greater than or equal to 30 kg/m^2 , in 2014, 39% of adults globally were overweight, and 13% were obese.[1] Obesity and overweight are risk factors for many of the non-communicable diseases (NCDs) contributing most to the global disease burden, including diabetes, hypertension and coronary heart disease.[2,3]

NCDs accounted for 68% of deaths worldwide in 2012,[4] and approximately 75% of those deaths came from low- and middle-income countries,[5] where attention and resources have traditionally been focused on infectious diseases. In Brazil between 1930 and 2007, the proportion of deaths attributable to infectious diseases dropped from 46% to 10%, and, by 2007, NCDs accounted for 72% of all deaths.[6] With the increase in NCD prevalence – and the significant role that the obesity epidemic plays in the disease burden of many NCDs – obesity and overweight were estimated in 2010 to cause at least 3.4 million adult deaths each year.[7] Most adults live in countries where these deaths now outnumber deaths related to being underweight.[1] So far, no country has succeeded in reversing upward trends in mean BMI.[8]

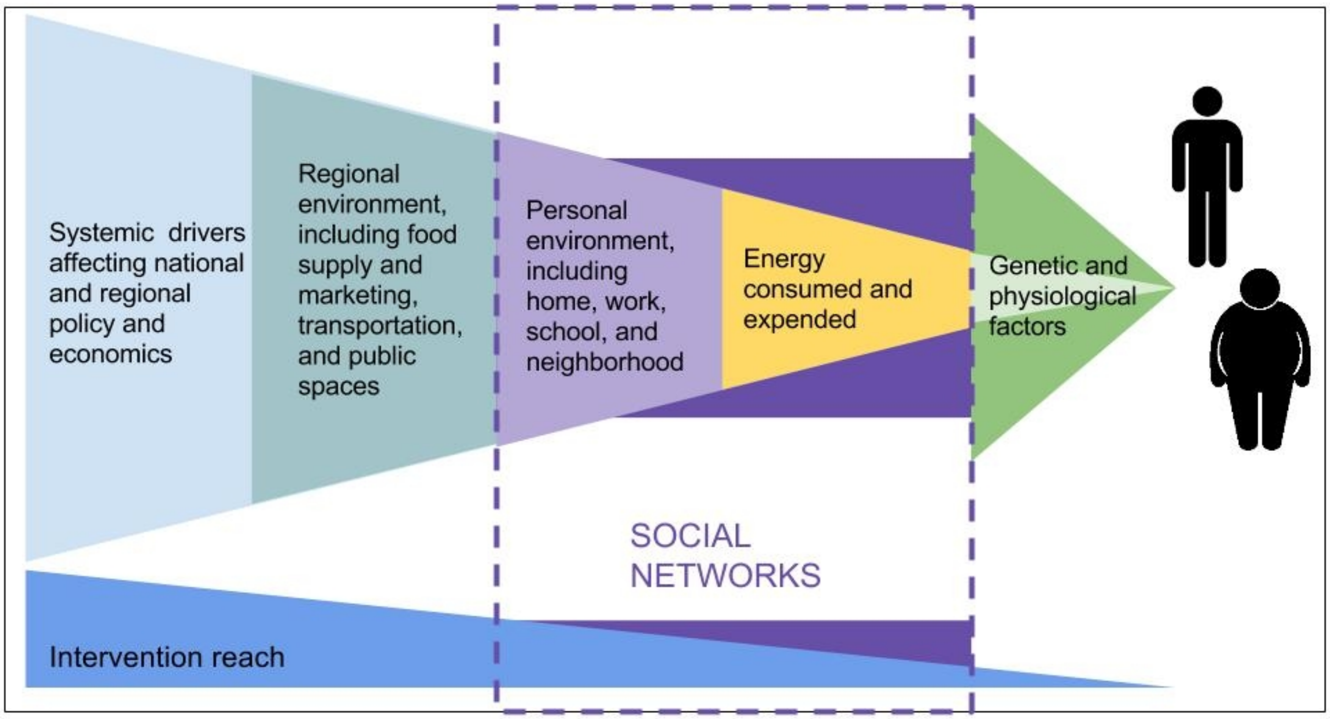
Fundamentally, an increase in BMI at the individual level results from an imbalance in personal energy consumption and energy expenditure: if more energy is consumed than is spent, an individual will gain weight.[1,9] But the factors influencing the availability and desirability of high-energy foods and the necessity or feasibility of physical exercise in labor and leisure extend well beyond the level of the individual. Global drivers interact with regional, local, and personal factors in determining individual risk of becoming overweight or obese.[8]

The near-universal increase in national and regional mean BMIs suggests that global drivers play an important role in the population-level shifts in BMI distributions. The obesity epidemic began in the 1970s and 1980s in high-income countries.[8] The beginning of the epidemic, and its subsequent spread to low- and middle-income countries, has coincided with an expanding transition in the global food landscape and a shift in patterns of physical activity; "development" has come with a higher-energy diet and lower-energy expenditure.[10] The industrialization of food production and consumption has greatly lowered the cost and increased the availability of energy-rich foods, and technological advances and urbanization have led to more sedentary forms of labor and leisure.[10,11] While humans have evolved important physiological defenses against weight-loss in nutrient-poor environments, we are ill-adapted as a species to avoiding weight gain in nutrient-rich environments. The obesity epidemic can thus be understood as a consequence of the juxtaposition between the environment in which human physiology evolved and the environment humans now face.[12]

Many studies have indicated that the built environment – including transportation, food availability, and access to public recreation spaces, among other factors – might contribute to the heterogeneity of the epidemic. For example, several studies have shown that neighborhoods with a greater number of large supermarkets (and therefore more diversity in cheap food options) have lower levels of obesity than areas with a greater number of small markets and fast food chains.[13,14] Other studies have suggested that more pedestrian friendly neighborhoods and neighborhoods with more access to physical activity facilities have lower prevalences of obesity.[15,16] In all of these associations, it is important to note both that a complicated relationship exists between the built environment and potentially confounding socioeconomic distributions, and that the direction of a cause-effect relationship is not clear. Beyond global drivers and the built environment, the risk of obesity at the individual level can depend on education; family and community eating and exercise habits; and genetic predispositions associated with race or family history.[17,18]

Local manifestations of the global obesity epidemic are thus unique products of the interplay between systemic, environmental and individual risk factors. This dynamic is summarized in Figure 1, adapted from Swinburn, et al.[8]. The figure shows how factors at multiple levels combine to produce changes in the BMI of an individual: system-level factors are at the most macroscopic scale, affecting populations, while genetic and physiological factors are at the most microscopic, affecting how one person responds to his or her energy intake and expenditure. The number of people impacted by any effort to control the obesity epidemic depends on the level at which change is effected, as indicated by the blue “intervention reach” triangle at the base of Figure 1; bariatric surgery reduces the BMI of a single person, while restrictions on advertising for ultra-processed foods could shift the average BMI of entire populations. As Geoffrey Rose has argued, systemic and environmental changes, while often more complicated to bring, represent a much more efficient and effective means by which to address NCDs than treatment strategies restricted to high-risk patients.[19]

Figure 1 – Simplified model of the factors affecting an individual's BMI



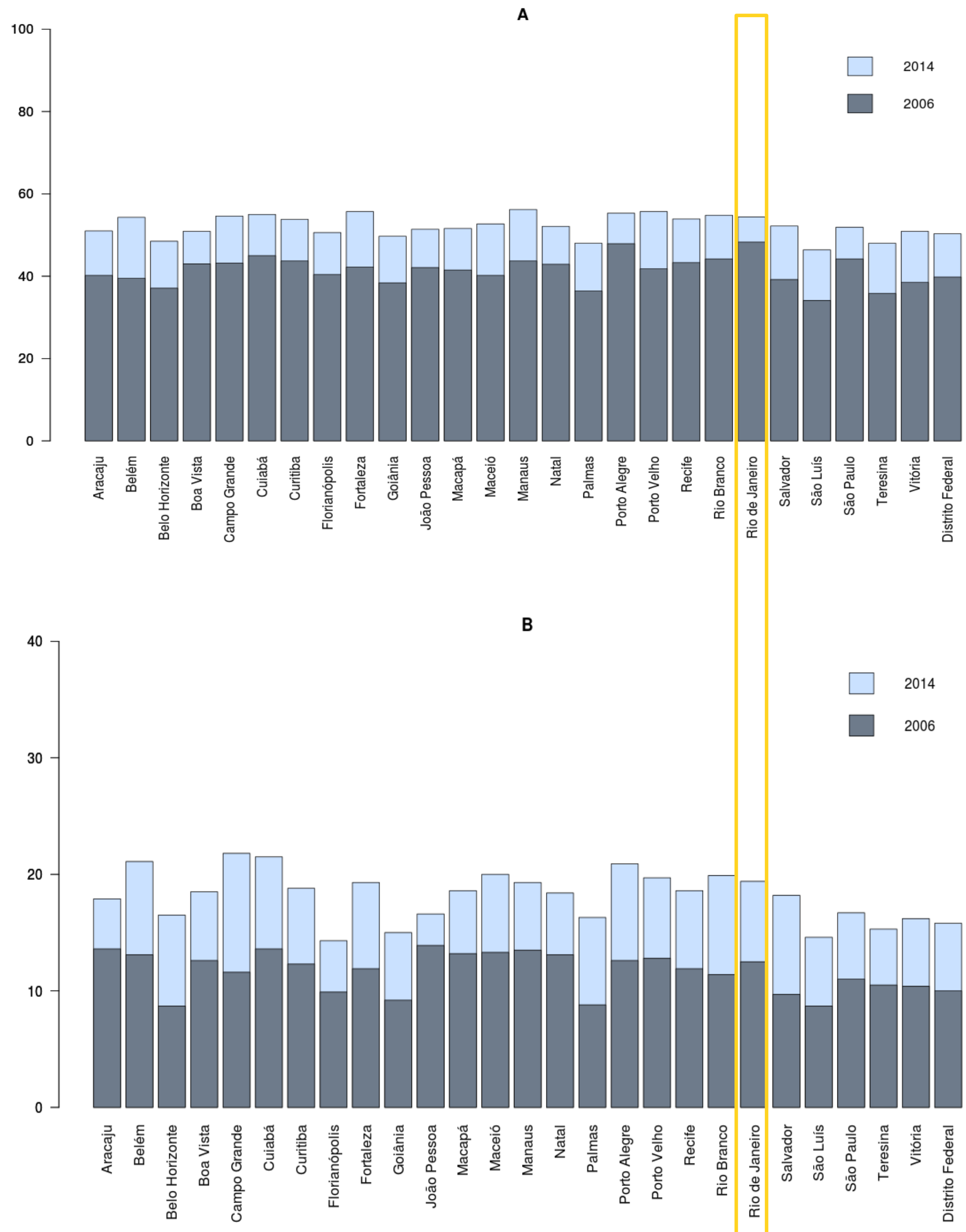
Global food systems affect local environments, and an individual's response to those environments is in turn affected by social factors, like upbringing, education and income level. The impact of energy consumption and expenditure on body type depends on genetic and physiological factors. Higher-level interventions impact more people, as symbolized by the "Intervention reach"triangle, but they are generally more difficult to enact. If energy consumption and expenditure are affected by interpersonal processes, this has important implications for understanding both the obesity epidemic and potential approaches to its control. This is shown in the figure by the purple adjustments in the "Social Networks" region. The figure was adapted from Swinburn, et al.[8]

ii. Overweight and obesity in Brazil

While obesity is still most prevalent in high-income countries, its prevalence in low- and middle-income countries is quickly rising, and the social distribution of obesity in these countries is often significantly different than in more developed ones. Based on data from 2007, mean national BMI in low-income countries tends to increase linearly as per capita Gross Domestic Product (GDP) increases, but for countries with per-capita annual GDPs above US$3000 dollars per year, mean BMI stabilizes around 25-26 kg/m^2.[20] In Brazil, where the 2014 per capita GDP was US$11,384,[21] the age-adjusted mean BMI for both males and females was 26 kg/m^2.[22] Data from telephone surveys conducted in Brazilian state capitals over the eight years between 2006 and 2014 indicates that the prevalence of obesity and overweight in adults (18 or older) rose in every city included in the study, as shown in Figure 2.[23,24] The number of times obesity was listed as one of the causes of death on a death certificate in the public health system more than doubled from 5.4 per million inhabitants in 2001 to

11.9 in 2011, and these figures ignore deaths from diseases for which obesity is a risk factor but not a direct cause, like diabetes and coronary heart disease.[25]

Figure 2 – Prevalences of obesity and overweight in Brazilian state capitals in 2006 and 2014



The prevalence of overweight (top) and obesity (bottom) in Brazilian capitals in 2006 and 2014. In all capitals, obesity and overweight became more prevalent in the eight year period between the two rounds of data collection. Rio de Janeiro data, of principal interest to this study, is highlighted in yellow.

The social distribution of obesity follows some predictable patterns in many low- and middle-income countries. In high-income countries, obesity is generally most prevalent in groups of low socioeconomic status (SES). In low- and middle-income countries, obesity has traditionally been a problem of high SES groups,[8] but in many of those countries, the burden of the epidemic is shifting to lower SES groups, especially in urban areas.[26,27] In Brazil between 1975 and 1985, the rate of obesity increased among all SES groups, but the increase was greater among those of low SES, and between 1985 and 2003, the obesity rate dropped among high SES women and remained stable in high SES men while continuing to rise among low SES individuals of both sexes.[28]

As in other countries, the obesity epidemic in Brazil has resulted from community-level and individual-level responses to increasingly obesogenic environments. Over the past few decades, Brazil, like many developing countries,[10,29] has experienced a “nutrition transition”: a shift away from traditional food systems and towards meals away from home,[30] ultra-processed foods[31] and foods rich in fats and sugars.[32,33] Transnational food corporations – almost all based in Europe and the United States – have expanded aggressively into developing countries like Brazil through mass-production and mass-marketing, greatly increasing the availability and visibility of low-cost, ultra-processed foods .[30,34] Ultra-processed foods are made from extracted and purified whole food ingredients and additives, and packaged as durable, convenient products designed to be visually appealing, highly palatable, and deceptively satisfying. The risk of these products, therefore, goes beyond higher levels of sugar and fat: they are also cheap, attractive, and aggressively marketed.[30,31,34]

Different socioeconomic strata experience the integration of the global food market unequally: while dietary choices may expand for wealthier populations, the diets of lower socioeconomic groups in developing countries across the world are likely to converge towards inexpensive, ultra-processed, obesogenic foods.[35] In high-income countries, dietary patterns have developed with and around industrialization, but in countries where this has not happened, trade liberalization has often allowed an influx of ultra-processed foods to abruptly displace traditional food systems. In Brazil, traditional food systems and family meals have proven relatively resistant to changes in the nutrition landscape,[30] but even so, consumption of traditional foods like rice and beans is declining in metropolitan areas, while consumption of ultra-processed foods is increasing across all socioeconomic groups.[31,33] In a 2008-2009 national survey, it was found that 52% of energy came from processed foods, defined as foods high in solid fats, *trans* fats, and added sugars.[32] Comparing the rates of obesity in low- and middle-income countries and regions where traditional food systems still dominate with those where they do not, it seems reasonable to conclude that the internationalization of ultra-processed foods and the displacement of traditional diets, especially in low-income urban populations, is a major driver of both the nutrition transition and the obesity epidemic in the developing world.[31,36,37]

The nutrition transition has been accompanied by urbanization and industrialization in Brazil, and as such, individuals are not only consuming more energy, but also generally expending less.[38] Estimates based on national and regional data – and, where such data was unavailable, international data for similar income groups – indicate that from 2002 to 2007, sedentary time increased and physical activity decreased, and these trends were predicted to continue into at least 2030.[39] Energy expenditure likely dropped even more significantly before 2002, when rates of rural-to-urban migration were especially high,[40] but no national data on physical activity exists for that time period.[6] While rates of "active leisure" physical activity have increased slightly in the time-period for which data is available (albeit only from São Paulo), that increase has been far out-weighed by estimated drops in physical activity in the home, at work, and in transit.[39] Longitudinal trends therefore indicate that energy expenditure continues to fall while energy consumption continues to rise. Without intervention, increasing prevalences of obesity and overweight, especially among the country's urban poor, show no signs of slowing.

iii. Efforts to address the obesity epidemic in Brazil

In recent years, the Brazilian government has been responsive to issues of energy consumption and expenditure. In 2010, Article 6 of the constitution was amended to include *alimentação* (or nourishment) on the list of each citizen's "social rights", a list that, among other items, includes a right to health. The constitution also declares that it is the duty of family, society, and the State to assure the health and nourishment of children and adolescents.[41] A document published by the Ministry of Social Services and the Battle Against Hunger elaborates, defining the right to *alimentação* as a right to "healthy, accessible, high-quality [food] in sufficient quantities and without interruption."[42]

Governmental programs have been introduced and adapted over the years to promote and protect these constitutional rights. *Fome Zero* (Zero Hunger), which began in 2003 under the presidency of Luiz Inácio da Silva and is now associated with the *Programa Bolsa Família* network of conditional cash transfer programs, aims to combat poverty and malnutrition. Most studies indicate that the program has resulted in increased family food security and improved nutrition for children among the poorest socioeconomic groups in Brazil, but some have also recognized that families covered by the program are consuming foods with higher caloric density and lower nutritious value.[43] The program has emphasized food availability rather than access to healthy foods, which ignores the present reality: even among the low-income populations that the social welfare programs target, obesity and overweight are increasingly more prevalent than underweight and hunger, especially among adults,[44] and as purchasing power in poor families increases, there is a disproportionately large increase in consumption of unhealthy, processed foods.[45]

The *National Policy on Food and Nutrition* (PNAN) was published in 2011 in response to the nutrition transition and the increasing prevalences of overweight and obesity. The document appropriately shifts the discussion away from increased consumption and towards healthier consumption. It also recognizes the role that food plays in the social context of Brazil: “food expresses social relations, values, and the histories of Brazilian individuals and peoples, and it carries direct implications for health and quality of life.”[46] The document further notes the government's responsibility to promote and prioritize national production of ecologically sustainable, nutritious foods over the capitalistic demands of marketplaces.[46] The *Guide to Nourishment for the Brazilian Populace*, published in 2014, stresses that minimally processed and unprocessed foods should be prioritized in the family diet.[47] These documents set important precedents for government approaches to addressing the role of changing diets in the epidemic of overweight and obesity, but how those approaches will be developed and whether or not they will be effective remains to be seen. Inês Rugani Ribeiro de Castro, a researcher at the State University of Rio de Janeiro's Nutrition Institute, stresses that an effective response must recognize that food is, and should be, social – “individual choices, while indispensable, are not enough to guarantee healthy and sustainable eating habits in a collective environment” - and that food systems in Brazil will not change spontaneously; government policies must be enacted to encourage the production and consumption of healthy foods.[48]

The government has also recognized the role of increasingly sedentary lifestyles in the prevalences of obesity and overweight. The 2006 *National Policy on Health Promotion*[49] and the 2014 *Intersectoral Strategy for Obesity Prevention and Control*[50] both highlight the importance of creating urban spaces designed to encourage physical activity and of organizing educational campaigns to teach people about the benefits of active lifestyles. The latter document suggests that institutional environments like schools, workplaces, and health outposts could play central roles in promoting and creating space and time for physical activities, and the “Healthy Weight Program” - another government initiative – specifically promotes weight monitoring and control in the work environment.

51

These governmental public health documents suggest a nuanced understanding of the complex interplay between systemic, environmental, and individual factors that has contributed to the rapidly increasing prevalence of obesity and overweight in Brazil; they recognize that the profile of the epidemic changes between males and females and between socioeconomic groups; and they suggest that a response to the epidemic must be as multi-tiered and multifaceted as its causes, a point on which many international recommendations agree.[52–54] But neither the Brazilian government's recommended response nor most international recommendations recognize the role that social networks might play in the epidemic and its control. Where social factors do enter the discussion, they are mentioned broadly under umbrella terms of “social support”, “social marketing”, “social norms”, or “sociocultural environments.”

In Swinburn et al.'s oft-cited theoretical framework for understanding the obesity epidemic, a distinction is made between interventions aimed at changing behaviors and interventions aimed at changing environments, and the authors conclude that the latter category encompasses those interventions truly capable of significantly altering the trajectory of the epidemic.[8] That study fails to note that the social environment of an individual is more than the sociocultural or socioeconomic context of his or her community; more concretely, an individual belongs to networks of specific people with specific attitudes and approaches to shared physical and cultural spaces. A growing body of research suggests that these relationships between individuals play an important role in the processes of gaining and losing weight.[55–57]

Network-based studies of the obesity epidemic recognize that friends' attitudes and behaviors can both reflect and contribute to an obesogenic environment. In this sense, there can be overlap between behavioral and environmental interventions: changing the behavior of one person affects the environments of other people. Returning to Figure 1, the purple sections indicate that social networks contribute to both the environment an individual experiences and to how an individual responds to that environment. The purple section of the “intervention reach” triangle at the bottom of the figure indicates that interventions targeting an individual's behavior extend beyond an individual; they also affect those within the individual's social network. Ultimately, people rarely make decisions regarding energy consumption and expenditure independently, and consideration of an individual's social network could be important in planning and executing effective and efficient approaches to addressing the epidemic of obesity and overweight in Brazil and internationally.

iv. A social network approach to understanding and responding to the obesity epidemic

As already mentioned, increases in BMI result from a combination of genetic, physiological, and behavioral factors at the individual level; local and regional factors at the environmental level; and national, international, and global factors at the systemic level. The disease process is further complicated by network interactions on every scale: metabolic pathways share components and diseases share genetic or functional origins; individuals exist and operate within social networks of friends and family; and national policies are established with an eye to international treaties and alliances. With a more specific understanding of how these networks impact individuals and populations, interventions can more appropriately address the global epidemic of obesity and overweight. Social networks in particular represent an under-explored aspect of the epidemic with potentially important implications for population-level weight control.[58]

Social network analyses acknowledge the role of relationships in individual-level outcomes and the role of distributions of relationships in community-level outcomes. The types of relationships and the positions of those relationships in a network are recognized as potentially relevant to health. In this sense, social network analyses are more informative than studies of social support, which are sometimes erroneously categorized as social network studies. Analyses of social support consider the impact of how supported an individual is or feels – often measured by an individual's claim to the number or perceived helpfulness of his or her close-friends – on a particular health outcome. [59] Social network data is more informative than social support data in that, in addition to individual-level analysis, it allows for dyad-level and community-level analyses, but it is also more complicated to obtain and interpret.

While early studies of social networks were generally sociological, social-network-based studies in the fields of epidemiology and public-health have become more common. In the 1970s and 1980s, several studies were published that suggested an individual's perceived level of social support had an impact on his or her mortality, and dyad-level studies indicated that, even after controlling for the role of health in spouse selection, marriage was associated with longer life. More recently, studies have extended beyond social support or dyadic effects to supradyadic, global networks, revealing more complicated associations between health and social networks, including new forms of contagion.[59] In network-based studies of infectious diseases, "contagion" implies the transfer from person to person of something physical, like a virus or a bacterium, sometimes via a non-human vector. On the other hand, network-based studies of noninfectious diseases focus on the spread of ideas and behaviors that in turn have the potential to affect health. This type of transmission is often referred to as “social contagion”.[60]

As the obesity epidemic has expanded, studies regarding social components to the processes of gaining and losing weight have become more common. A study published in 1988 showed a relationship between binge-eating in sororities and popularity,[61] and another published in 1991 showed an association between loneliness and dietary inadequacy in independently living elderly individuals.[62] Other studies in the 1990s showed that obese children were less likely to become obese adults when supported socially in healthy eating and exercise by friends and family;[63] and that individuals recruited into a weight-loss program with friends and given social support during and after the program were more likely both to complete the program and to have maintained their weight-loss at a 10 month follow-up.[64] These studies and others[55–57] demonstrate that even before the first supradyadic study of the association between obesity and social networks, published in 2007 by Nicholas Christakis and James Fowler,[65] researchers had begun exploring the social aspects of weight loss and weight gain.

Christakis and Fowler created a dynamic social network graph from longitudinal friendship data collected over 32 years through the Framingham Heart Study (FHS), and, from an analysis of the changing BMIs of individuals within that network, they concluded that obesity is socially contagious. They arrived at this conclusion through several pieces of evidence. First, they noticed clusters of obese friends in the network, and identified three possible explanations for this clustering: homophily, or obese individuals seeking out friendships with other obese individuals; confounding, or factors associated with obesity – such as shared obesogenic environments or shared genetic predispositions – leading two friends to become obese together; and induction, or obesity spreading from one individual to another through the social contagion of unhealthy eating and exercise habits. Three other observations led them to conclude that induction was at least playing a part in the observed clustering of obesity: weight gain in a geographic neighbor did not affect the probability of an individual becoming obese; the chance of an individual becoming obese from one point in time to the next rose if one of his or her friends had recently become obese, and this effect was most pronounced in mutual, same-sex friendships. If the increased risk of obesity was a result of homophily, Christakis and Fowler reasoned, there would be no time lag, and if it were a result of confounding, friendship type should not have made a difference in the probability of an individual becoming obese, and geographic neighbors might have been more likely to grow obese together.[65]

Christakis and Fowler's study inspired further exploration into the relationship between an individuals supradyadic social network and his or her BMI, and while comprehensive network data is still relatively rare, several recent review articles indicate that the number of studies utilizing such data is growing.[55–57] These studies generally support the conclusion that obesity clusters in social networks, but the causes of the social clustering of obese individuals are still poorly understood.[55] Despite the conclusions Christakis and Fowler draw regarding the contagion of obesity, their results could still potentially be explained by a shared environment (confounding) – mutual friends could tend to share more similar environments than non-mutual friends – or by homophily – individuals could form friendships based on shared unhealthy behaviors rather than shared body types.[66,67] Researchers have also warned that, without controlling for endogenous network factors, there exists a tendency to overestimate obesity clustering.[68] Endogenous factors – explained in greater detail in the following sections – are network-specific factors that might affect friendship formation. For instance, if two individuals share a common friend, they are more likely to become friends themselves. Still, at least one study that did control for endogenous factors still found evidence of obesity clustering.[68]

Despite remaining uncertainties regarding the mechanisms by which clusters of obese friends form, associations between social networks and BMI could carry important implications for the propagation of obesity and the epidemic's control. Several studies have incorporated social contagion into models estimating the future trajectory of the epidemic.[69–71] Others have estimated how social components to obesity might be exploited in curbing the epidemic,[71,72] and how the cost effectiveness of potential prevention approaches change when social network effects are taken into consideration.[73] Despite this research, however, social-network-based obesity interventions remain rare, and the effectiveness of such interventions largely remains to be seen.[74]

In Brazil, eating is more social than it is in many of the North American and European countries from which most network data has thus far been gathered During the work day, lunch breaks often allow time for employees to enjoy a meal with colleagues at on-site dining halls or in local restaurants. These observations justify the government's workplace-based initiatives aimed at promoting healthy eating and exercise habits, but they also justify social network analyses specific to overweight and obesity in Brazilian institutional settings. Such analyses have the potential to inform more nuanced and efficient institutional anti-obesity programs.

In exploring the relationship between the BMIs and social networks of civil servants in a Rio de Janeiro work-based cohort, this study aims to begin a discussion regarding the role that interpersonal factors might play in the proliferation and control of obesity in Brazil. Sectional friendship data collected between 2012 and 2014 was used to construct a sociocentric social network, here-in referred to as ELSA-RioSC, of the 1,521 actively-employed Rio de Janeiro participants of the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil).The relationships between individuals' sectional and longitudinal BMIs and their social positions and friends' BMIs were assessed in various ways, controlling for potentially confounding exogenous and endogenous factors. Given the potentially different roles eating, exercise and body image might play in the social lives of Brazilian men and women, analyses were stratified by sex where possible. Specifics regarding methods, results, conclusions, limitations, and future research directions are described in the following sections. Despite this studies limitations, it contributes important data and results to the sparse research on social networks in Brazil and, more broadly, on social networks in developing countries. It also adds to the growing global collection of supradyadic social network studies of obesity. Most specifically, it serves as an introductory exploration into the role of interpersonal relationships in the Brazilian obesity epidemic and its workplace-based control.

IV. Theoretical Foundations and Model Rationale

Note: The information summarized in the following sections is limited to the principles and processes particularly relevant to this study. For a more complete review of basic graph theory, see Wasserman and Faust, 1994[75] *or Kolaczyk, 2009.*[76]*For specifics regarding the construction and interpretation of social network graphs in R, see Kolaczyk and Csárdi, 2014.*[77]

i. Basic graph terminology

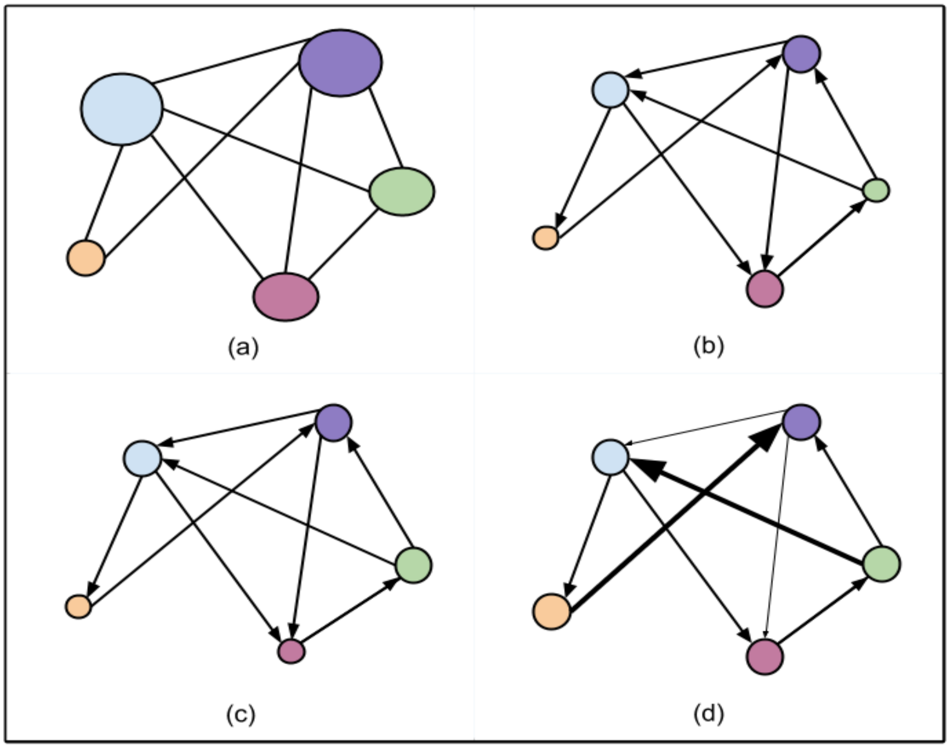
A graph is a visual mapping of a network. Network graphs are comprised of two principal features: *vertices* (or nodes) and *edges*. Mathematically, a graph is expressed as $G=(V, E)$ where *V* is the set of vertices and *E* is the set of edges. In a social network graph, vertices represent people, and edges represent the relationships between people. An edge can be directed, as in Figure 3b-d, or undirected, as in Figure 3a. A directed edge represents a directed relationship. In some cases two individuals necessarily agree on the type of relationship they share, like a brother and sister, and an undirected edge is most appropriate, but in other cases directed edges can contribute important information, like in a network of friends. In such a network, a friendship could be mutual, with both individuals naming the other as a friend, or one-sided, with only one of the two individuals considering the other to be a friend. In the case of a directed friendship, the vertex from which an edge emanates can also be referred to as the "ego", and the friend being cited can be referred to as the "alter".

Beyond the direction of a relationship, both vertices and edges can carry information. Edges can be given varied levels of importance through a vector of edge weights, $[w_e]_{e \in E}$, where relationships deemed more important are given a greater weight. Figure 3d shows a visual representation of a network with weighted edges. For vertices, a vector of attributes, $[x_v]_{v \in V}$, can carry information like the sex, age, and BMI of each individual in a network.

Graphs can be either sociocentric or egocentric. An egocentric graph is one emanating from a single vertex, and the relationships between the referenced vertices, or alters, are unknown. The unit of analysis in an egocentric graph is the individual: what do an individual's friends tell us about his or her health? A sociocentric graph is one in which the relationships between all vertices are known, and analysis extends beyond the individual to the distribution edges: what does a set of edges tell us about the role relationships play in health or the role health plays in relationships?

The properties of a social network, or "network graph characteristics", can include distributions of characteristics measured at the levels of vertices and edges, like local indicators of centrality and connectivity, and characteristics that apply to the graph as a whole, like measures of reciprocity and density. The specific measures included in the present study are described in the following sections.

Figure 3 – Basic graph vertex and edge characteristics



Graph (a) shows an undirected graph where the sizes of the vertices are proportional to their degrees. Graphs (b) and (c) show directed graphs with vertices proportional to in-degrees and out-degrees, respectively. In graph (d), the vertices are equally sized, but the edge sizes reflect arbitrarily assigned edge weights.

ii. Vertex characteristics

Three measures of vertex centrality were considered in this study: degree, betweenness centrality, and eigenvector centrality. The “in-degree”of a vertex is a count of the number of edges incident upon the vertex, the “out-degree”is a count of the number of edges emanating from the vertex, and the “undirected degree” is the sum of the in-degree and the out-degree.

Betweenness centrality is the proportion of shortest “paths” that include a given vertex, where a path is an alternating edge-vertex route between two vertices. In equation form,

[2] $$c_B = \sum_{s \neq t \neq v \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

where $\sigma(s,t|v)$ is the number of shortest paths between *s* and *t* that include *v* and $\sigma(s,t)$ is the total number of shortest paths between s and t. Betweenness centrality is therefore a quantification of the extent to which an individual is a "connector" of different sub-graphs.

Eigenvector centrality is a score calculated from the values of the first eigenvector of the graph adjacency matrix. A vertex with a large score is connected to many vertices with large degrees. Eigenvector centrality therefore captures the idea that a person with many well-connected friends might be important to a network. It is calculated according to Equation [3],

[3] $$c_{E_i}=\alpha\sum_{\{u,v\}\in E} c_{E_i}(u)$$

where the vector $\boldsymbol{c}_{E_i}=(c_{E_i}(1),\ldots,c_{E_i}(N_v))^T$ is the solution to $\boldsymbol{A}\boldsymbol{c}_{E_i}=\alpha^{-1}\boldsymbol{c}_{E_i}$. ***A*** is the graph adjacency matrix and α^{-1} is the largest eigenvalue of ***A***.

Finally, graph connectivity can also be measured at the individual level through a local clustering coefficient. A clustering coefficient has a value between zero and one and measures the proportion of "triangles" to "connected triples", where a triangle is a collection of three individuals who are all friends with one another and a connected triple is a collection of three vertices connected by at least two unique edges. The local clustering coefficient therefore measures the proportion of pairs of an individual's friends who are also friends with one another. In equation form:

[4] $$cl(v)=\frac{\tau_\Delta(v)}{\tau_3(v)}$$

where, for vertex *v*, $\tau_\Delta(v)$ is the number of triangles and $\tau_3(v)$ is the number of connected triples.

iii. Graph characteristics and comparison to random graphs

Some graph characteristics are summaries of vertex characteristics. For instance, an out-, in- or undirected degree distribution shows the frequencies of each degree value in a graph. An average degree is simply the mean of the vertex degrees.

Other graph characteristics are global by definition. Several measures of graph cohesion fall into this category. Graph density is measured as the ratio of existing edges to possible edges. The density of a graph $H=(V_H,E_H)$ is thus defined as:

[5] $$den(H)=\frac{|E_H|}{|V_H|(|V_H|-1)/2}$$

A maximum density of one would correspond to a "clique" - a graph or subgraph in which each vertex is connected by edges to every other vertex. In the case of a "fixed choice" network, where there is a limit on the maximum number of edges emanating from each node, the maximum density would be

$den(H)=\frac{|V_H|\times OUT_{max}}{|V_H|(|V_H|-1)/2}$ where OUT_{max} is the maximum out-degree.

Another measure of the density of edges in a graph is the global version of the clustering coefficient:

$$[6] \qquad cl(G)=\frac{1}{V'}\sum_{v\in V'} cl(v)$$

where G is the graph of interest and $V'\subseteq V$ is the set of vertices with a degree of at least two.

In a directed graph (like a friendship network), "reciprocity" can also give an indication of graph cohesion. Reciprocity is a measure of the probability that two individuals are mutual friends conditional on the fact that they are at least non-mutual friends.

A "connected" graph is one in which a path connects every vertex to every other vertex. The "giant component"of a graph is the largest connected subgraph, and for a directed graph, it can be defined as either "weak" or "strong". The weak giant component assumes that all connections are undirected, and a path can run through an edge in either direction, whereas the strong giant component only allows for paths to follow the direction of the relationships. The strong giant component will therefore always be smaller than the weak giant component. One measure of connectivity is the fraction of vertices included in the giant component (weak or strong). Another is the average path length between all connected vertices.

Graph attributes can be compared to those of random graphs in order to ascertain whether or not an observed network is likely to have formed by a mechanism similar to that used to generate a random network. If a characteristic of an observed graph is very unlikely given the probability distribution of that same characteristic in a collection of random graphs generated by a defined mechanism, it can be concluded that the observed graph is unlikely to have developed by that same mechanism. Many mechanisms exist for the generation of random graphs with distinctive characteristics, but the three utilized here were classical random graph models, small-world models, and preferential attachment models.

The Erdős–Rényi model is a model for generating classical random graphs. Given a specific number of vertices and edges, in an Erdős–Rényi model, the edges are simply distributed amongst pairs of vertices at random. All possible distributions of edges are therefore equally likely.

Small-world models were developed to account for the fact that most real-world networks have higher degrees of clustering than classical random graph models. In the Watts-Strogatz model, used in this analysis, vertices are arranged in a circular array and connected to *K* neighbors, *K/2* on each side. This leads to a high degree of clustering, but the low average path length – also typical of real-world networks and well captured by classical random graph models – is lost. To recover a low average path length, each edge in the graph is either rewired or left unchanged with a defined constant probability. If the edge is rewired, one end remains fixed on a vertex, and the other end is moved at random to any of the other vertices in the graph.

Preferential attachment models attempt to account for another discrepancy between some real-world networks and most random graph models: many networks have hubs – vertices with very high degrees – and broad degree distributions that drop precipitously (often following power laws). In a friendship network, this would mean many individuals having very few friends and a few individuals having many friends. In one such preferential attachment model, the Barabási-Albert model, the graph begins with a small connected network and grows over time. As new vertices are added to the graph, they are connected to existing vertices with a probability proportional to the degrees of those existing vertices. In other words, vertices rich in edges tend to get richer, creating well-connected hubs. Barabási-Albert models show the desired power law degree distributions, but they have much lower levels of clustering than most observed networks.

iv. Network modeling

While random graph models are useful in determining whether or not graph characteristics of an observed network would be unlikely under certain defined mechanisms of stochastic network construction, they are generally too simple to be useful in the statistical modeling of an observed network. More appropriate approaches have been developed to model both the formation of a network and the distribution of a variable within an existing network. In the case of BMI, the former approach would involve determination of whether or not similarity in the BMIs of two vertices increases the probability of an edge forming between those vertices. The latter approach would involve determining whether or not the BMI of an individual is affected by the BMI of his or her friends. The appropriateness of each approach therefore depends on an assumption regarding the mechanism of BMI in friendships: does BMI play a role in the formation of friendships (homophily), do existing friendships play a role in an individual's BMI (induction), are both important, or is neither?

Exponential random graph models

Development, fit and comparison of exponential random graph models (ERGMs) is in many ways analogous to those same processes in generalized linear models.[77,78] ERGMs assume that, given a collection of vertices, an observed set of edges is the product of an unknown stochastic process, and that set represents one possibility of the many possible graphs that might have been generated through the same process. An ERGM aims to define a process that may have plausibly led to the observed set of edges given the set of vertices. The existence or non-existence of an edge is assumed to be a random variable: a function defines the conditional probability of edge formation rather than defining concretely whether or not an edge exists. This assumption recognizes that part of the process of edge formation will be unknown and undefined. As with the assumption in a generalized linear model that the relationship defined by a coefficient is the same for all observations, in exponential random graph models, there is an assumption of homogeneity: the probability is equal of an edge forming between any two vertices within a subset to which a given set of parameters applies.

The conditions on which the probability of edge formation in an ERGM depends can include both endogenous, or structural, effects and exogenous, or node-level, effects. Endogenous effects are parameters that allow for the probability of edge formation to depend on the existence of other edges. Two examples would be reciprocity, or the idea that the probability of a directed edge forming between two vertices is greater if an edge in the opposite direction already exists, and transitivity, or the idea that the probability of an edge forming between two vertices is greater if those vertices share a common contact. These effects can be conditioned on a variable: perhaps reciprocity effects are more significant for edges within age groups than for those between age groups, for instance. Models that include endogenous effects are considered “dyad dependent”, meaning that the formation of each edge is dependent on the existence of other edges in the graph.

Exogenous effects are parameters that allow for the probability of edge formation to depend on individual vertex characteristics or on comparative vertex characteristics. For example, an exogenous effect might be the difference in BMI between two individuals – if the difference is small, a friendship might be more likely to form. The non-comparative BMI of an individual could also be taken into consideration – perhaps individuals with large BMIs are generally less likely to form friendships than those with average BMIs. A model that only includes exogenous parameters is considered “dyad independent”. A well defined model will likely include both exogenous and endogenous effects. For a dyad-independent model, parameter estimation can be accomplished through maximum pseudolikelihood estimation (MPLE), which is a more localized and less computationally expensive approach than maximum likelihood estimation. For a dyad-dependent model, however,, such estimations perform poorly, and Markov chain Monte Carlo (MCMC) simulation methods are preferred.[79]

The estimated coefficients are interpreted as a conditional log-odds ratio for friendship formation. Equation [7] expresses one formulation of an ERGM:

[7] $$logit(P(Y_{ij}=1|n\,actors, Y_{ij}^{c}))=\sum_{k=1}^{K}\theta_k\,\delta z_k(y)$$

where Y_{ij} is a dyad between vertices *i* and *j*, and Y_{ij}^{c} represents all other dyads. The inclusion of the latter term indicates that the probability of a dyad is dependent on other dyads, and is therefore only necessary for models that include endogenous effects. $z_k(y)$ are network statistics calculated on the observed adjacency matrix, *y*. These statistics can include, for example, the number of edges, the number of triangles, or the number of edges between individuals of the same sex. *K* is the set of all network statistics, θ_k is a parameter determining the impact of each statistic for a specific data set, and δ dictates the change in $z_k(y)$ when Y_{ij} is toggled from 0 to 1. ERGM parameters are interpreted as the conditional log odds of a new tie forming. For example, if the coefficient for sharing the same categorical BMI were 0.31, when all other variables are controlled for, sharing a BMI category increases the odds of a friendship forming by exp(0.31)=1.363, or approximately 36%.

The goodness-of-fit of an ERGM can be determined in a manner similar to that described for the comparison of observed graph attributes to those of random graphs: the proposed model is used to generate a series of graphs, and characteristics of the observed graph are compared to the distribution of characteristics of the generated graphs. If the observed graph characteristics are not extremely unlikely given the distributions generated by the proposed model, the model fits the data reasonably well.

For a more complete treatment of ERGM theory, see Robins, et al. or Goodreau, et al. [78,79]

Generalized linear models and latent Gaussian models

As previously noted, it is also possible to model BMI within an existing network, thereby testing the theory that friends affect the BMI of an individual, rather than (or in addition to) the BMI of an individual affecting the formation of friendships. In modeling an individual's sectional BMI and change in BMI over time, two methods of accounting for the potential non-independence in the response variable were tested: the average BMIs of an individual's friends, stratified by friendship type, were included as an explanatory variables in both generalized linear and latent Gaussian models; and the impact on the final models of including a latent Gaussian field with a network dependency structure was assessed.

Latent Gaussian models are a class of structured additive regression models in which the latent field is Gaussian, even when the response variable is not. Using a Bayesian analysis, the posterior marginals on such a model can be estimated either through an MCMC algorithm or, more efficiently, through an integrated nested Laplace approximation (INLA).[80] The latter approach was utilized in the present study.

Specific models can be specified for the latent Gaussian field, and three models were considered in this analysis: one defining the latent Gaussian field to be a vector of independent and normally distributed random variables, one assigning a network dependency to the latent Gaussian field, and one defined by the union of the previous two.

v. Model rationale

Three classes of models were fit to the ELSA-Rio data: male and female sectional models for Wave 2 BMI, male and female models for the average annual percent change in BMI between Waves 1 and 2, and an exponential random graph model. In total, therefore, five final models were developed. Latent Gaussian models (all but the ERGM) were fit to each sex independently because of previously reported differences in associations between BMI and socioeconomic variables for male and female professionals in Rio de Janeiro.[81] Fonseca, et al. showed that degree of education was inversely associated with BMI among female employees but not male employees, and income level was not associated with BMI for either sex. It was also considered likely that the relationship between an individual's BMI and his or her social network would differ between males and females.

As noted above, two approaches to the inclusion of friend BMI information in the latent Gaussian models were considered: through a latent spacial effect and through variables summarizing nearest neighbor averages. These approaches are defined in greater detail in the Methods section below. The variables considered for inclusion in the models are defined in Table 1, and the rational behind those considerations is described here. Specifics regarding data collection, model selection, and model evaluation can again be found in the Methods section.

Given that the ERGM was fit to the network and not the BMI of individuals within the network, the data could not be reasonably separated by sex. The role of sex in friendship formation was, however, controlled for in the model, as described in the “Exponential random graph model” subsection.

Wave 2 BMI

A sectional analysis of Wave 2 BMI was undertaken to assess whether or not, after controlling for potentially confounding variables, an association exists between the BMI of an individual and his or her social position or friends' BMIs at a particular point in time. This analysis would show if BMI tends to cluster in ELSA-RioSC, as has been observed in other social networks of both adults[65,82] and adolescents,[57] and if those with a higher BMI tend to be more socially marginalized, another observation common to several network studies, especially in regards to adolescent school networks.[83–85]

The explanatory variables taken into consideration could thus be divided into three categories, namely: controls, network vertex characteristics, and characteristics of friends. The control variables were sex, age, education, and per-capita monthly family income, all collected through the Wave 2 questionnaire, as discussed in the Methods section. These factors have been shown to be associated with BMI on an individual level in Brazilian populations.[86–88] Their inclusion therefore controls for possible confounding effects. For example, if BMI and degree are negatively correlated in a model that includes education, the correlation cannot be explained by both lower BMI and greater degree resulting from a greater level of education; the correlation exists even within a fixed education level.

The network vertex characteristics taken into consideration were the in-degree, out-degree, undirected-degree, betweenness centrality, eigenvector centrality, and local clustering coefficient. As described in the Theoretical Foundation section, in-degree is a measure of the number of people who consider you a friend, and it can be thought of as a measure of popularity. A low in-degree could serve as an indicator of social marginalization.[83] When paired with a low out-degree, it could indicate social isolation, and when paired with a normal or large out-degree, it could indicate a discrepancy between mutual and ego-perceived friendships.[85] Given that friendship nominations were limited to five, and not confined to active Wave 2 participants, inclusion of undirected degree is a method of measuring social connectedness under the assumption that all friendships, regardless of their directionality, might in fact tend to be mutual. Aside from degree, measures of centrality – namely betweenness centrality and eigenvector centrality – were included to assess whether those with larger BMIs might be more or less central to an at-work social network. If betweenness centrality were negatively correlated with BMI, it would suggest that more overweight individuals are less likely to connect distinct friend groups. If eigenvector centrality were negatively correlated with BMI, it would suggest that more overweight individuals are less likely to make "popular" friends. These variables thus provide more nuance in an assessment of the social marginalization – or lack thereof – of overweight and obese adults in a workplace.

Finally, the friend BMI characteristics were included to assess whether or not friends generally share similar BMIs. While a sectional analysis would not permit conclusions regarding the cause of BMI clustering, it could identify whether or not such clustering exists in ELSA-RioSC, as it does in other adult social networks.[65,82] Studies of those networks and others have also suggested that the directionality of friendships is important. Among adults, evidence suggests that BMI is more strongly associated between mutual friends,[65,82] while at least one study of an adolescent network has suggested that the opposite is true.[89] Furthermore, in unidirectional friendships, the BMI of an individual is more likely to be associated with the BMIs of those he or she cites as friends than with those who cite the individual as a friend.[65,90] Inclusion of the average BMI of the friends an individual cites, the average BMI of those who cite an individual as a friend, and the average BMI of bi-directional (mutual) friendships – all measured in Wave 2 – in the model for individual Wave 2 BMIs allowed for an assessment of the importance of friendship directionality in associations between friends' BMIs in ELSA-RioSC. The average BMI of cited friends was weighted by the order in which friends were cited, as described in Equation [9].

[9] $$OF_W = \frac{\sum_1^5 \frac{1}{i} \times BMI_i}{\sum_1^5 \frac{1}{i}}$$

where OF_W represents the weighted average BMI of "out", or cited, friends, and BMI_i represents the BMI of the i^{th} cited friend. The hypothesis was that an individual would cite closer friends first, and that the BMI of closer friends, like mutual friends, might be more highly correlated with an individual's own BMI, as was found in at least one other study.[82] A measure of the proportion of an individual's friends who are overweight or obese was also included, the hypothesis being that an individual with a greater proportion of overweight friends might tend to have a greater BMI, even when the average BMI of his or friends is normal. This was found to be the case in Trogdon, et al.[91]

A summary of the control characteristics, network vertex characteristics, and characteristics of friends can be found in Table 1.

Average annual percent change in BMI

Regardless of whether or not BMI clusters in a friendship network at a specific point in time, it is possible that friends' BMIs tend to change in similar ways. While friendship tie data was only collected for Wave 2, if at-work friendships are assumed to stay relatively constant over 2-6 years (the time between the two ELSA Waves), the association between friends' BMIs – or the average change in friends' BMIs – and the average annual percent change in an individual's BMI can be estimated. As in the case of the Wave 2 sectional analysis, several possible mechanisms could lead to correlated changes in BMI among friends: friends could respond to a shared environment similarly (confounding), friendships could form around similarities in exercise and eating habits (homophily), or friends could affect one another's exercise and eating habits over time (induction).

As described in the Introduction, one study of an adult social network concluded from longitudinal data that an individual is more likely to gain weight after his or her friends gain weight, and changes in mutual friends' weights are more strongly correlated than changes in non-mutual friends' weights. Christakis and Fowler interpreted these results as suggestive of induction,[65] but, again as noted in the Introduction, other studies questioned that interpretation, recognizing that similar patterns of directionality and time-lag might have resulted from behavioral homophily or shared environment.[66,67] In the case of behavioral homophily, a friendship could form around shared unhealthy eating habits, which would lead to higher probabilities of both one friend already being obese and of the other friend, if non-obese, becoming obese at a later point in time. A shared environment could also increase the probability of both friends becoming obese, even if that process took longer in one friend than in the other. In either case, if friendships are assumed to form on the basis of shared environments or behaviors, the strongest friendships would be expected to show the strongest correlations.[67]

Even where the causes of friend-associated weight change are debated, research has indicated that the association exists in both adolescent and adult networks.[65,90,92] Social networks and social support have also been found to affect the success of weight-loss interventions,[64,93] and models have indicated that public health policies targeting obesity within a social context would prove more effective in addressing the epidemic than current strategies.[71]

The average annual percent change in an individual's BMI was calculated according to Equation [10]:

[10] $$Average\,\%\,Change\ per\ Year = \frac{Variable_2 - Variable_1}{Variable_1 * (Age_2 - Age_1)} * 100\%$$

where the subscript indicates the wave in which the data was collected, and where “Variable” in this case refers to BMI. In determining the most appropriate model, the same three types of explanatory variables were considered as in the sectional analysis, namely: controls, network vertex characteristics, and characteristics of friends.

The control variables were categorical age, education level, per-capita monthly family income (now at Wave 1), and percent annual change in per-capita monthly family income (again calculated according to Equation [10]). As in the sectional analysis, these variables were included to control for some of the socioeconomic factors previously reported to correlate with BMI in Brazilian adults. The change in per-capita income was included to control for the possibility that an increase or decrease in available funds might lead to a change in diet, activity, or lifestyle, thereby affecting BMI over time.

The network vertex characteristics were the same as those considered in the sectional analysis and described in Table 1. The hypothesis was that socially marginalized individuals might be more inclined to gain weight, independent of their current weight status. Several studies have shown that people with greater levels of social support are less likely to regain weight following weight-loss treatment,[64,94] but those studies applied specifically to individuals trying to maintain new weights. It was assumed that most of the individuals in ELSA-RioSC were not in this position, and, given that eating is a very social activity in Brazil,[48] a correlation between social prominence and weight-gain was also considered plausible in the study population.

The friend characteristics were again equivalent to those considered for inclusion in the sectional model, and the data was considered both sectionally (using information from Wave 1 in this case, rather than Wave 2) and by categorical percent annual change. The former was included to test the hypothesis that the sectional BMI of a friend might lead an individual to change his or her own BMI, and the latter to test the hypothesis that independent of the relative BMIs of two friends at a given point in time, they might experience similar longitudinal changes in BMI. Studies have indicated both that the sectional BMIs of an individuals' friends might affect the individual's BMI longitudinally[92] and that the change in the BMI of an individual is correlated with the change in the BMI of his or her friends.[65,90] As in the sectional analysis, the directionality of friendships was taken into consideration, as several studies found that changes in BMI in mutual friendships and in ego-perceived friendships are correlated while those in alter-perceived friendships are not.[65,90]

Table 1 – Description of the latent Gaussian model variables		
	Type	Definition
Dependent Variables		
BMI.2	Continuous	Wave 2 BMI
Average annual percent change in BMI	Continuous	Calculated according to Equation [10]
Control Explanatory Variables		
Sex	Binary	Male or female, collected in both waves with no change.. Models fit to data for each sex separately.
Age	Discrete/ Categorical	Reported in years for Waves 1 and 2. The variable was considered in discrete form for the Wave 2 sectional analysis and categorical form for the longitudinal analysis with the following categories: (35,45], (45,55], and (55,70]
Education	Categorical	Collected in Wave 2. Three categories: No college; At least some college but no graduate school; College and graduate school
Per-capita monthly family income	Categorical	Reported as a continuous variable in Waves 1 and 2. The variable was considered in categorical form with the following categories: (0, 1244], (1244,2487],(2487,3731], (3731,20000]. Categories were defined according to agglomerations of the income categories specified in the ELSA questionnaire.[95]
Average annual percent change in per-capita monthly family income	Continuous	Calculated according to Equation [10].
Network Explanatory Variables		
Undirected degree	Categorical	Node undirected degree, ≥ 0. The categories were: 0, 1, 2, 3+
In-degree	Categorical	Node in-degree, ≥ 0. The categories were: 0, 1, 2, 3+
Out-degree	Categorical	Node out-degree, 0-5. The categories were: 0, 1-2, 3-4, 5+
Local clustering coefficient	Continuous	Calculated according to Equation [4]. When an individual was too disconnected to allow for the calculation, the value was taken as 0.
Betweenness centrality	Continuous	Calculated according to Equation [2]. When an individual was too disconnected to allow for the calculation, the value was taken as 0.
Eigenvector centrality	Continuous	Calculated according to Equation [3]. When an individual was too disconnected to allow for the calculation, the value was taken as 0.
Friend Explanatory Variables *		
Undirected friend average BMI	Categorical**	Average BMI of all friends (both cited and cited-by).
Out friend average BMI	Categorical	Weighted average BMI of all those cited as friends by the individual, calculated according to Equation [9].

CONTINUED ON FOLLOWING PAGE

*Friend Explanatory Variables **		
In friend average BMI	Categorical	Average BMI of all those who cited the individual as a friend.
Mutual friend average BMI	Categorical	Average BMI of all those who both cited the individual as a friend and were cited as a friend by the individual.
Proportion of overweight friends	Categorical	Proportion of cited friends who are overweight or obese, divided into four categories: 0, [1/5 - 2/5], [1/2-2/3], and [3/4,1]. The variable was only considered sectionally; Wave 2 data was used for the sectional analysis and Wave 1 data for the longitudinal analysis.

**For all of the friend variables, where not otherwise specified, the sectional analysis utilized Wave 2 data and the longitudinal analysis utilized both Wave 1 data and the average annual percent change from Wave 1 to Wave 2.*
***All "friend average BMI"variables were divided into three categories:average BMI<25 kg/m^2 (underweight or normal weight), 25 kg/m^2≤ average BMI<30 kg/m^2 (overweight), and average BMI≥30kg/ m^2 (obese).*

Exponential random graph model

As previously described, ERGM involves estimation of a set of endogenous (dyad-dependent) and exogenous (dyad-independent, node- or dyad-level) parameters defining a process that could, with a reasonable probability, have generated the observed network. Through the inclusion of endogenous effects, such as transitivity and mutuality, the ERGM approach avoids over-estimation of network effects: often new friendships develop in large part as a function of existing friendships rather than as a function of exogenous factors.[84,89] For example, if an overweight individual is friends with two other overweight individuals, and those individuals become friends – a form of transitivity known as "triad closure"- that homophilous friendship is likely to have developed at least in part through the shared friend rather than exclusively through shared attributes, behaviors, or ideas.

Four endogenous parameters were considered for inclusion in the model, as described in Table 2. The most basic endogenous parameter is a count of the number of edges, which, when combined with the number of nodes, defines the graph density, as per Equation [5]. The "geometrically weighted edgewise shared partners distribution" is included as a measure of transitivity that estimates the probability of a friendship forming given the number of friends the individuals share (i.e. the number of triad closures that would result from the addition of a new edge).[84] The probability of a directed edge forming where one in the opposite direction already exists is captured in the "mutual" term, and the probability that an edge forms given the expected final proportion of isolates to nodes in a network is captured in the "isolates" term. Inclusion of these parameters is consistent with endogenous parameter selections in other ERGM analyses of BMI in social networks.[68,84] An endogenous constraint of a maximum out degree of 5 was included to account for the "fixed choice" question format in the ELSA Wave 2 questionnaire.[96]

The exogenous parameters of principal interest were those associated with BMI. The parameter for “small BMI difference” accounted for the conditional probability of a friendship forming given that two individuals share a relatively similar BMI. As with the friendship variables of the latent Gaussian models, this parameter was included to assess whether or not the social BMI clustering observed in other networks is also present in ELSA-RioSC. The “node sex match”, “node education match”, “node department match”, and “small age difference” terms control for homophily in some of the factors associated with obesity.[86–88] Inclusion of these terms was therefore analogous to the inclusion of “control variables” in the latent Gaussian models.

The strength of the ERGM approach was that it allowed for simultaneous inclusion of both potentially confounding variables and endogenous effects in a model assessing the role of BMI in ELSA-RioSC friendship formation.[65,82]

Table 2 – Description of the ERGM variables

Endogenous Variables	
Edges	Baseline. Adds one parameter equal to the density of the graph.
Geometrically weighted edgewise shared partners distribution	Adds a statistic based on alternating sums of k-triangles; accounts for the role of transitivity in the formation of new friendship ties
Isolates	Adjusts the probability of a new edge forming according to the expected proportion of isolates in the final network
Mutual	A parameter affecting the probability of a new edge forming given that an edge in the opposite direction already exists
Endogenous Constraints	
Max out = 5	A node was restricted to a maximum out-degree of 5.
Exogenous Variables	
Node sex match	The conditional probability of an edge forming given that the nodes it would join are of the same sex.
Node education match	The conditional probability of an edge forming given that the nodes it would join have the same education level.
Node department match	The conditional probability of an edge forming given that the nodes it would join work in the same department.
Small age difference	The conditional probability of an edge forming given that the nodes it would join are close in age.
Small BMI difference	The conditional probability of an edge forming given that the nodes it would join have similar BMIs.

V. Methods

i. Data collection

<u>*ELSA Wave 1*</u>

The baseline ELSA-Brasil cohort (Wave 1) included 15,105 participants from five universities and one research institute, namely: the federal universities of Bahia, Espirito Santo, Minas Gerais, and Rio Grande do Sul; the University of São Paulo; and the Oswaldo Cruz Foundation (Fiocruz), in the city of Rio de Janeiro. Wave 1 data was collected between 2008 and 2010 and included a detailed interview, clinical inspection, and medical laboratory work. Recruitment was limited to those individuals between 35 and 74 years of age either actively employed by or retired from one of the six institutions. Only the 1,784 Rio de Janeiro (Fiocruz) ELSA participants were considered for the present study.[97]

The Wave 1 variables included in this study were: sex, BMI, age, and family income. BMI was calculated according to Equation [1], and height and weight were both measured by trained professionals, using standard techniques and equipment. Sex (binary), age (discrete), work type (categorical) and estimated per-capita monthly family income (continuous) were self-reported in the interview portion of data collection.[97]

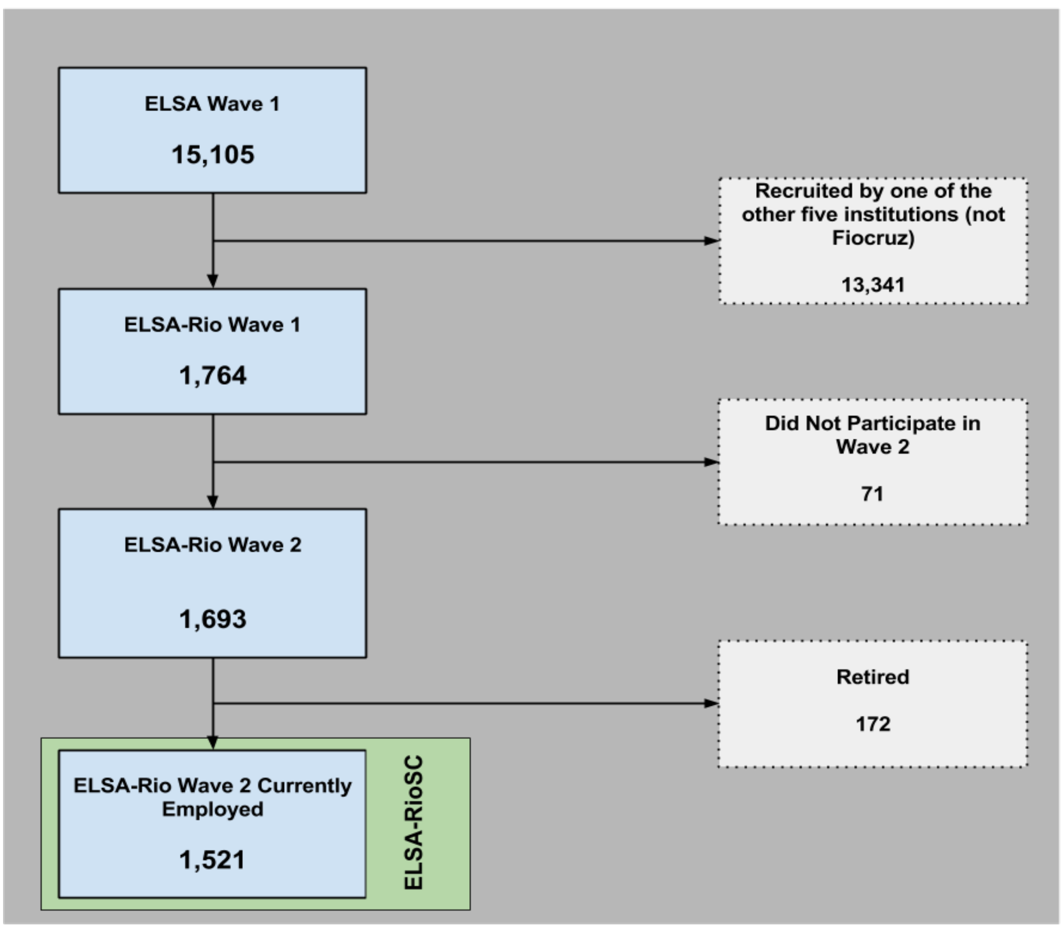
<u>*ELSA Wave 2*</u>

Wave 2 was undertaken between 2012 and 2014, with an average of four years (and a minimum of two and maximum of six) between the two Waves for each participant. Wave 2 again included a detailed interview, clinical inspection, and medical laboratory work. Of the 1,784 original Rio participants, 1,693 (94.9%) were recovered in Wave 2. As with Wave 1, Wave 2 measurements of BMI, age, and income were included in the present analysis, along with self-reported education level (categorical), and data from the newly added "Social Networks" portion of the Wave 2 interview questionnaire. The Wave 1 variables re-collected in Wave 2 – namely per capita monthly family income, age, weight, and height – were collected in equivalent fashions.

The "Social Networks" section of the Wave 2 questionnaire lay the foundation for this study.[95,96] The section included the question "Could you please tell us the complete names of your five closest friends [at work], in addition to the department and sector they work in?", which allowed for the construction of a friendship network, as described in the "Network Construction and Characteristics" section below. The question format was "free recall" for the names of cited friends, meaning participants were not given a roster on which to base their friend selections. The question was also "fixed choice", meaning participants were limited in the number of friends they could cite, with a maximum of five.[75,96] These formating decisions had important implications for the social network analysis, as described in the "Conclusions and Discussion" section below. The most immediate implication was that names listed by free recall had to be probabilistically matched to official names through a linkage process described in the following subsection.

The "Social Networks" questions were only presented to to those ELSA-Brasil participants actively employed by an ELSA institution at the time of their Wave 2 interview, which further limited the size of ELSA-RioSC to 1,521 individuals. The changes in the size of the study population with each additional eligibility constraint are summarized in Figure 4.

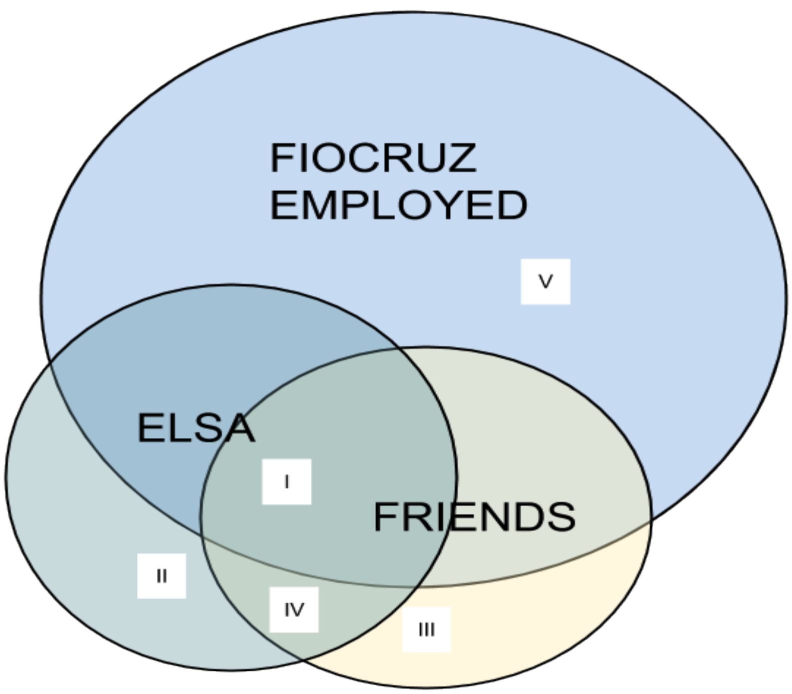
Figure 4 – Flow chart of the change in ELSA-RioSC population size with each additional eligibility constraint



ii. Linkage

Participants reported the names, work departments, and work sectors of their friends during interviews without consultation of an official registry of Fiocruz employees. Therefore, the participant-reported names rarely corresponded exactly to official names. The official names participants were likely referring to were retroactively determined from two databases: the directory of Fiocruz employees, and the directory of ELSA participants. The latter was nearly a subset of the former, but retired ELSA participants, while still eligible to participate in the cohort study, were not included in the directory of Fiocruz employees. Figure 5 summarizes the relationship between the three databases. Section I in Figure 5 is the only section that can be considered sociocentric, and the ELSA-RioSC edge list was comprised only of those edges between individuals in Section 1. The only relevant linkage for ELSA-RioSC construction, therefore, was the linkage between the free recall list of names and the directory of ELSA participants, and subsequent discussion will be limited to that linkage process.

Figure 5 – Venn diagram of the "cited friend", "ELSA participant", and "Fiocruz employee" databases



Section I represents participant-cited friends who were also ELSA participants and actively employed by Fiocruz; section II represents retired ELSA participants who were no longer in the database of Fiocruz employees; section III represents all participant-cited friends who were not in the Fiocruz or ELSA databases, including both retired non-ELSA Fiocruz employees and individuals working on the Fiocruz campus but employed by a third party; section IV represents retired ELSA participants who were cited as a friend by an active participant; and section V represents Fiocruz employees who were not ELSA participants and were not cited as a friend by any ELSA participants.

Likely matches were identified with the help of a probabilistic linkage software, OpenRecLink.[98,99] OpenRecLink was used to facilitate the name revision process, but all matches were inspected individually and accepted or rejected according to researchers' evaluations of the uniqueness of the match. The OpenRecLink software compared two databases – in this case, the database of participant-reported friend names and the database of ELSA participants – and for each entry in one it found the entries most likely to represent the same individual in the other. To improve the efficiency of the program, comparisons were limited by "blocks", where blocks were defined according to exact matches in one or more of the available fields, namely: work department, work sector, first name, and last name. In data collection, friends' work departments and sectors were both selected from a list, but the large number of available options and the wide variation in selected answers for work sector made it a bad choice for blocking. Furthermore, work sector was only reported for 42% of cited friends. By comparison, work department was reported for 95% of cited friends. For 94% of cited friends, more than one name was given. Blocking was therefore done in several steps by work department and by the Soundex codes of the first and last cited names of friends.

First, blocks were defined by all three factors, and the high probability matches were either accepted or rejected. Next, blocks were defined by work department and first name Soundex code; by first and last name Soundex codes; and by work department and last name Soundex code. High probability matches were again inspected and accepted or rejected. Remaining cited friend names were then compared to ELSA participant names without blocking. Finally, any cited friend names that had not been matched with the assistance of the OpenRecLink program were manually inspected and compared to the list of ELSA participants to identify remaining likely matches. This process ensured that the possibility of finding a convincing match was not dismissed even for those cited friends for whom limited available information made it difficult for OpenRecLink to identify high-probability matches.

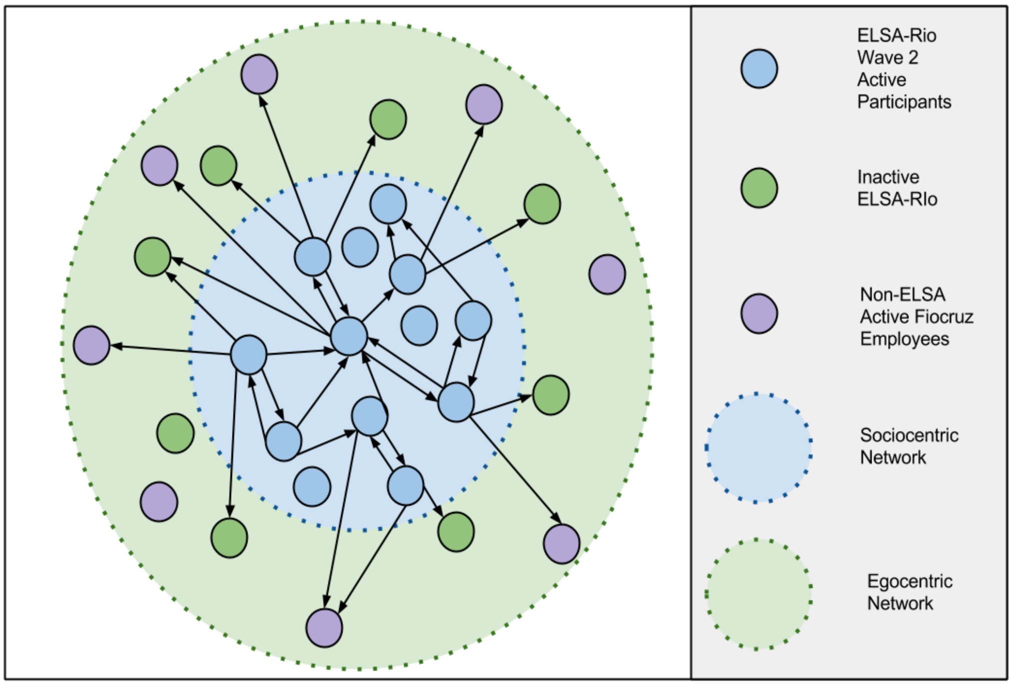
Following linkage, a two-column list of edges was generated, with the first column containing the interviewed individual (or "ego), the second column containing the cited friend (or "alter"), and each row representing one friendship. This edge list was sent to the ELSA data center, where it was codified and returned, along with the variables described in the previous sections, namely: sex, type of work, employment status, age, BMI, family income, and education. These variables were associated with the codes rather than the names of the participants. The separation of the linkage process from the modeling process via codification of the edge-list made it impossible to match participant attributes to their names, minimizing any ethical concerns.

iii. Network construction and characteristics

ELSA-RioSC was limited to those ELSA participants given the opportunity to list their at-work friends (i.e. to whom the "Social Network" section was administered). ELSA-RioSC edges were directed, with some individuals only being cited as a friend, some only citing friends, some both citing and being cited, and some neither citing nor being cited (isolates). Figure 6 shows a simplified model of the directional friendship data. The area in blue represents the ELSA-RioSC sociocentric network, whereas the area in green, which includes non-ELSA Fiocruz employees and ELSA participants who had retired from Fiocruz or did not participate in Wave 2, represents a mixed sociocentric and egocentric network. The individuals from Section III of Figure 5 are not included in the Figure 6 model, which includes only those participant-reported friends who were reasonably matched to current or retired Fiocruz employees. Given the fixed choice format of the social network section of the ELSA Wave 2 questionnaire, the sociocentric graph had a maximum possible density of 0.0066, which would have resulted from all 1,521 individuals citing exactly five friends, all also actively employed Wave 2 ELSA participants.

Several graph characteristics were calculated for ELSA-RioSC – namely the in-degree, out-degree, and undirected degree distributions, the graph density, the global clustering coefficient, the degree of reciprocity, the average path length, and the proportion of vertices in the weak giant component. Using Monte Carlo methods, some of these statistics were compared to those of appropriate random graph models, as described in the "Theoretical Foundations and Model Rationale" section.[65,77] Classical random graph models were generated with the Erdős–Rényi algorithm, preferential attachment models were generated with the Barabási-Albert algorithm, and small-world models were generated with the Watts-Strogatz algorithm, all using the "igraph" package of *R*.[100] In each case, 1,000 random graph models were generated, and the distributions of the relevant statistics were compared to the observed statistics in ELSA-RioSC. The Erdős–Rényi and Barabási-Albert algorithms allowed for generation of directed graphs, while the Watts-Strogatz algorithm did not. For the Erdős–Rényi models, the numbers of nodes and edges were set equal to those of ELSA-RioSC. For the Barabási-Albert models, the number of nodes was set equal to the node count of ELSA-RioSC, and the remaining parameters were left at their default values, as these were found to generate an edge count similar to that of ELSA-RioSC. For the Watts-Strogatz model, the number of nodes was again fixed to the node count of ELSA-RioSC, each vertex was joined to two of its neighbors (one on each side) in the original lattice, and the rewiring probability was set to 0.35. These parameters were found to generate models with undirected degree distributions that best approximated the ELSA-RioSC degree distribution.

Figure 6 – Simplified graph of egocentric and sociocentric ELSA-obtained friendship data



The area in blue shows a simplified model of ELSA-RioSC, the sociocentric network. The area in green shows that the egocentric networks of each ELSA participant extends beyond the Wave 2 actively employed ELSA community.

iv. Generalized linear models and latent Gaussian models

The BMI data was analyzed separately for men and women, and for each sex, two models were developed: one for Wave 2 sectional BMI, and the other for the average annual percent change in BMI from Wave 1 to Wave 2, as previously described in the "Model Rationale" subsection. For both the sectional and longitudinal data, decisions regarding which variables to include in the models were made through generalized linear modeling. Likelihood ratio testing (LRT) was used to determine whether or not the addition of a given variable significantly improved the model fit, with a p-value less than or equal to 0.20 deemed significant. Variables were added in a step-wise fashion: beginning from models including only age, the independent addition of each remaining control variable was tested, and the variable that led to the greatest improvement in fit (the lowest LRT p-value) was added to the model. The process was then repeated until no further control variables improved the model fit significantly, then the inclusion of network and friend variables were tested via the same mechanism. Once the variables for the final model had been determined, Bayesian methods were used to estimate coefficients for a latent Gaussian model with the same set of variables, and the impact of including three different latent Gaussian effect models was assessed through between-model comparisons of the Deviance Information Criteria (DIC) and the Watanabe-Akaike Information Criterion (WAIC). The latent effect models tested were: an independent random variable model ("iid"), a spacial effect model ("besag"), and a model combining spacial and random effects ("bym").

Variable inclusion determinations were made using the "glm" and "add1" functions of the R package. Latent Gaussian models were fit using the "INLA" package.[80]

Missing Data

Wave 2 BMI data was unavailable for 20 of the 1,521 ELSA-RioSC nodes. The average of the Wave 1 BMIs of these individuals, 29.4 kg/m^2, was potentially significantly greater than the ELSA-RioSC overall Wave 1 BMI average, 27.1 kg/m^2, with a one-sided t-test returning a p-value of 0.081. It was therefore deemed important to address the missing BMI data where possible. For 15 of the 20 individuals, Wave 2 weight was available but height was not. In these cases, Wave 2 BMI was calculated from Wave 1 height under the assumption that height would not have changed significantly for most adults in a two to six year time frame. The remaining 5 individuals were not included in the models. Their Wave 1 BMIs were still greater on average than those of the network at large, at 29.3 kg/m^2 (t-test p-value of 0.372), but a loss of five individuals was considered less likely to significantly impact the results than a loss of twenty individuals.

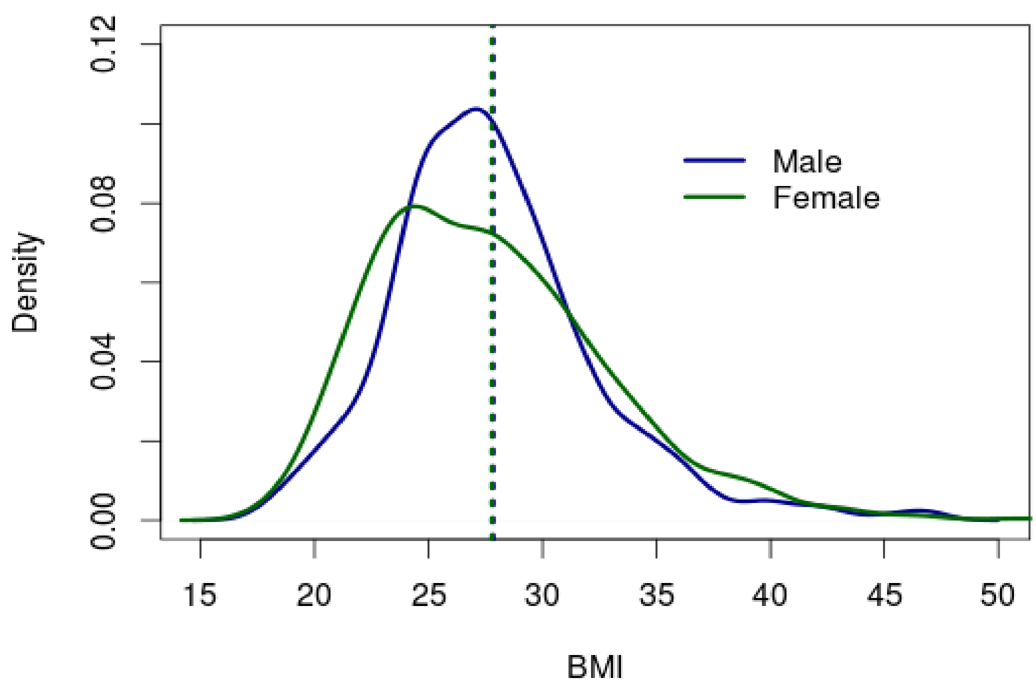
Aside from Wave 2 BMI, the only other categories with missing values were the age at Wave 2 (one missing) and the per-capita monthly family income at Wave 1 (two missing). For the missing Wave 2 age , the average age difference from Wave 1 to Wave 2 of 4 years was added to the age at Wave 1. For the two missing Wave 1 income data points, the average annual change in income for the age groups to which the individuals belonged was multiplied by their age changes from Wave 1 to Wave 2. This product was then subtracted from the Wave 2 income to arrive at an estimate of the Wave 1 income. Given the limited extent of the missing age and income data – affecting a total of 3 individuals – the impact of these estimations on model fits was likely minimal.

Of the 1,521 individuals in ELSA-RioSC, therefore, 1,516 were ultimately included in the sectional and longitudinal generalized linear and latent Gaussian models.

Sectional Wave 2 BMI

Visual inspection of the male and female Wave 2 BMI probability densities – Figure 7 – showed a positive skew. Both generalized linear models and latent Gaussian models operate on an assumption of normally distributed residuals, and while a skew in the BMI distribution does not necessarily correspond to a skew in the residual distribution after a model has been fit, it seemed possible that a gamma distribution would better fit the data than a Gaussian one. The Shapiro-Wilk test returned p-values of 7.934e-13 and 8.246e-13 for the male and the female data, respectively, leading to rejection of the null hypothesis that the data was sampled from a population with normally distributed BMI. Generalized linear models were fit using both gamma and Gaussian distributions (both with identity link functions) and compared through analyses of the residuals and the Akaike Information Criterion (AIC). The distribution deemed most appropriate was then used for the latent Gaussian model.

Figure 7 – Probability densities for male and female sectional Wave 2 BMI

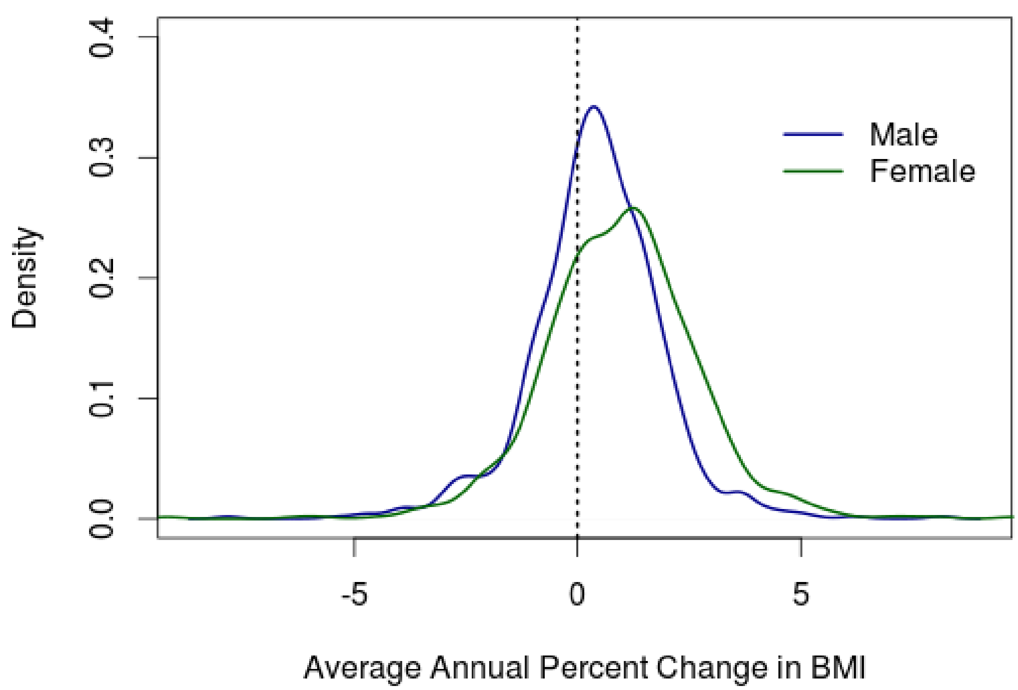


The figure clearly indicates the rightward skews of the distributions. The vertical dotted lines marking the average BMI for each sex show that males and females shared nearly identical averages, but the distributions for each sex were markedly different.

Average Annual Percent Change in BMI from Wave 1 to Wave 2

The average annual percent change in BMI was calculated according to Equation [10]. It was thus a continuous variable with a range including negative and positive values. As with the Wave 2 sectional data, the densities of the average annual percent change in BMI for both men and women were skewed to the right, as shown in Figure 8. The Shapiro-Wilk test returned a p-value of less than 2.2e-16 for both sexes. It was again suspected that a non-Gaussian probability distribution might better model the data than a Gaussian one, but a gamma distribution – favored for the sectional analysis – does not accept negative values. To account for this fact, the data was shifted to the right by 10 annual percentage points such that every value became positive. When the shifted data was then modeled with a gamma distribution with an identity link function, the intercept was the only coefficient affected, and it simply required a 10 unit reduction for proper interpretation. The GLM results obtained using gamma and Gaussian probability distributions were evaluated and compared through analysis of the residuals and relative AICs. The more appropriate of the two distributions was then used in the latent Gaussian model, as in the Wave 2 sectional analysis.

Figure 8 – Probability densities for male and female average annual percent change in BMI



Distribution of the average annual percent change in BMI between Wave 1 and Wave 2 for males (blue) and females (green). The figure shows that the BMIs of most individuals increased between the two Waves, and the distributions for both sexes appear somewhat skewed to the right.

vi. *Exponential random graph model*

Three exponential random graph models were fit to the ELSA-RioSC Wave 2 data. The first ERGM included only exogenous variables (those listed in Table 2). The “small BMI difference” parameter was defined by a difference in BMI of four or less kg/m^2, and 10 years or less was considered a “small age difference.” The exogenous variables deemed significant in the first model – as with the generalized linear and latent Gaussian models, based on a p-value of 0.20 or less – were included in the second and third models. P-values were calculated from the sum of the likelihood variations and the MCMC variations. The “small BMI difference” parameter was included in all models regardless of significance as it was of principle interest to this study.

The second model included the significant exogenous variables from Model 1, the “mutual” and “isolates” endogenous variables, and the endogenous constraint of a maximum out-degree of five. Model 3 included the same variables as Model 2 with the further addition of the geometrically-weighted edgewise shared partners distribution (another endogenous variable). The significance of the “Small BMI difference” parameter was compared across all three models. If the parameter became insignificant or less significant with the inclusion of endogenous effects, those effects might partially explain any apparent social clustering of BMI.

As the first model only included exogenous factors, it was estimated by MPLE, as described in the “Network Modeling¨ section, Section *III.iv*. The second and third models, including both exogenous and endogenous variables, were fit using MCMC estimation procedures. The fits of the models were evaluated according to the procedure again described in Section *III.iv*: a series of graphs was generated from the estimated parameters, and the in-degrees, out-degrees, and edgewise shared-partner values in ELSA-RioSC were compared to the distributions of those same characteristics in the generated series. If the observed values fell within the 95% confidence intervals of the distributions of generated values, the ERGM was deemed a reasonable estimation of one process that may have led to the observed ELSA-RioSC friendship network.

VI. Results

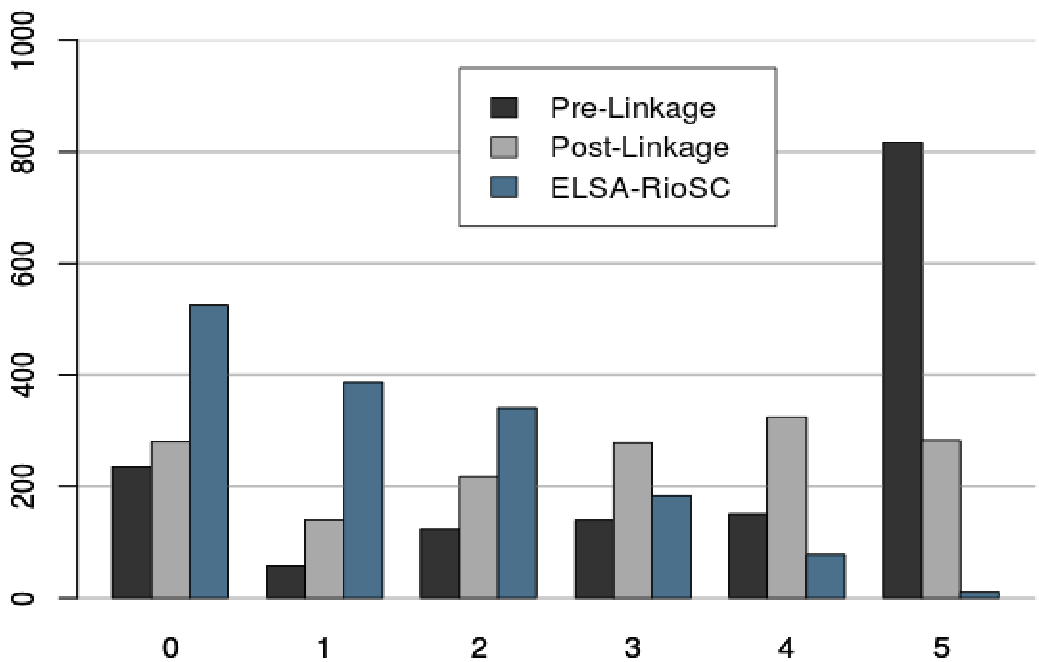
i. Data collection, linkage and network construction

The 1,521 active Fiocruz employees who were recovered in Wave 2 of ELSA listed a total of 5,405 names in response to the question: “Could you please tell us the complete names of your five closest friends [at work], in addition to the department and sector they work in?”. The 5,405 pre-linkage edge total included repeated names, mis-remembered or mis-recorded names, and names of individuals who worked on the Fiocruz campus but were not employed by the institute. Given the free-recall nature of the question, and the limited extent of identifying data (name and work department), even some names meant to refer to individuals employed by Fiocruz could not be linked with sufficient confidence to exactly one employee. The linkage process was therefore not expected to match all of the participant-reported names with Fiocruz employee names. Ultimately, 4,147 reasonable name matches were found (76.7% of reported friendships), again counting a name multiple times if it was cited by multiple individuals.

Of the 4,147 post-linkage friendships, or edges, 1,973 (47.6%) were between Wave 2 active employees, and these comprised the ELSA-RioSC edge list. 1,189 of the 1,521 individuals in ELSA-RioSC were connected by an edge to at least one other individual in the network, leaving 332 isolates (21.8% of the network). Of those 1,189 individuals, 258 were cited as a friend but did not cite any friends in ELSA-RioSC, 193 cited a friend in ELSA-RioSC but were not cited as a friend, and the remaining 738 both cited friends within the network and were cited by others as a friend.

Figure 9 shows the out degree distribution pre-linkage, post-linkage, and post-restriction to ELSA-RioSC. While most ELSA participants who gave responses to the “Social Networks” section of the questionnaire provided five names, the distribution shifted post-linkage, as not all cited friends could be matched to Fiocruz employee names. Once the population of accepted friends was limited to actively employed Wave 2 participants, the distribution became inverted, with most individuals having one or two cited friends in the network, and very few having five. The ELSA-RioSC out-degree, shown in blue in Figure 9, is discussed further in the following section, as are several other graph characteristics.

Figure 9 – The shift in the out-degree distribution from pre-linkage to post-linkage to sociocentric restriction

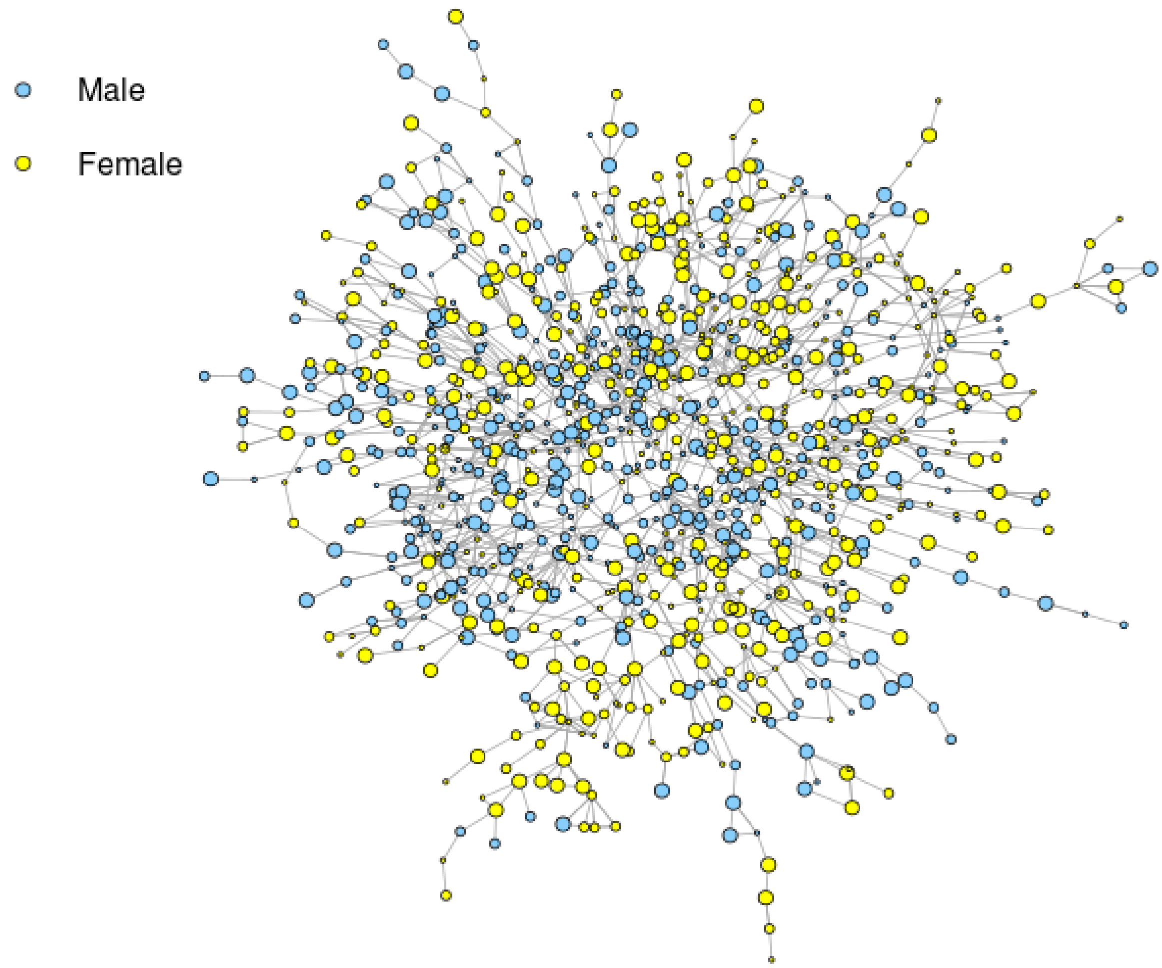


The figure shows the shift in the out-degree distribution from pre-linkage to post-linkage (from black to gray), and, following the linkage process, the shift with the restriction from the egocentric network to the sociocentric ELSA-RioSC network (from gray to blue).

ii. Exploratory network characteristics

ELSA-RioSC is a directed graph with 1,521 vertices and 1,973 edges, corresponding to an average undirected degree for each node of 2.59. The reciprocity of the directed edges, defined in the "Theoretical Foundations" section (*IV.iii*), was relatively high at 0.386, indicating that friendships were often reciprocated, or mutual. The graph density, as defined by Equation [5], was low (0.00085). Given the fixed choice question design, however, the maximum possible graph density – if each of the 1,521 members of ELSA-RioSC had cited exactly five other ELSA-RioSC members as friends – was only 0.0066, as noted in Section *IV.iii*. The observed density was therefore 13.0% of the maximum density. The degree of transitivity in ELSA-RioSC– as measured by the global clustering coefficient (Equation [6]) – was 0.198. The average undirected path length for vertices in the weak giant component, which included 70.0 % of ELSA-RioSC vertices, was 12.6. The weak giant component is shown visually in Figure 10, with the vertices colored by sex and with the sizes of the vertices proportional to categorical BMI. The in-degree, out-degree, and undirected degree distributions are shown in Figure 11, and the basic network characteristics are summarized in Table 3.

Figure 10 – The weakly connected giant component of ELSA-RioSC



The colors of the nodes correspond to the sexes of the participants, while the sizes of the nodes are proportional to BMI categories (normal weight, overweight, or obese).

Figure 12 shows the average degree of an individual's nearest neighbors by the individual's own degree, where the nearest neighbors are defined as those nodes to which an ego is connected, regardless of the directionality of that connection. Using degree as a measure of connectivity, the most connected individuals (those with the highest degrees) tended to have more highly connected friends than the least connected individuals. For an increase in one in an individual's own degree, the mean degree of his or her nearest neighbors increased, on average, by 0.190.

Figure 11 – ELSA-RioSC out-degree, in-degree and undirected-degree distributions

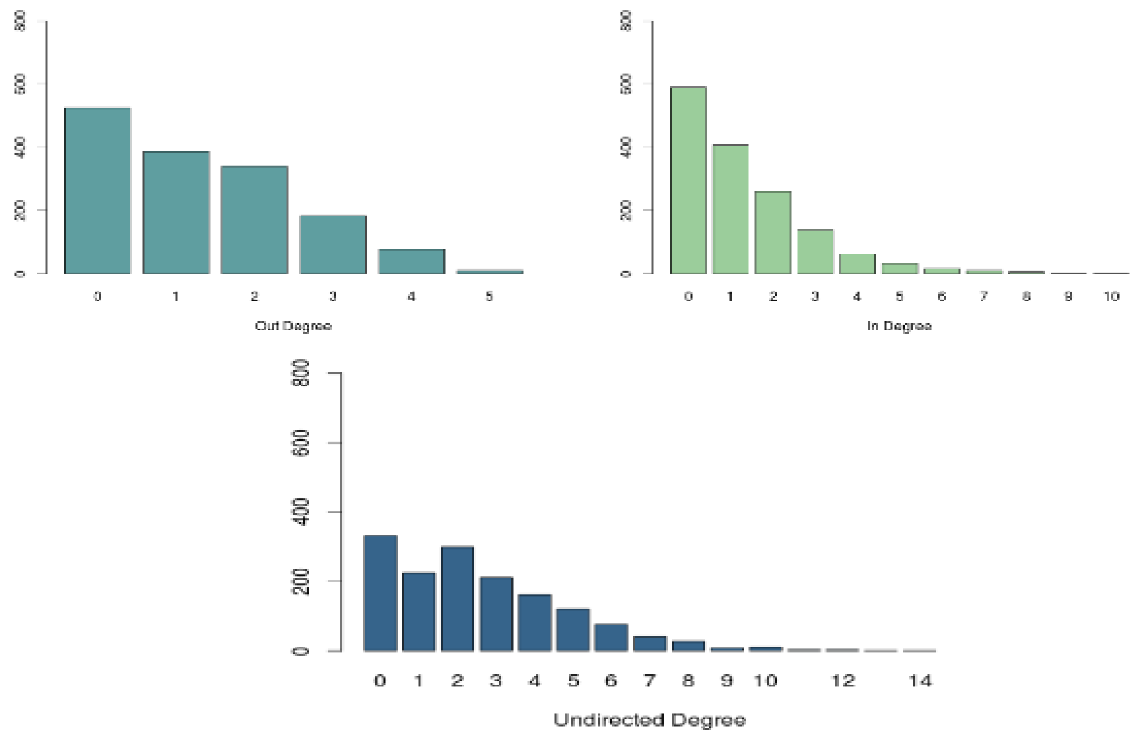
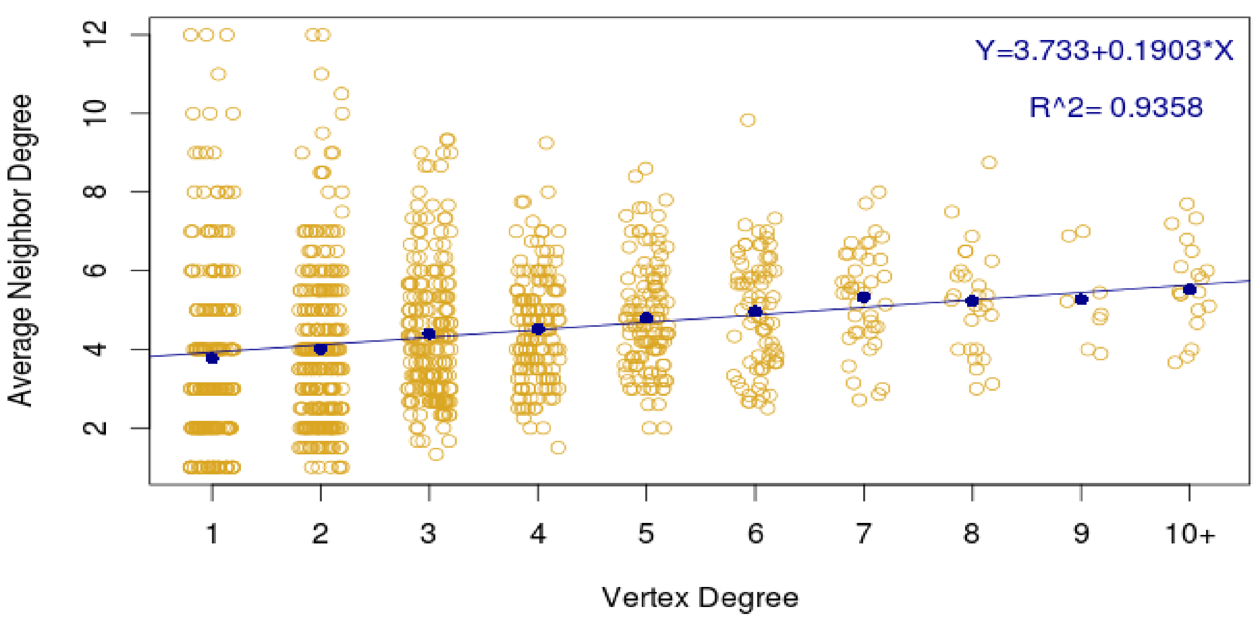


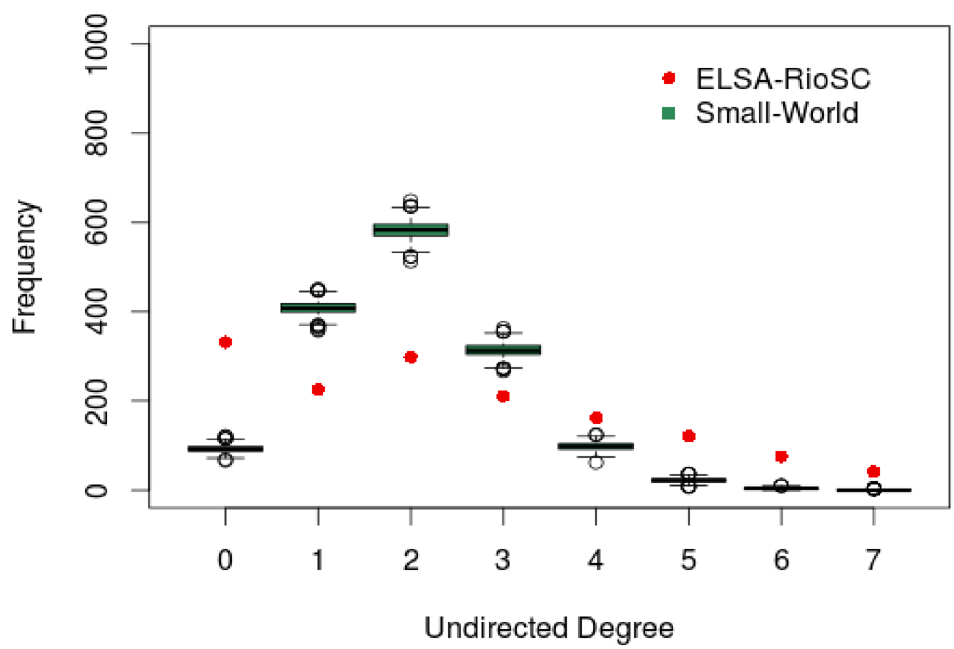
Figure 12 – Average alter degree by ego degree



The blue trend-line shows that individuals with higher degrees (undirected) were more likely to have well-connected friends than individuals with lower degrees.

The graph characteristics were compared to those of various random graph models through Monte Carlo methods, as described in the Methods section. The degree number and edge number of the Erdős–Rényi models were defined by the ELSA-RioSC values, so the density in these models was equal to the observed graph density. The algorithms for the Barabási-Albert and Watts-Strogatz models did not allow for precise density specification, but parameters were adjusted to approximate the observed density. The density of the Barabási-Albert models was 0.000657, and the density of the Watts-Strogatz models was 0.00132. Furthermore, the parameters of the Watts-Strogatz algorithm were adjusted so that the model produced undirected degree distributions that approximated the ELSA-RioSC undirected degree distribution. As noted in the Methods section, this was accomplished through a K value of two and a rewiring probability of 0.35. Even with these adjustments, the algorithm did not produce degree distributions that corresponded well to the observed data, as shown in Figure 13.

Figure 13 – Observed undirected-degree frequencies relative to the distribution of undirected degree frequencies in 1,000 iterations of the Watts-Strogatz model

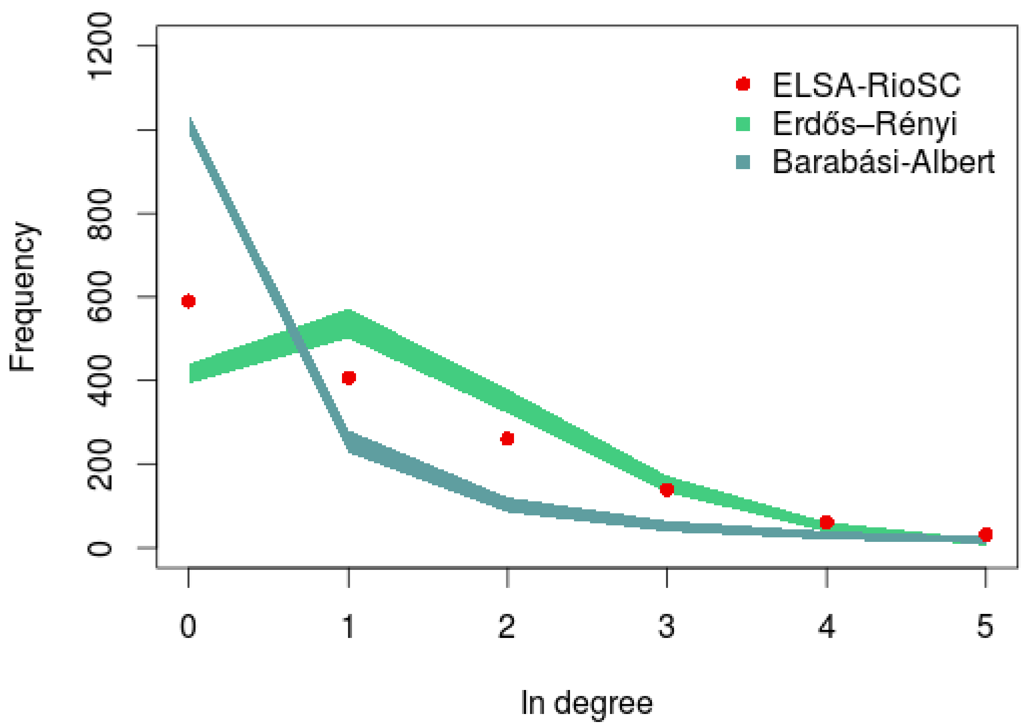


The observed frequency of undirected degrees (in red) and the distribution of undirected degrees from 1,000 iterations of the Watts-Strogatz small-world algorithm with K=2 and a rewiring probability of 0.35 (in green). While these model parameters were found to most reasonably approximate the observed data, the figure indicates that it is extremely unlikely that ELSA-RioSC resulted from a small-world mechanism.

For the two algorithms that allowed for the generation of directed graphs – namely, the Erdős–Rényi and Barabási-Albert models – generated in-degree frequencies were compared with the observed ELSA-RioSC in-degree distribution. Figure 14 shows the intervals that captured 95% of the generated frequency values for each in-degree for the Erdős–Rényi (classical random graph) and Barabási-Albert (preferential attachment) models in green and blue respectively. The observed in-degree frequencies of ELSA-RioSC, shown in red in Figure 14, generally did not fall within those confidence intervals, but at higher in-degrees (3+), the observed in-degree distribution fell within the range of relatively likely Erdős–Rényi values.

The high degree of reciprocity observed in ELSA-RioSC (0.386) was well outside of the interval capturing 95% of the degrees of reciprocity for the Erdős–Rényi models: [0, 0.00304]. By algorithm design, Barabási-Albert models contain no reciprocated friendships.

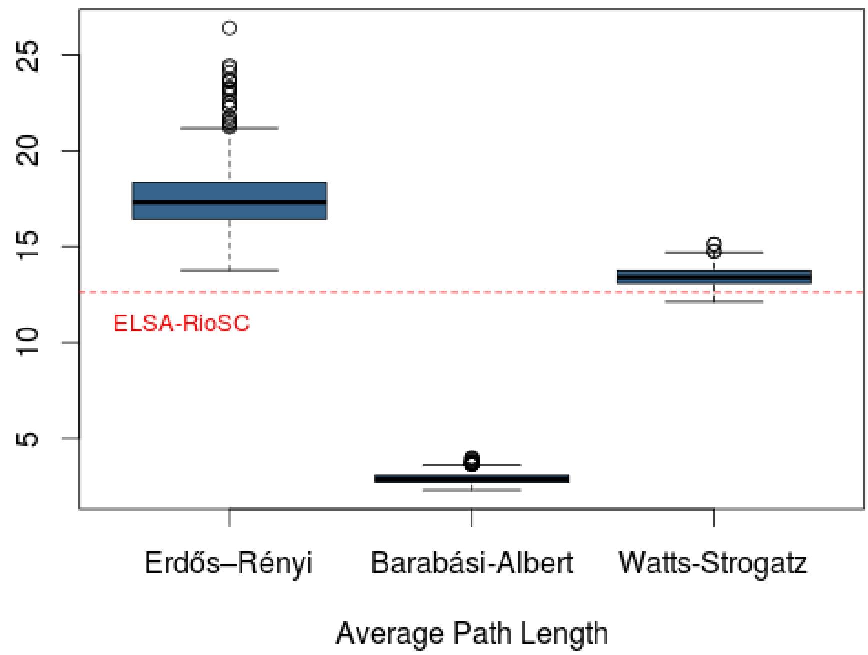
Figure 14 – Observed in-degree frequencies relative to the distributions of in-degree frequencies in iterations of the Barabási-Albert and Erdős–Rényi models



Comparison of the ELSA-RioSC observed in-degree frequencies (in red) with the ranges that included 95% of the frequencies generated through Monte Carlo methods with the Erdős–Rényi (in green) and Barabási-Albert (in blue) algorithms.

The statistics calculated under the interpretation of all ELSA-RioSC edges as undirected – namely the average path length, and the global clustering coefficient – further demonstrated that none of the three relatively simplistic random graph models described the observed data appropriately. The transitivity in ELSA-RioSC, 0.198, was again well outside of the 95% interval for all of the random graph models: [0, 0.00402], [0.00235, 0.00235], and [0, 0.00254] for the classical, preferential attachment, and small-world models, respectively. The ELSA-RioSC average path length (12.6) fell within the relatively likely range of values for the small-world model, but not for either of the other two model types, as shown in Figure 15.

Figure 15 – Observed average path length relative to the distributions of average path lengths in iterations of three random graph models



The figure shows the ELSA-RioSC average path length – 12.6, denoted by the red dashed line – relative to the distributions of average path lengths generated with three random graph model algorithms, as labeled on the x-axis.

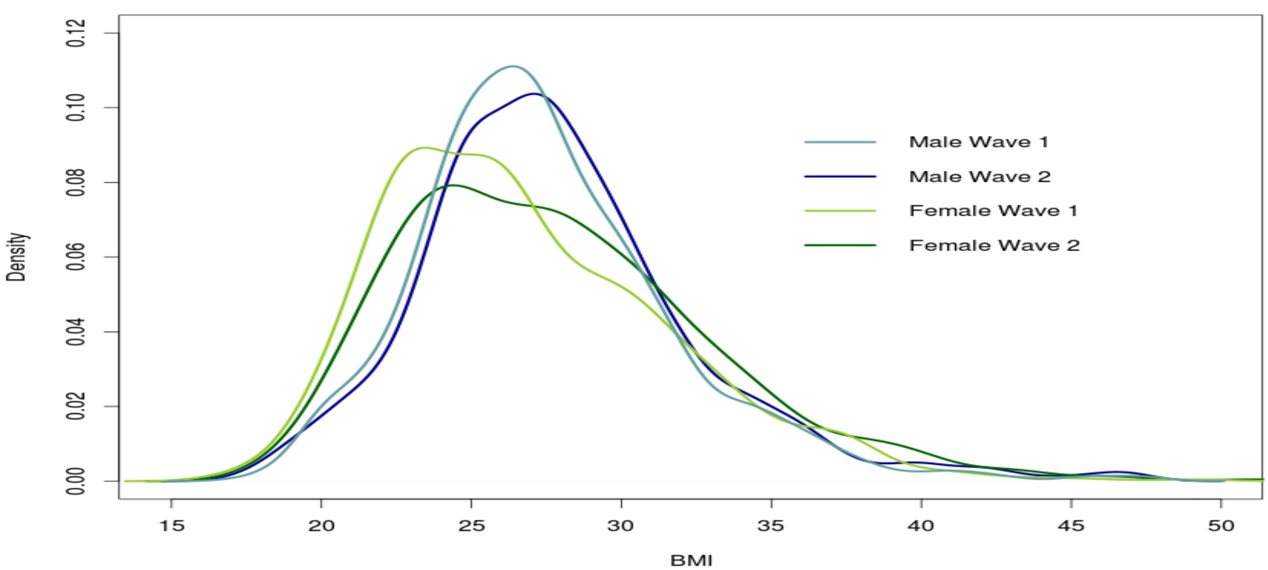
Table 3 – Summary characteristics of ELSA-RioSC		
General	*Type*	Directed, sociocentric
	Vertices	1,521
	Edges	1,973
Centrality	Average degree	2.59
Density	Graph density	0.000853
	Global clustering coefficient	0.198
	Reciprocity	0.386
Connectivity	Proportion of vertices in weak giant component	0.700
	Average path length	12.6

In summary, the observed in-degree was best approximated by the Erdős–Rényi model, although that approximation was only reasonable for in-degrees above two, and the average path length was best approximated by the Watts-Strogatz model. None of the models appropriately approximated more than one of the relevant network statistics. It was therefore deemed very unlikely that ELSA-RioSC developed via a a mechanism similar to those used to generate these three types of random graph models. These results justified the development of an ERGM that might better fit the data. The ERGM results are described below.

iii. Vertex attributes

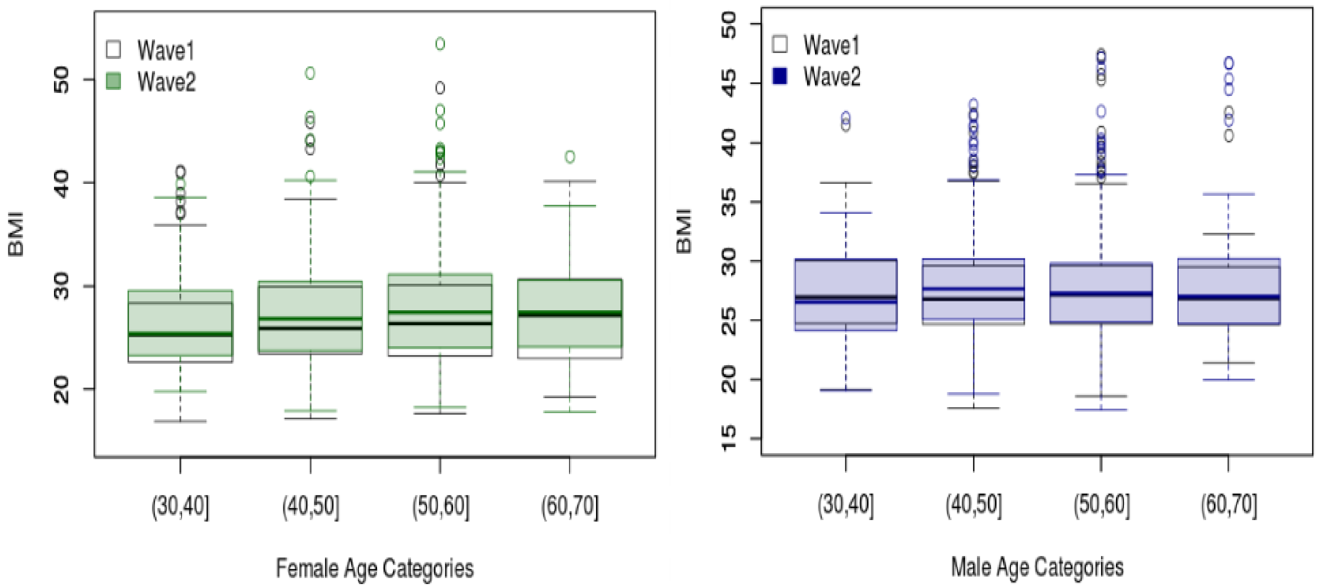
The distributions of attributes within the ELSA-RioSC network are summarized in Table 4. The distribution of BMI among men was significantly different than among women, as shown in Figure 16, and individuals' BMIs generally increased from Wave 1 to Wave 2, as shown in Figure 8. The unidirectional paired Wilcoxon test gave p-values of less than 2.2e-16 for both the male and female BMI data, leading to rejection of the null hypotheses that, on average, BMIs did not significantly increase from Wave 1 to Wave 2. Figure 17 shows that this increase was not simply due to an aging cohort; the distribution of BMI for both sexes was not significantly different across age groups at either Wave 1 or Wave 2, and within age groups, the BMI distribution either did not change or else shifted slightly towards higher values from the first wave to the second.

Figure 16 – Male and Female BMI probability densities at Waves 1 and 2



The Wave 1 and Wave 2 BMI distributions for both males (blues) and females (greens). For both sexes, the BMI distribution shifted significantly to the right from Wave 1 to Wave 2.

Figure 17 – Wave 1 and Wave 2 BMI distributions by age group for each sex



The figure shows that the BMI distributions either remained relatively constant or shifted towards higher BMIs within all age groups for both males and females.

Table 4 provides a summary of the distributions of each of the vertex attributes by category, stratified by sex. ELSA-RioSC was 48.4% male and 51.6% female. At Wave 1, 64.9% of participants were either overweight or obese, and by Wave 2, 69.4% of participants fell into one of those two categories. The proportion of participants who were obese also rose from 23.5% in Wave 1 to 28.2% in Wave 2. A smaller proportion of females than males were overweight at both Wave 1 and Wave 2, but a much greater proportion of females than males experienced a "large gain" in BMI – defined as a 2-20% average annual increase - between the first and second Waves (10.2% and 24.3% for males and females, respectively). Whereas the proportions of males and females who were obese at Wave 1 were relatively similar (23.0% and 24.1%, respectively), by Wave 2, 30.2% of females were obese compared to only 26.0% of males. Even so, at Wave 2 a greater percentage of males were either overweight or obese (73.7%) than females (65.1%), owing to the significantly greater proportion of overweight males.

Table 4 – Summary of the categorical distributions of ELSA-RioSC nodal attributes by sex			
Variable	Male	Female	Total
	n = 748	n = 773	N = 1521 (100%)
Wave 2			
Weight Status[1] *			
Underweight or Normal	196 (26.3)	269 (34.9)	465 (30.7)
Overweight	355 (47.7)	269 (34.9)	624 (41.2)
Obese	194 (26.0)	233 (30.2)	427 (28.2)
Age Group			
(30,45]	144 (19.3)	159 (20.6)	303 (19.9)
(45,55]	351 (46.9)	369 (47.7)	720 (47.3)
(55,75]	253 (33.8)	245 (31.7)	498 (32.7)
Family Income Class[101]			
≤R$1,244	114 (15.2)	59 (7.63)	173 (11.4)
R$1,245-2,487	279 (37.3)	226 (29.2)	505 (33.2)
R$2,488-3,741	177 (23.7)	219 (28.3)	396 (26.0)
≥R$3,742	178 (23.8)	269 (34.8)	447 (29.4)
Education Level			
No college	242 (32.4)	153 (19.8)	395 (26.0)
College +/- Specialization	235 (31.4)	252 (32.6)	487 (32.0)
Graduate degree	271 (36.2)	368 (47.6)	639 (42.0)
Wave 1			
Weight Status[1]			
Underweight or Normal	218 (29.1)	316 (40.9)	534 (35.1)
Overweight	358 (47.9)	271 (35.1)	629 (41.4)
Obese	172 (23.0)	186 (24.1)	358 (23.5)
Family Income Class[4]			
≤R$1,244	232 (31.0)	145 (18.8)	377 (24.8)
R$1,245-2,487	323 (43.2)	383 (49.5)	706 (46.4)
R$2,488-3,741	121 (16.2)	135 (17.5)	256 (16.8)
≥R$3,742	72 (9.63)	110 (14.2)	182 (12.0)

CONTINUED ON FOLLOWING PAGE

Variable	Male	Female	Total
Wave 1 → 2			
Average annual percent change in BMI *****			
Large loss (-20, -2]	39 (5.23)	31 (4.02)	70 (4.62)
Small loss (-2, -0.5]	120 (16.1)	98 (12.7)	218 (14.4)
No change (-0.5, 0.5]	232 (31.1)	176 (22.8)	408 (26.9)
Small gain (0.5, 2]	278 (37.3)	279 (36.2)	557 (36.7)
Large gain (2, 20]	76 (10.2)	187 (24.3)	263 (17.3)
Average income change per year			
Loss (-25, -5]	51 (6.81)	52 (6.73)	103 (6.77)
No change (-5, 5]	170 (22.7)	137 (17.7)	307 (20.2)
Small gain (5, 15]	220 (29.4)	238 (30.8)	458 (30.1)
Medium gain (15, 25]	152 (20.3)	170 (22.0)	322 (21.2)
Large gain (25, 500]	155 (20.7)	176 (22.8)	331 (21.8)

*****The 5 individuals (3 male, 2 female) for whom Wave 2 BMI was unavailable and could not be reasonably estimated were omitted from these sections. They therefore total to 1,516, whereas the categories of every other variable total to 1,521.

Of the 1,521 individuals in ELSA-RioSC, 332 were isolates, meaning they were not connected by an edge to any other individual in the network. The average BMIs of both the 172 male and 160 female isolates were greater than the average BMIs of the non-isolates, but possibly not significantly so, with one-sided t-test p-values of 0.110 and 0.360 for males and females, respectively. Income, age, and education level were similar for the isolates and the non-isolates.

iv. Generalized linear models and latent Gaussian models

Sectional Wave 2 BMI

As described in the Methods section, generalized linear modeling was used to determine which variables significantly improved the sectional Wave 2 BMI model fits for both males and females. Models with Gaussian and gamma probability distributions (with identity link functions) were considered, and analysis of the residuals and comparison of the AICs of the models indicated that the gamma distribution was more appropriate for both the male and female data. Latent Gaussian models for each sex were then fit with those variables determined through GLM to be important, again using a gamma distribution with an identity link function, and the models with no latent effects were compared to models with random latent effects, network latent effects, and both network and random latent effects.

Table 5 shows the results for the crude model – where each variable was considered in isolation – and for the final model. The crude coefficients are those of the generalized linear models, whereas the final model coefficients are those of the Bayesian latent Gaussian models with random latent effects, estimated through integrated nested Laplace approximation. The models with network latent effects were deemed inappropriate and rejected, as discussed below. Figures 18 and 19 show the quality of the fits of the final generalized linear models for males and females, respectively, before the conversion to latent Gaussian models and the inclusion of structured latent effects. In both figures, the “Normal Q-Q” plots show that the residuals were not normally distributed at relatively extreme values (2 or more standard deviations from the mean). The Shapiro-Wilk test for the residuals of the final generalized linear models gave a p-value of 1.98e-14 for the male model and a p-value of 2.54e-13 for the female model, suggesting that the GLM assumption of normally distributed residuals was not met in either case.

Table 5 – Results for the sectional models of Wave 2 BMI

Variables	Male		Female	
	Crude Coefficient (Stand. Dev.)	Final Coefficient (Stand. Dev.)	Crude Coefficient (Stand. Dev.)	Final Coefficient (Stand. Dev.)
Intercept	27.8 (0.167) ***	26.6 (1.343) ***	27.8 (0.188) ***	30.1 (1.79) ***
Age	0.00304 (0.0231)	0.006 (0.0225)	0.0274 (0.0268)	0.0034 (0.0266)
Categorical income				
≤1244	---------------------	---------------------	---------------------	---------------------
(1244, 2487]	0.0734 (0.504)	0.210 (0.494)	-2.286(0.819)***	-1.583 (0.778) **
(2488, 3741]	0.944 (0.551) **	1.01 (0.540) **	-3.358 (0.816) ***	-1.970 (0.807) **
≥3742	-0.513 (0.539)	-0.355 (0.529)	-3.641 (0.801) ***	-2.03 (0.811) **
Education level				
No college	---------------------		---------------------	---------------------
Some college → College complete with specialization	0.0461 (0.419)		-1.942 (0.549) ***	-1.53 (0.561) ***
Graduate degree	0.123 (0.405)		-3.330 (0.512) ***	-2.62 (0.566) ***
Out degree				
0	---------------------		---------------------	
1	-0.381 (0.438)		-0.0459 (0.493)	
2	-0.289 (0.452)		-0.0996 (0.514)	
3+	0.0709 (0.490)		-0.398 (0.549)	

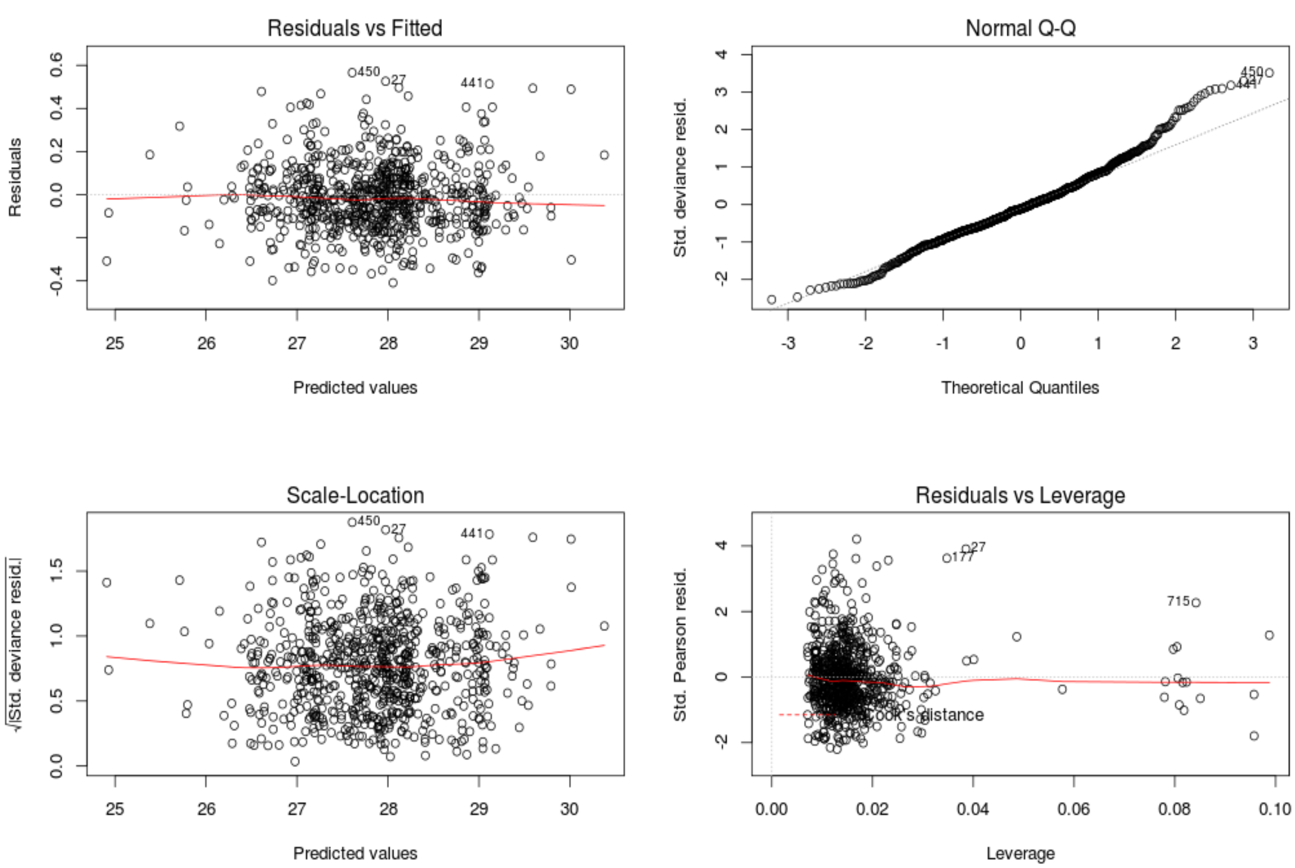
CONTINUED ON FOLLOWING PAGE

Variables	Male		Female	
	Crude Coefficient (Stand. Dev.)	Final Coefficient (Stand. Dev.)	Crude Coefficient (Stand. Dev.)	Final Coefficient (Stand. Dev.)
In degree				
0	---------------------		---------------------	
1	-0.108 (0.415)		-0.308 (0.482)	
2	-0.455 (0.503)		-0.416 (0.529)	
3+	-0.332 (0.482)		-0.304 (0.545)	
Undirected degree				
0	---------------------		---------------------	---------------------
1-2	-0.551 (0.455)		-0.186 (0.524)	0.221 (0.590)
3-4	-0.453 (0.488)		0.5614 (0.570)	1.11 (0.735) *
5+	-0.490 (0.530)		-0.991 (0.578) **	-0.232 (0.785)
Betweenness centrality	37.0 (32.0)	37.4 (31.9) *	-32.0 (33.6)	
Eigenvector centrality	-0.232 (2.08)		0.750 (6.50)	
Clustering coefficient	-0.694 (0.664)		0.415 (0.671)	
Undirected friend average BMI				
(1,25]	---------------------	---------------------	---------------------	
(25,30]	0.688 (0.483) *	0.906 (0.541) *	0.260 (0.529)	
(30,60]	1.08 (0.587) **	1.51 (0.648) **	0.411 (0.641)	
none	1.14 (0.546) **	1.09 (0.544) **	0.400 (0.614)	
In friend average BMI				
(1,25]	---------------------		---------------------	---------------------
(25,30]	0.600 (0.535)		1.18 (0.560) **	0.987 (0.536) **
(30,60]	1.27 (0.639) **		0.935 (0.674)*	0.195 (0.640)
none	0.885 (0.520) **		1.17 (0.549) **	0.944 (0.645) *
Out friend weighted average BMI				
(1,25]	---------------------		---------------------	
(25,30]	0.174 (0.499)		0.561 (0.547)	
(30,60]	0.518 (0.593)		0.530 (0.632)	
none	0.435 (0.494)		0.556 (0.540)	
Mutual friend average BMI				
(1,25]	---------------------		---------------------	
(25,30]	0.652 (0.672)		0.915 (0.685) *	
(30,60]	0.342 (0.792)		0.683 (0.789)	
none	0.811 (0.564) *		1.14 (0.552) **	
Proportion of friends cited who were overweight				
0	---------------------	---------------------	---------------------	
[1/5, 2/5]	-1.91 (1.21) *	-1.75 (1.22) *	-0.0421 (0.879)	
[1/2, 2/3]	0.319 (0.494)	0.00840 (0.571)	-0.425 (0.550)	
[3/4, 1]	-0.385 (0.370)	-0.794 (0.501) *	0.213 (0.431)	

*The superscripts in the tablet indicate significance based on the p-value of the Likelihood Ratio Test, where *** indicates p<0.01; ** indicates p<0.1; and * indicates p<0.2.*

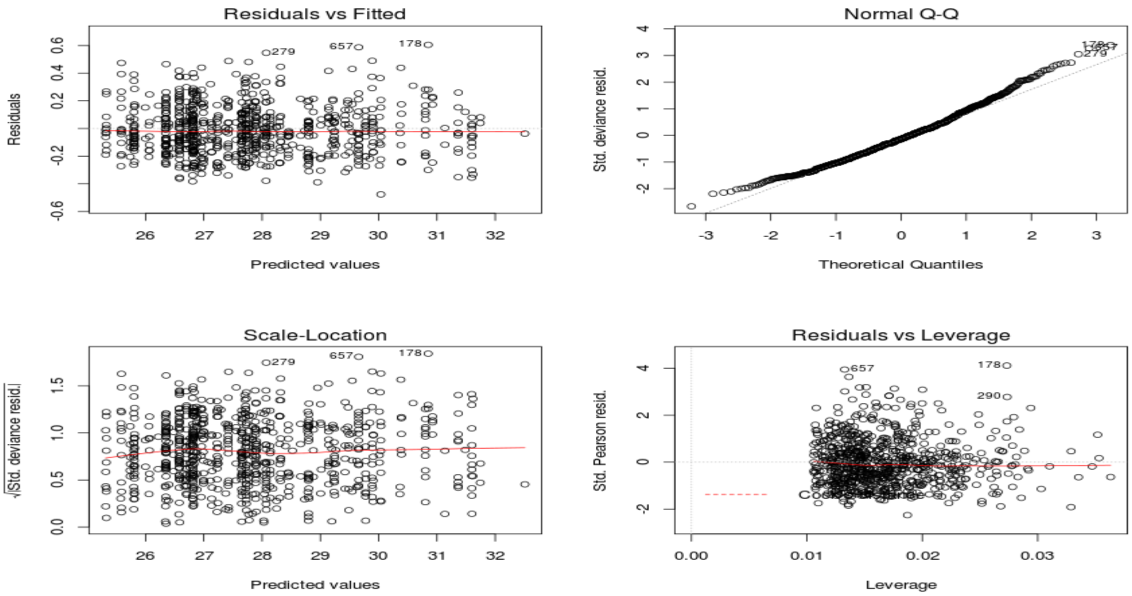
The change in the coefficient estimates was extremely small in the transition from the final generalized linear models to the latent Gaussian models with no latent effects. Figures 20 (male) and 21 (female) show the change in the coefficients from the latent Gaussian models with no latent effects to the models with random latent effects, network-structured latent effects, and both random and network latent effects. As those figures show, inclusion of a neighborhood structure in the latent effect significantly shifted the coefficient estimates and decreased the precision of those estimates, especially for network-related variables, like undirected degree in the female model and average BMI of undirected friends in the male model. The use of these network-structured latent effects was ultimately rejected. Network-based variables were already included in the final models, and to also include network information through latent effects was deemed redundant; it added little or no information and greatly reduced the precision of coefficient estimates.

Figure 18 – Analysis of the residuals of the final GLM male Wave 2 BMI model



The plots show that none of the points had an out-sized impact on the model fit, but also that the residuals were not normally distributed at the extremes (past two standard deviations from the mean).

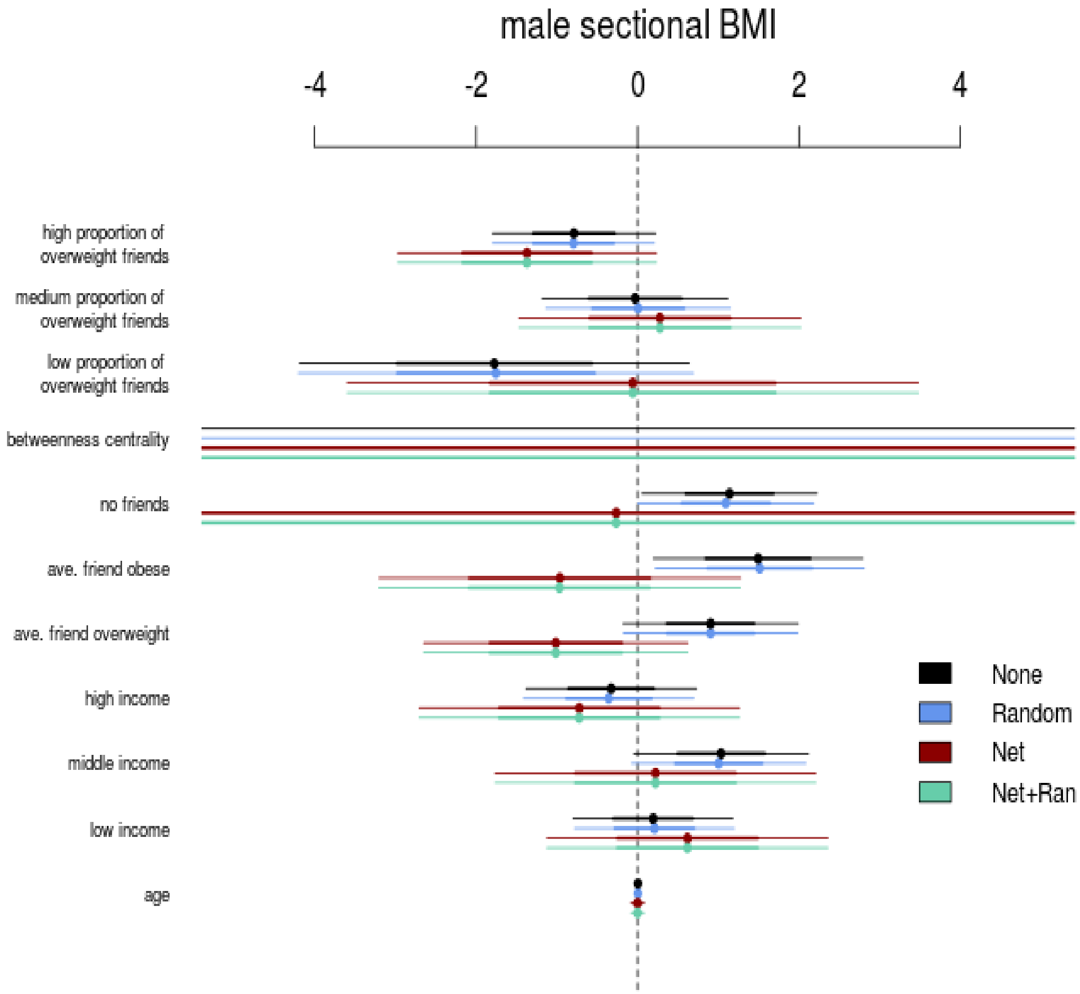
Figure 19 – Analysis of the residuals of the final GLM female Wave 2 BMI model



As with the male model, none of the points had an out-sized impact on the model fit, but again, the residuals were not normally distributed at the extremes (past two standard deviations from the mean). This was not as significant in the female model as in the male model.

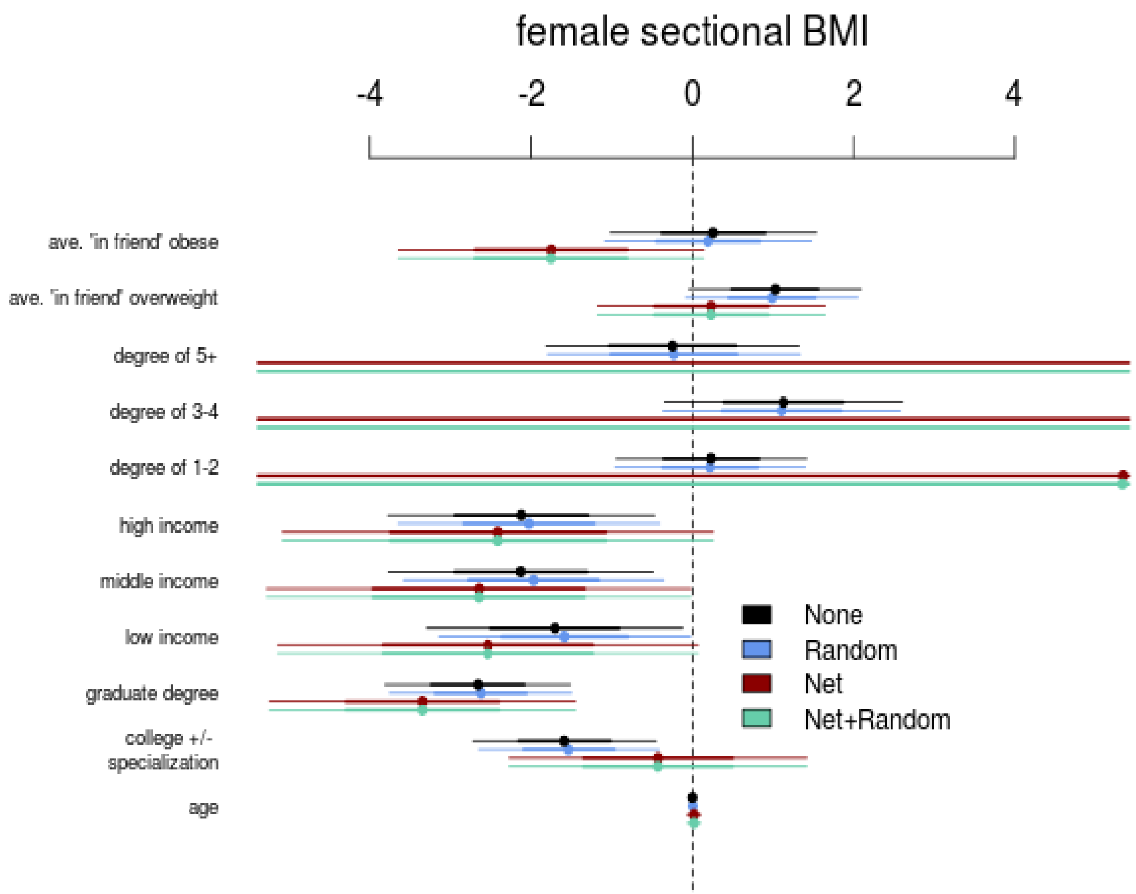
Figures 20 and 21 also show that the coefficient estimates changed only slightly with the inclusion of a latent independent and normally distributed random variable (the shifts from black to blue). Comparison of the WAICs of the models with and without this random latent effect indicated that both the male and female models benefited from its inclusion, with the male model WAIC improving from 4324 to 4284 and the female model WAIC improving from 4631 to 4598. The coefficient estimates in the “Final” model columns of Table 5 are therefore from Bayesian latent Gaussian models with random latent effects.

Figure 20 – Shifts in the latent Gaussian coefficient estimates for the male Wave 2 BMI model with the inclusion of latent field models



The change in the coefficient estimates for the male latent Gaussian models from a model with no latent effect (black) to one with random (blue), network (red), or both random and network (green) latent effects.

Figure 21 – Shifts in the latent Gaussian coefficient estimates for the female Wave 2 BMI model with the inclusion of latent field models



The change in the coefficient estimates for the female latent Gaussian models from a model with no latent effect (black) to one with random (blue), network (red), or both random and network (green) latent effects.

The final male model included the control variables age and income level; the network variable betweenness centrality; and the friend variables "undirected friend average BMI"and "proportion of overweight friends", as defined in Table 1. The final female model included the control variables age, education, and income level; the network variable undirected degree; and the friend variable "in-friend average BMI", again as defined in Table 1. For both males and females, age was not significantly associated with Wave 2 BMI, but as with sex, age is a biological property known to affect how one responds to energy consumption and expenditure. It was therefore included in all possible models, regardless of LRT p-values. The only other variable included in both the male and female final sectional models was income level. In women, income level and BMI were negatively correlated, with lower income levels generally corresponding to greater BMIs. The BMIs of those women in the highest income group were, on average, 2.03 kg/m^2 less than the BMIs of those women in the lowest income group , when all other variables were kept constant. The association between BMI and income for males was less clear: BMI was significantly greater only for those males in the upper-middle income group (compared with those in the lowest income group). Beyond age and income level, the final female model also included the additional control variable “education level”, which again showed an inverse relationship to BMI and was positively associated with income level: women with greater levels of education tended to earn more and weigh less.

In regards to network variables – variables measuring an individual's role or relative position in ELSA-RioSC – males with greater betweenness centralities tended to have greater BMIs, and females with intermediate undirected degrees tended to have greater BMIs than females with no ELSA-RioSC friends or five or more friends in the network. Neither of these associations was particularly significant – the p-value for the inclusion of betweenness centrality in the male model was 0.156 and the smallest p-value for inclusion of categorical undirected degree in the female model was 0.138 (for those individuals with a degree of 3 or 4) – but both indicated that, if anything, greater popularity or centrality in ELSA-RioSC was associated with greater weight.

The association between an individual's BMI and the BMI of his or her friends was more clear in the male model than in the female model. For males, an individual's BMI was positively associated with the BMI of his friends. If the mean of his friends' BMIs fell within the “obese” category, the individual had, on average, a BMI 1.51 kg/m^2 greater than if the mean of his friends' BMIs fell within the “normal” weight category, controlling for all other variables. Males with no friends in the network on average had a BMI 1.15 kg/m^2 greater than those with a normal “undirected friend average BMI”. The “in-friend average BMI” term in the female model indicated that if the mean BMI of the individuals who cited a woman as a friend fell in the overweight category, the BMI of the woman was, on average, 0.987 kg/m^2 greater than the BMI of women cited as friends by individuals whose average BMI fell in the normal weight category. The BMI of women not cited as a friend by anyone was also on average significantly greater (p-value 0.129) than the BMI of women in the baseline group (those with a “normal” in-friend average BMI).

Average Yearly BMI Change from Wave 1 to Wave 2:

The process for determining a final model for the male and female average yearly change in BMI was analogous to the process described in the previous section for the sectional models. In this case, comparison of the AICs and residual analyses indicated that for both the male and female data, a Gaussian probability distribution was more appropriate than a gamma distribution once the significant variables had been included. In Table 6, the variables in the “Crude” columns were therefore estimated using GLM with a Gaussian distribution and an identity link function. As with the sectional Wave 2 analysis, the “Normal Q-Q” plots showed that the residuals were not normally distributed at relatively extreme values, and Shapiro-Wilk testing of the residuals returned p-values of 3.17e-12 and 5.01e-11 for the male and female models, respectively. The GLM assumption of normally distributed residuals was thus again violated.

As in the sectional models, a latent Gaussian model was fit with the same variables, and again, the latent effects with network structures (spacial latent effects) were rejected, as they were deemed to contribute redundant information while greatly increasing the uncertainty of estimates, as shown in Figures 22 and 23 for the male and female models, respectively. Those figures also show that inclusion of a latent independent and normally distributed random effect had very little impact on the estimated coefficients, but the WAIC values improved from 2684 to 2644 for the male model and from 3028 to 3015 for the female model. The latent effect was therefore included in the final model, and the coefficients in the “Final” columns of Table 6 were estimated using INLA with a latent Gaussian model with a Gaussian probability distribution and an independent and normally distributed random latent effect.

Table 6 – Results for the average annual percent change in BMI from Wave 1 to Wave 2

Variables	Male		Female	
	Crude Coefficient (Stand. Dev.)	Final Coefficient (Stand. Dev.)	Crude Coefficient (Stand. Dev.)	Final Coefficient (Stand. Dev.)
Intercept	0.418 (0.0538) ***	0.539 (0.158) ***	0.956 (0.0623) ***	1.04 (0.153) ***
Categorical Age				
(30, 45]	---------------------	---------------------	---------------------	---------------------
(45, 55]	-0.234 (0.118)**	-0.261 (0.119) **	-0.176 (0.135) *	-0.175 (0.135) *
(55, 75]	-0.496 (0.154)***	-0.547 (0.155) ***	-0.686 (0.197) ***	-0.721 (0.197) ***
Categorical income				
≤1244	---------------------	---------------------	---------------------	
(1244, 2487]	0.0283 (0.127)	0.0449 (0.127)	0.127 (0.169)	
(2488, 3741]	-0.224 (0.165) *	-0.232 (0.169) *	0.154 (0.208)	
≥3742	0.297 (0.198) *	0.296 (0.202) *	-0.0702 (0.219)	
Average annual percent change in income	0.00306 (0.00203) *	0.00400 (0.00210) **	0.000471 (0.00196)	
Education level				
No college	---------------------		---------------------	
Some college → College complete with specialization	-0.0825 (0.135)		0.306 (0.178) **	
Graduate degree	0.101 (0.130)		0.141 (0.167)	
Out degree				
0	---------------------		---------------------	
1	-0.0116 (0.141)		-0.190 (0.163)	
2	-0.0920 (0.146)		-0.101 (0.170)	
3+	-0.201 (0.157)		-0.335 (0.182) **	
In degree				
0	---------------------		---------------------	
1	-0.0174 (0.133)		0.136 (0.159)	
2	-0.0687 (0.163)		-0.238 (0.175) *	
3+	-0.0525 (0.156)		-0.00520 (0.180)	
Undirected degree				
0	---------------------		---------------------	
1-2	-0.106 (0.146)		-0.231 (0.174) *	
3-4	-0.160 (0.156)		-0.0720 (0.187)	
5+	-0.0999 (0.170)		-0.345 (0.195) **	
Betweenness centrality	3.83 (10.1)		-9.49 (11.3)	
Eigenvector centrality				
0	---------------------		---------------------	
(0,2e-05]	-0.0587 (0.134)		-0.0649 (0.159)	
(2e-05,1.0]	-0.0259 (0.131)		-0.0358 (0.156)	

CONTINUED ON FOLLOWING PAGE

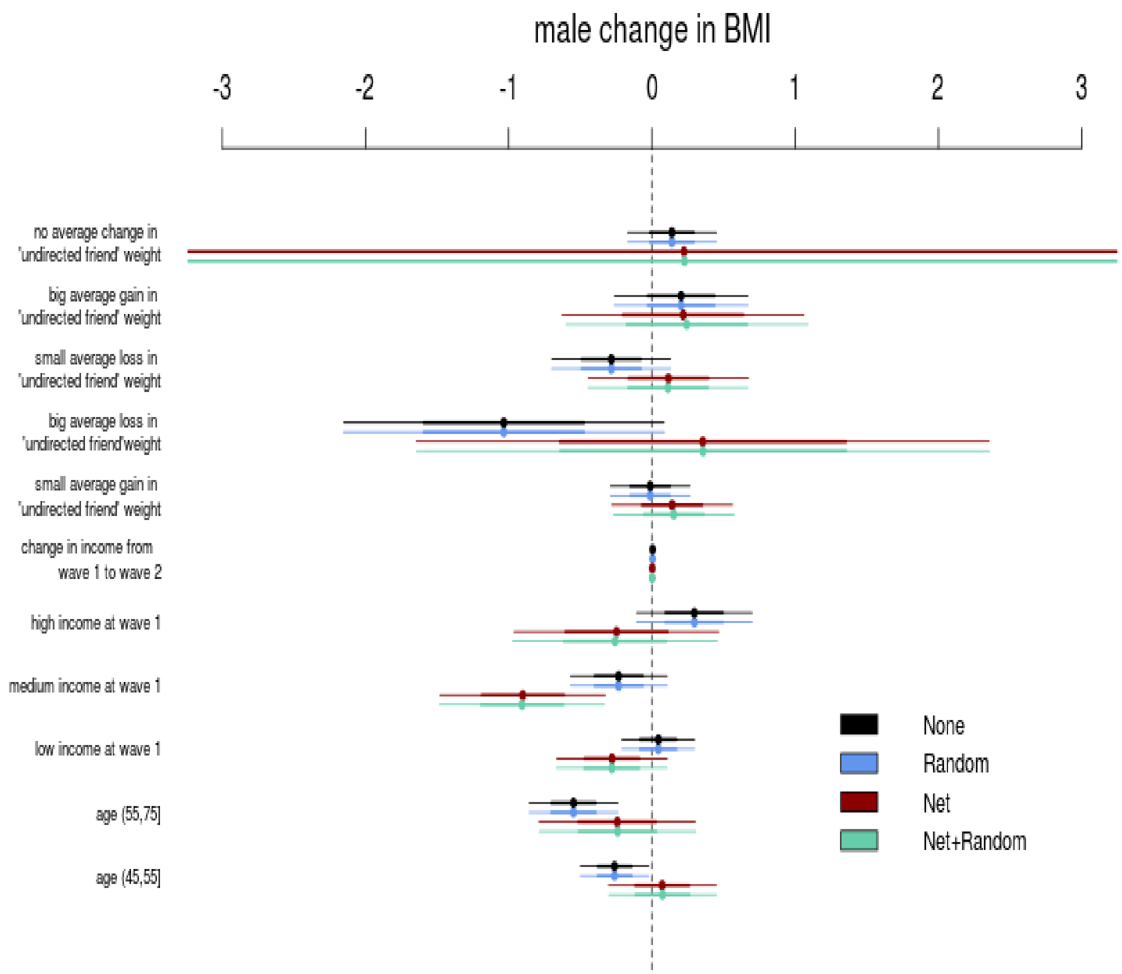
Variables	Male		Female	
	Crude Coefficient (Stand. Dev.)	Final Coefficient (Stand. Dev.)	Crude Coefficient (Stand. Dev.)	Final Coefficient (Stand. Dev.)
Clustering coefficient				
0	---------------------		---------------------	
(0,0.5]	-0.0168 (0.135)		-0.273 (0.153) **	
(0.5,1]	-0.137 (0.219)		0.00295 (0.212)	
Undirected friend average BMI				
(1,25]	---------------------		---------------------	
(25,30]	0.103 (0.148)		0.116 (0.160)	
(30,60]	0.161 (0.195)		-0.100 (0.219)	
none	0.209 (0.168)		0.260 (0.189) *	
Average annual percent change in undirected friend average BMI				
No change (-0.5, 0.5]	---------------------	---------------------	---------------------	
Small loss (-2, -0.5]	-0.357 (0.208) **	-0.284 (0.208) *	0.253 (0.267)	
Large loss (-20,-2]	-1.07 (0.564) **	-1.03 (0.560) **	-0.147 (0.562)	
Small gain (0.5, 2]	-0.0203 (0.140)	-0.0116 (0.139)	0.0220 (0.162)	
Large gain (2, 20]	0.237 (0.233)	0.204 (0.234)	-0.0168 (0.237)	
Isolate (No friends)	0.0786 (0.155)	0.140 (0.155)	0.240 (0.186) *	
In friend average BMI				
(1,25]	---------------------		---------------------	
(25,30]	0.254 (0.170) *		0.171 (0.180)	
(30,60]	0.253 (0.205)		0.117 (0.232)	
none	0.232 (0.164) *		0.126 (0.175)	
Average annual percent change in in friend average BMI				
No change (-0.5, 0.5]	---------------------		---------------------	
Small loss (-2, -0.5]	-0.224 (0.231)		-0.0590 (0.282)	
Large loss (-20,-2]	-0.829 (0.442) **		0.0324 (0.456)	
Small gain (0.5, 2]	0.00279 (0.164)		-0.121 (0.184)	
Large gain (2, 20]	0.0959 (0.254)		-0.170 (0.269)	
No in friends	-0.0000214 (0.151)		-0.0592 (0.173)	
Out friend average BMI				
(1,25]	---------------------		---------------------	---------------------
(25,30]	-0.0643 (0.153)		0.127 (0.173)	0.168 (0.172)
(30,60]	0.0339 (0.195)		-0.251 (0.217)	-0.263 (0.215)
none	0.0654 (0.151)		0.209 (0.170)	0.231 (0.169) *

CONTINUED ON FOLLOWING PAGE

Variables	Male: Crude Coefficient (Stand. Dev.)	Male: Final Coefficient (Stand. Dev.)	Female: Crude Coefficient (Stand. Dev.)	Female: Final Coefficient (Stand. Dev.)
Average annual percent change in out friend average BMI				
No change (-0.5, 0.5]	---------------------		---------------------	
Small loss (-2, -0.5]	-0.463 (0.215) **		0.319 (0.287)	
Large loss (-20,-2]	-0.792 (0.503) *		-0.0246 (0.540)	
Small gain (0.5, 2]	-0.221 (0.158) *		-0.0327 (0.182)	
Large gain (2, 20]	0.0751 (0.234)		-0.0928 (0.230)	
No out friends	-0.0726 (0.150)		0.198 (0.174)	
Mutual friend average BMI				
(1,25]	---------------------		---------------------	
(25,30]	-0.0891 (0.216)		-0.0217 (0.226)	
(30,60]	-0.0262 (0.260)		0.109 (0.275)	
none	0.104 (0.181)		0.00567 (0.181)	
Average annual percent change in mutual friend average BMI				
No change (-0.5, 0.5]	---------------------		---------------------	
Small loss (-2, -0.5]	-0.275 (0.295)		0.299 (0.354)	
Large loss (-20,-2]	-0.195 (0.491)		-0.185 (0.605)	
Small gain (0.5, 2]	-0.289 (0.219) *		0.0445 (0.242)	
Large gain (2, 20]	0.281 (0.345)		-0.274 (0.307)	
No mutual friends	0.0204 (0.172)		-0.00879 (0.195)	
Proportion of friends cited who were overweight				
0	---------------------		---------------------	
[1/5, 2/5]	-0.317 (0.294)		-0.221 (0.285)	
[1/2, 2/3]	-0.0112 (0.156)		-0.215 (0.179)	
[3/4, 1]	-0.0926 (0.123)		-0.137 (0.145)	

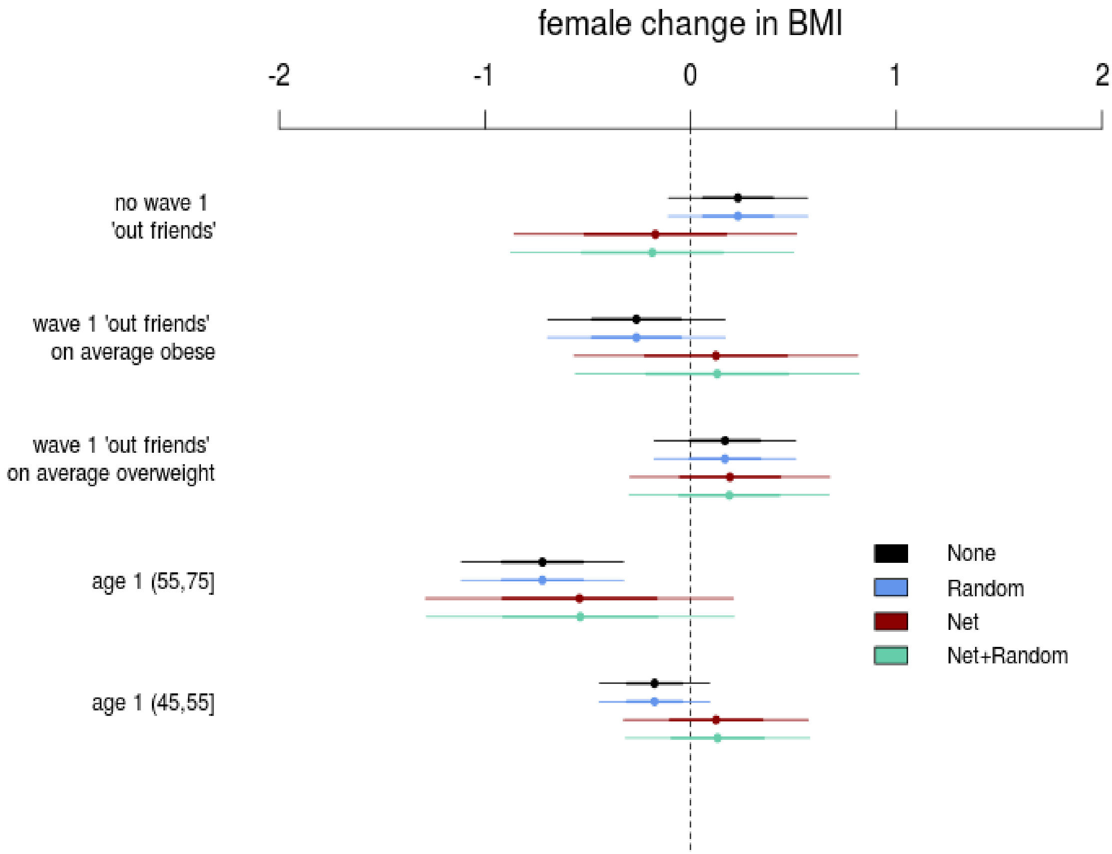
*The superscripts in the table indicate significance based on the p-value of the Likelihood Ratio Test, where *** indicates p<0.01; ** indicates p<0.1; and * indicates p<0.2.*

Figure 22 – Shifts in the latent Gaussian coefficient estimates for the male average annual percent change in BMI model with the inclusion of latent field models



The figure shows that there was very little difference in estimates with the inclusion of an independent and normally distributed random latent effect (black to blue and red to green), whereas the shift and increase in uncertainty with the inclusion of network structure was significant (blue to green and black to red). The coefficients in blue represent those included in the "Final" column of Table 6.

Figure 23 – Shifts in the latent Gaussian coefficient estimates for the female average annual percent change in BMI model with the inclusion of latent field models



As with the male data, the figure shows a significant shift in the female model coefficient estimates with the inclusion of network-structured latent effects but not with independent and normally distributed random latent effects. Again, the blue estimates represent the coefficients ultimately selected for the final female model.

Age was the first variable to enter both the male and female models, again because of the biological precedent, but whereas in the sectional models that inclusion was not significant, it was significant in the models for change in BMI. Both males and females in the middle and upper age groups on average experienced a lesser percent annual increase in BMI than their peers in the lowest age group. In the case of males aged 55 to 75, there was even a slight annual percent decrease in BMI on average, as the absolute value of the negative coefficient (-0.547) was greater than that of the positive intercept (0.539). Income variables entered the final male model but not the final female model: relative to the lowest income category, individuals from families with average monthly per capita incomes in the high-intermediate range of R$ 2,488-3,741 showed a tendency to gain less weight, whereas those in the highest income category showed a tendency to gain more weight.

Additionally, as the average annual change in per-capita monthly family income increased for a male, he became more likely to have also experienced a more positive change in weight. According to the model, those men who experienced a 10% average annual increase in income would be expected to also have experienced, on average, a mean annual BMI change 0.031 percentage points more positive than the BMI change in men who experienced no alteration in income.

No network position variables entered either of the final models, but each included one variable pertaining to friends' weight statuses. For the female model, that variable was the out-friend average BMI at Wave 1: women who cited no friends in the network tended to experience greater increases in BMI than women who cited friends with, on average, normal BMIs. For the male model, that variable was the average annual percent change in undirected friend BMIs: men whose friends, on average, experienced large annual percent losses in weight also tended to experience less positive (or even negative) changes in BMI compared to men whose friends, on average, experienced little-to-no change in BMI. The direction of this correlation held for weight gain – men whose friends, on average, experienced large annual percent increases in BMI also tended to experience greater increases in BMI – but the p-values only indicated significance for the coefficients related to weight-loss in friends.

v. Exponential random graph model – exploratory analysis

Results from the three exponential random graph models are summarized in Table 7. In the models including only exogenous effects and in those including both endogenous and exogenous effects, two individuals with relatively similar BMIs – within 4 kg/m^2 of one another – were not significantly more likely to become friends than individuals with less similar BMIs. Different approaches to comparative BMI were tested – namely, matched categorical BMIs (normal weight and underweight, overweight, and obese) and matched obesity status (obese/non-obese) – and those parameters also did not significantly improve the model fit. Even in a model with only the “small BMI difference” parameter (and no controls for other exogenous or endogenous effects) the parameter proved insignificant (p-value 0.698). The ERGM results therefore indicated that similarity in BMI was not a significant factor in predicting friendships in ELSA-RioSC.

Model 1 showed that, of the exogenous parameters, if two individuals were within 10 years of one another in age, if they were of the same sex, if they worked in the same department, or if they had approximately the same level of education , they were significantly more likely to become friends. If individuals were of the same sex, for instance, the probability of a friendship existing between them increased by 159% (exp(0.950) = 2.59), given no change in the other statistics. The probability increased even more if individuals worked in the same department. The Model 3 results indicate that these exogenous factors remained significant even after controlling for endogenous effects. In the final model, the endogenous parameters most significantly affecting the probability of edge formation were those controlling for "mutual" and transitive (GWESP) effects, but the "isolates" parameter also proved significant. As in Model 1, in Model 3 the most significant exogenous parameter was shared work department.

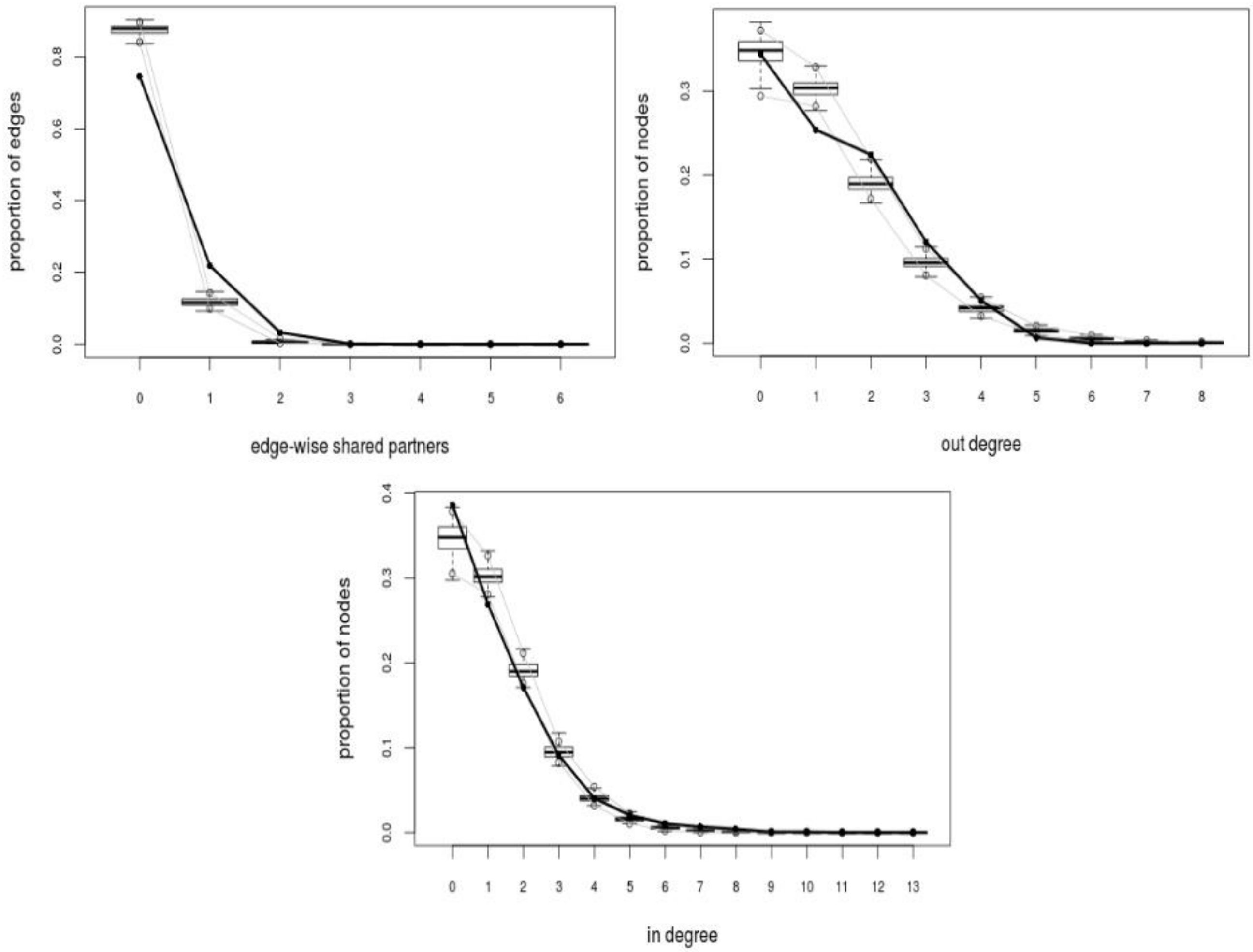
Table 7 – Exponential random graph model results

Variables	Model 1: Exogenous effects	Model 2: Exogenous & mutual/isolates effects	Model 3: Complete model
Edges	-9.68 (0.0801) ***	-9.25 (0.106) ***	-9.01 (0.104) ***
Small BMI difference	-0.00730 (0.0453)	0.0334 (0.0593)	-0.00776 (0.0718)
Node sex match	0.950 (0.0512) ***	0.705 (0.0685) ***	0.656 (0.0735) ***
Node department match	3.02 (0.0500) ***	2.36 (0.0682) ***	2.27 (0.0707) ***
Small age difference	0.596 (0.0575) ***	0.514 (0.0855) ***	0.380 (0.0848) ***
Node education match	0.979 (0.0465) ***	0.788 (0.0735) ***	0.598 (0.0624) ***
Isolates	-------------------	0.495 (0.0833) ***	0.289 (0.0975) ***
Mutual	-------------------	5.505 (0.193) ***	4.77 (0.441) ***
Geometrically weighted edgewise shared partners distribution	-------------------	-------------------	1.13 (0.190) ***

*For the superscripts in the table, *** indicates p<0.01; ** indicates p<0.1; and * indicates p<0.2.*

Figure 24 indicates that Model 3 much more appropriately described the observed ELSA-RioSC network than did the three algorithm-based random graph models from the *Exploratory network characteristics* section. The observed values of edgewise shared partners, in-degrees, and out-degrees generally fell within the distributions of relatively common values generated from the estimated parameters. This was not the case for low values of edgewise shared partners, with the model overestimating the number of isolated edges and underestimating the number of edges sharing only one partner, but the shape of the edge-wise shared partners distribution still corresponded relatively well to the observed distribution.

Figure 24 – Analysis of the goodness-of-fit of the final ERGM model



Goodness-of-fit of the complete ERGM (Model 3) for several graph characteristics. Starting in the upper left corner and moving clockwise, the graphs show the edge-wise shared partner distribution, the out-degree distribution, and the in-degree distribution. The boxplots in each figure indicate the distributions of the proportion of edges, dyads or nodes with a given value in the series of networks generated from the ERGM parameters, and the bold, black points and lines indicate the observed values in ELSA-RioSC.

VII. Discussion and Conclusions

This study aimed to construct a sociocentric social network graph from responses to the "Social Network" section of the ELSA Wave 2 Questionnaire and to use that graph to better understand the relationship between an individual's at-work social environment and his or her BMI. The former was achieved through a probabilistic linkage process, and the latter through a combination of generalized linear, latent Gaussian, and exponential random graph modeling.

The prevalences of obesity and overweight in the ELSA-RioSC network were well above those in the city of Rio de Janeiro as a whole: in Rio in 2014, 54.4% of adults were above normal weight and 19.4% were obese,[23,24] while during approximately the same time period in ELSA-RioSC, 69.4% of participants were above normal weight and 28.2% were obese. These results alone justify the use of anti-obesity strategies targeting institutional environments in the city. An understanding of the social dynamics of weight gain and weight loss at work will help to inform those strategies.

The results from the modeling processes, which were of principal interest, are considered first, followed by a discussion of the representativeness and accuracy of those results given the methods utilized in data collection and analysis. Possible improvements to those methods and directions for future research are described next. Finally, the study concludes in considering the implications that the results presented here may carry for work-place anti-obesity interventions at Fiocruz and beyond.

i. Discussion of the results

The final latent Gaussian models suggest a sex-dependent relationship between an individual's BMI and the BMIs of those in his or her social network. In general, males and females with no friends or few friends showed a tendency for both greater BMIs at Wave 2 and greater average annual percent increases in BMI from Wave 1 to Wave 2 than did those individuals with some or many friends in the network. These results might be explained in one of three ways: individuals practicing behaviors that made them more likely to gain weight were more likely to be socially isolated at work, socially isolated individuals were more likely to develop behaviors associated with weight gain, or individuals already perceived as overweight were both more likely to continue gaining weight and less likely to be popular. To distinguish between these possible explanations, longitudinal friendship data would be required. If an individual was found to lose friends as he or she gained weight, the third explanation would seem most reasonable, and if individuals who lost friends over time tended to subsequently gain weight, evidence would support the second explanation. If individuals with few or no friends were found to both remain unpopular in the network and to gain weight, the first explanation would be favored. While studies of adolescent social networks have shown obese adolescents[84,85] and adolescents who practice some obesity-related behaviors to be more socially isolated,[68] similar studies are needed for adult social networks and more specifically for employee social networks. Studies have also shown that individuals with less social support are more likely to re-gain weight after a weight-loss intervention,[64,94] but again, more research is needed regarding the role of social isolation (both at work and at home) in normally changing weight statuses.

Both the sectional Wave 2 BMI and average annual percent change in BMI for the men in the network showed a positive correlation with the BMIs of their friends (not separated by sex), but the relationship between a woman's BMI and the BMIs of her friends was less clear. The data for the men agrees with data from other studies of adult social networks.[65,82] The results do not, however allow for conclusions to be drawn regarding the directionality of these associations: friendships could form around similar BMIs or similar eating and exercise habits, friends could respond to obesogenic environments similarly, or friends could affect one another's attitudes about food and exercise. The observed relationships between men's BMIs and the BMIs of their friends in the ELSA-RioSC network could also have resulted from some combination of these factors. Further analysis of homophily and change in eating and exercise habits could help to distinguish between these competing theories. For instance, one ego-centric at-work social network study concluded that individuals who perceived healthier eating habits in their coworkers were more likely to consume more fruits and vegetables and less fat, and individuals who perceived greater levels of coworker physical activity were likely to exercise more.[102]

It is possible that the BMIs of female ELSA-RioSC participants would be associated with the BMIs of their female friends. A study of the social clustering of obesity noted associations were stronger in same-sex friendships,[65] while another did not find that such limitations had a statistically significant effect.[82] A sex-restricted analysis represents an important future extension on the present study. It is also possible that at-work friendships play a bigger part in men's overall social lives than in women's overall social lives. If this were the case, female civil servants' BMIs might still reflect the BMIs of those in their social networks, but the influential parts of their social networks might exist predominantly outside of the workplace. An egocentric analysis including individuals' at-work and at-home relationships could help to test this theory. For instance, a future questionnaire could first ask participants to identify the number of individuals they consider to be close friends and then to indicate how many of those close friends are coworkers.

The strong negative association between the Wave 2 sectional BMIs of women, but not men, and their levels of education and income – more educated, higher-earning women showed, on average, lower BMIs – agrees with previous research on adults in Brazil. A study of data from 1999 of adult employees at a university in Rio de Janeiro also showed an inverse association between schooling and BMI among women, but not men. That study did not, however, find an association between per capita monthly family income and BMI in either sex.[81] A study of nationally representative data from 2006 found an inverse association between BMI and education level among women and a direct association among men.[86] Another study of national, longitudinal data showed that while in recent decades rates of obesity have tended to increase across all socioeconomic groups for Brazilian men, rates have decreased in women of high socioeconomic status and increased in women of low socioeconomic status.[28] Those trends support the inverse association observed here between socioeconomic status and BMI in women but not men.

Few studies were found against which to compare the results from the exponential random graph model, and none were found specific to Brazil. One ERGM study found that avoidance of overweight friends was a determining factor in adolescent friendship formation: non-overweight adolescents were less likely to nominate overweight adolescents as friends, but the reverse was not true.[84] Another ERGM study found that same-sex friends shared some obesity-related eating and exercise behaviors.[68] The ERGM results for ELSA-RioSC did not address either of those questions, instead focusing on whether or not, after controlling for potentially confounding exogenous and endogenous variables, similarity in BMI increases the probability of friendship formation. The results showed that, both before and after the inclusion of exogenous and endogenous controls, individuals with similar BMIs (defined by a difference of $4kg/m^2$ or less) were not significantly more likely to be friends. On the other hand, the results indicated that all other factors included in the model – namely, shared sex, shared work department, shared level of education, a small difference in age, the number of isolates in the network, mutuality, and transitivity – significantly improved the models ability to predict friendships in ELSA-RioSC. It is possible that conditioning endogenous effects on BMI would reveal a significant role for BMI in friendship formation, or that similarity in BMI is a predictor of same-sex friendship formation but not opposite-sex friendship formation. Both of these possibilities represent potentially important future directions for ELSA-RioSC ERGM-based research.

Together, the exponential random graph and latent Gaussian model results indicate that while two individuals with similar BMIs are not necessarily more likely to become friends than two individuals with dissimilar BMIs, among men (but not women), individuals whose friends' average BMIs are greater tend to have slightly greater BMIs than individuals whose friends average BMIs are in the normal to low range. Men's annual percent changes in BMI are also associated with the average annual percent changes in the BMIs of their friends. Women's BMIs in ELSA-RioSC are more closely associated with socioeconomic control factors than with at-work social factors.

ii. Methods, limitations and representativeness

Missing data was extremely limited, as noted in the Results section, and is therefore unlikely to have affected the analysis, but the lack of Wave 1 friendship information may have had important implications. In modeling the changes in BMI between Wave 1 and Wave 2, friendships were assumed to have remained constant over the 2-6 years between data collections. This was a more reasonable assumption than those made in other studies spanning greater stretches of time,[66] but given the relative fluidity of a work environment, it is likely that some of those friendships did change. If friendship data is collected again in future ELSA waves, the extent to which friendships change over time at Fiocruz can be assessed, and the data presented here can be retroactively adjusted to account for a certain proportion of friendships incorrectly assumed to have remained unchanged.

The methods in which friendship data was collected and assessed may also have affected the validity of the analyses and results. The question format allowed for the construction of a sociocentric ELSA-Rio network, which is more informative than egocentric or dyadic data. The fact that ELSA participants could name at-work friends regardless of those friends' own participation in ELSA also insured that ELSA-RioSC was a relatively representative subgraph of the greater unobserved Fiocruz friendship network: participants weren't asked to confine their friendship nominations to ELSA participants, so the ELSA participants they did name as friends were in fact within their five closest at-work friends. On the other hand, the fixed choice question format – restricting friendship nominations to a maximum of five – may have led to individuals limiting their definitions of important friendships (if they had more than five close friends at work, as 45.6% of individuals claimed to have). Additionally, approximately 3% of individuals cited fewer than five friends by name despite claiming to have more than five close friends at work, indicating either an exaggeration in their original estimation of close friends, an inability to recollect enough information by which to cite those friends, or an unwillingness to identify friends by name.

The free recall approach to friendship nominations was considered more viable than a roster-based approach given the large number of Fiocruz employees,[96] but it led to a significant amount of lost information from individuals mis-remembering or inadequately citing names; for 6% of cited friends, only a first name was given. In some cases, this meant that the friend to whom an individual was referring could not, with sufficient confidence, be identified through linkage. In other cases, incorrect linkages may have been drawn from false interpretation of the limited available data or from inaccurate participant-reported data. Together, missing and incorrect friendships, if relatively extensive, may have hidden true associations in the network or created false associations. Even with a “roster” approach to friend collection, however, some degree of false identification would have been expected. If free recall is used again in the future, questionnaire administrators should encourage participants to share as much accurate information as possible about their friends, allowing individuals to look up their friends' names on social media sites or to call them for proper identification information.

In regards to the linkage process, in future analyses, linkage should be focused on the ELSA database. Given the large number of outsourced workers not listed in official registries, linkage beyond ELSA is difficult; it is also unnecessary. Friendships between ELSA participants are the only ones relevant to a sociocentric network, and this subgraph is more informative than the egocentric network that would result from inclusion of relationships between ELSA participants and non-ELSA employees, for whom no health-related information was available. Given current software limitations that make linkage based solely on names difficult, with a well-defined process and criteria, limiting linkage to ELSA participants would allow for more a more focused and intensive use of resources. As mentioned above, however, for the sake of accuracy and representativeness, friendship nomination should not be limited to ELSA, even if the linkage process were to be.

The BMI distribution of ELSA-Rio Wave 2 participants was not representative of national or city-wide BMI distributions; prevalences of overweight and obesity in both sexes in ELSA-RioSC were significantly higher than those in Rio or in Brazil.[23,24] As expected, the observed rates were also well above those reported in a 1999 study of adult employees at a university in Rio. In that study, as in this study, overweight was more prevalent in men than in women, but rates of obesity were similar between both sexes.[81] Data was unavailable regarding the prevalences of overweight and obesity in the entire Fiocruz employee population, so no conclusions could be drawn regarding how well the study population represents the greater institutional population. ELSA-Brasil was relatively successful in meeting its recruitment goals regarding gender, age, and occupation type, but it is unclear how well those goals reflect the population of Fiocruz employees. Participation was also largely volunteer-based (76%), with the remaining 24% recruitment-based,[97] and it is possible that healthier or more socially connected individuals were more likely to learn about – or agree to participate in - the study. If this was the case, the associations observed in ELSA-RioSC may not apply to the overall Fiocruz employee social network.

iii. Implications for the obesity epidemic and its control

Assuming that the associations observed in the ELSA-RioSC network do, in fact, represent real dynamics in the larger Fiocruz employee social network, these results have the potential to inform future work-based efforts to curb the obesity epidemic. The association between mens' BMIs and the BMIs of their friends implies that weight-loss efforts aimed at a single person may lead to change in multiple people, and recruiting friends to at-work weight-loss programs together may be especially helpful for men. But some overweight men and women may be difficult to reach through socially-based weight loss programs, as the data showed evidence that men and women with fewer at-work friends tend to have slightly larger BMIs and to gain more weight. The BMIs of women who cited no friends in ELSA-RioSC, for instance, increased by 0.231 percentage points more per year than those of women who cited, on average, low or normal weight friends in the network. The BMIs of isolated men in ELSA-RioSC also increased by 0.140 percentage points more per year than those of men with friends whose weights on average did not change significantly. If social isolation leads to weight gain (rather than weight-gain leading to social isolation), these results indicate that efforts to provide participants in work-based weight-loss programs with social support may help to make those programs more effective for both sexes.

The inverse associations between women's BMIs and their income levels and education levels are likely reflective of realities outside of work, like the relatively low cost of ultra-processed, obesogenic foods, the limited availability and visibility of healthy food options in lower-income communities, and the insufficient propagation of information on nutritious eating. Still, much of an employees' day is spent at work, and work shifts often include at least one meal. The workplace could therefore represent an important platform through which to affect change that extends to homes and communities. Food options at on-site eating establishments should be healthy and paired with information on balanced, nutritious diets, and employers should continue to provide employees with ample time for sit-down meals with friends to preserve traditional food systems in Brazil. Workplaces could also provide employees with opportunities to purchase unprocessed groceries on site, thereby helping to address the inequality in access to healthy food options. If it is the case, as proposed above, that women are still affected in their eating and exercise habits by those in their social networks, but their meaningful social networks exist primarily outside of work, then it is possible that workplace based initiatives to improve access to lower-cost, nutritious foods and information regarding the benefits of balanced diets would spread to communities through the social networks of female employees.

The results presented here justify further research regarding the associations between an individual's social networks and his or her body weight. In particular, longitudinal friendship data and detailed information on the eating and exercise habits of employees would be useful in understanding some of the mechanisms that led to the observed associations in ELSA-RioSC. Fundamentally, the epidemic of overweight and obesity in Brazil is a product of individuals living in unhealthy environments and adopting unhealthy behaviors. As Ribeiro de Castro notes: “The only actions that have a chance of being effective are those that integrate measures aimed at people...with measures aimed at the environments in which they live.”[48] What social network analyses recognize is that the environment in which a person lives includes his or her friends and family: actions aimed at one person can impact multiple people, and obesogenic environments can include intrapersonal environments. In Brazil, where eating has traditionally been – and largely continues to be – a social act, an understanding of the role of social networks in weight-gain, and perhaps more importantly in weight loss, is critical in planning an effective strategy to combat the so far uncontrolled epidemic of obesity.

VIII. References

1. WHO | Obesity and overweight. *WHO* at <http://www.who.int/mediacentre/factsheets/fs311/en/>
2. Organisation mondiale de la santé. *Global status report on noncommunicable diseases 2014: attaining the nine global noncommunicable diseases targets; a shared responsibility.* (World Health Organization, 2014).
3. Ezzati, M. & Riboli, E. Behavioral and Dietary Risk Factors for Noncommunicable Diseases. *N. Engl. J. Med.* **369,** 954–964 (2013).
4. WHO | NCD mortality and morbidity. *WHO* at <http://www.who.int/gho/ncd/mortality_morbidity/en/>
5. WHO | Noncommunicable diseases. at <http://www.who.int/mediacentre/factsheets/fs355/en/>
6. Maria Inês Schmidt *et al.* Chronic non-communicable diseases in Brazil: burden and current challenges. *The Lancet* (2011). at <http://www.sbh.org.br/pdf/lancet_collection.pdf>
7. Ng, M. *et al.* Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet* **384,** 766–781 (2014).
8. B. Swinburn *et al.* The global obesity pandemic: shaped by global drivers and local environments. *The Lancet* **378,** 804–818 (2011).
9. Williams, E. P., Mesidor, M., Winters, K., Dubbert, P. M. & Wyatt, S. B. Overweight and Obesity: Prevalence, Consequences, and Causes of a Growing Public Health Problem. *Curr. Obes. Rep.* **4,** 363–370 (2015).
10. Barry M. Popkin. The Nutrition Transition and Obesity in the Developing World. *J. Nutr.* **131,** 871S–873S (2001).
11. Chou, S.-Y., Grossman, M. & Saffer, H. An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System. *J. Health Econ.* **23,** 565–587 (2004).
12. James O. Hill & John C. Peters. Environmental Contributions to the Obesity Epidemic. *Science* **280,** 1371–1374 (1998).

13. Morland, K. B. & Evenson, K. R. Obesity prevalence and the local food environment. *Health Place* **15,** 491–495 (2009).

14. Spence, J. C., Cutumisu, N., Edwards, J., Raine, K. D. & Smoyer-Tomic, K. Relation between local food environments and obesity among adults. *BMC Public Health* **9,** 192 (2009).

15. Papas, M. A. *et al.* The Built Environment and Obesity. *Epidemiol. Rev.* **29,** 129–143 (2007).

16. Feng, J., Glass, T. A., Curriero, F. C., Stewart, W. F. & Schwartz, B. S. The built environment and obesity: a systematic review of the epidemiologic evidence. *Health Place* **16,** 175–190 (2010).

17. Weinsier, R. L., Hunter, G. R., Heini, A. F., Goran, M. I. & Sell, S. M. The etiology of obesity: relative contribution of metabolic factors, diet, and physical activity. *Am. J. Med.* **105,** 145–150 (1998).

18. Clement, K., Boutin, P. & Froguel, D. P. Genetics of Obesity. *Am. J. Pharmacogenomics* **2,** 177–187 (2002).

19. Rose, G. Sick individuals and sick populations. *Int. J. Epidemiol.* **30,** 427–432 (2001).

20. Egger, G., Swinburn, B. & Amirul Islam, F. M. Economic growth and obesity: An interesting relationship with world-wide implications. *Econ. Hum. Biol.* **10,** 147–153 (2012).

21. GDP per capita (current US$). *The World Bank* (2014). at <http://data.worldbank.org/indicator/NY.GDP.PCAP.CD>

22. Global Health Observatory Data Repository. *WHO* at <http://apps.who.int/gho/data/view.main.NCD1760?lang=en>

23. Núcleo de Comunicação/GAB/SVS. *VIGITEL Brasil 2013: protective and risk factors for chronic diseases by telephone sruvey*. 32–33 (Ministério da Saúde, 2014). at <http://biavati.files.wordpress.com/2014/05/vigitel-2013.pdf>

24. *VIGITEL Brasil 2006: protective and risk factors for chronic diseases by telephone survey*. 39 (Ministério da Saúde, 2007). at <http://bvsms.saude.gov.br/bvs/publicacoes/relatorio_vigitel_2006_marco_2007.pdf>

25. Epidemia triplica o número de mortes por obesidade em dez anos no país. As informações são do DATASUS - DATASUS. at <http://datasus.saude.gov.br/noticias/atualizacoes/447-epidemia-

triplica-o-numero-de-mortes-por-obesidade-em-dez-anos-no-pais-as-informacoes-sao-do-datasus>

26. Jones-Smith, J. C., Gordon-Larsen, P., Siddiqi, A. & Popkin, B. M. Is the burden of overweight shifting to the poor across the globe? Time trends among women in 39 low- and middle-income countries (1991–2008). *Int. J. Obes. 2005* **36,** 1114–1120 (2012).

27. Monteiro, C. A., Conde, W. L. & Popkin, B. M. The Burden of Disease From Undernutrition and Overnutrition in Countries Undergoing Rapid Nutrition Transition: A View From Brazil. *Am. J. Public Health* **94,** 433–434 (2004).

28. Carlos Augusto Monteiro, Wolney L. Conde & Barry M. Popkin. Income-specific trends in obesity in Brazil: 1975-2003. *Am. J. Public Health* **97,** 1808–1812 (2007).

29. Popkin, B. M., Adair, L. S. & Ng, S. W. Global nutrition transition and the pandemic of obesity in developing countries. *Nutr. Rev.* **70,** 3–21 (2012).

30. Carlos A. Monteiro & Geoffrey Cannon. The Impact of Transnational 'Big Food' Companies on the South: A view from Brazil. *PLoS Med* **9,**

31. Carlos Augusto Monteiro, Renata Bertazzi Levy, Rafael Moreira Claro, Inês Rugani Ribeiro de Castro & Geoffrey Cannon. Increasing consumption of ultra-processed foods and likely impact on human health: evidence from Brazil. *Public Health Nutr.* **14,** 5–13 (2010).

32. Pereira, R. A., Duffey, K. J., Sichieri, R. & Popkin, B. M. Sources of excessive saturated fat, trans fat and sugar consumption in Brazil: an analysis of the first Brazilian nationwide individual dietary survey. *Public Health Nutr.* **17,** 113–121 (2014).

33. Levy-Costa, R. B., Sichieri, R., Pontes, N. dos S. & Monteiro, C. A. Household food availability in Brazil: distribution and trends (1974-2003). *Rev. Saúde Pública* **39,** 530–540 (2005).

34. Monteiro, C. A., Moubarac, J.-C., Cannon, G., Ng, S. W. & Popkin, B. Ultra-processed products are becoming dominant in the global food system. *Obes. Rev.* **14,** 21–28 (2013).

35. Corinna Hawkes. Uneven dietary development: linking the policies and processes of globalization with the nutrition transition, obesity and diet-related chronic diseases. *Glob. Health 2006* **2,** (2006).

36. Sarah E. Clark, Corinna Hawkes, Sophia M. E. Murphy, Karen A. Hansen-Kuhn & David

Wallinga. Exporting obesity: US farm and trade policy and the transformation of the Mexican consumer food environment. *Int. J. Occup. Environ. Health* **18,** 53–64 (2012).

37. Anne Marie Thow & Corinna Hawkes. The implications of trade liberalization for diet an heath: a case study from Central America. *Glob. Health* **5,** (2009).

38. Popkin, B. M. Urbanization, lifestyle changes and the nutrition transition. *World Dev.* **27,** 1905–1916 (1999).

39. Ng, S. W. & Popkin, B. M. Time use and physical activity: a shift away from movement across the globe. *Obes. Rev.* **13,** 659–680 (2012).

40. Brito, F. The displacement of the Brazilian population to the metropolitan areas. *Estud. Av.* **20,** 221–236 (2006).

41. Brasil. Constituição da República Federativa do Brasil de 1988. at <http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm>

42. Ministério do desenvolvimento social e combate à fome. SEGURANÇA ALIMENTAR E NUTRICIONAL: ACESSO À ALIMENTAÇÃO. at <http://mdspravoce.mds.gov.br/seguranca-alimentar-e-nutricional/acesso-a-alimentacao/>

43. Rosângela Minardi Mitre Cotta & Juliana Costa Machado. Programa Bolsa Família e segurança alimentar e nutricional no Brasil: revisão crítica da literatura. *Rev. Panam. Salud Pública* **33,** 54–60 (2013).

44. Conde, W. L. & Monteiro, C. A. Nutrition transition and double burden of undernutrition and excess of weight in Brazil. *Am. J. Clin. Nutr.* **100,** 1617S–1622S (2014).

45. De Bem Lignani, J., Sichieri, R., Burlandy, L. & Salles-Costa, R. Changes in food consumption among the Programa Bolsa Família participant families in Brazil. *Public Health Nutr.* **14,** 785–792 (2011).

46. Política Nacional de Alimentação e Nutrição. (2011). at <http://dab.saude.gov.br/portaldab/diretrizes_pnan.php>

47. Ministério da Saúde. Guia Alimentar Para a População Brasileira. (2014).

48. De Castro, I. R. R. Desafios e perspectivas para a promoção da alimentação adequada e

saudável no Brasil Challenges and perspectives for the promotion of adequate and healthy food in Brazil Desafíos y perspectivas para la promoción de una. *Cad Saúde Pública* **31,** 7–9 (2015).

49. Ministério da Saúde. *Política Nacional de Promoção da Saúde*. (2006). at <http://bvsms.saude.gov.br/bvs/publicacoes/politica_nacional_promocao_saude_3ed.pdf>

50. Câmara Interministerial de Segurança Alimentar e Nutricional. *Estratégia Intersetorial de Prevenção e Controle da Obesidade: Recomendações para estados e municípios*. (CAISAN, 2014). at <http://www.cfn.org.br/eficiente/repositorio/documentos/814.pdf>

51. Programa Peso Saudável. at <http://dabsistemas.saude.gov.br/sistemas/pesosaudavel/>

52. Global Strategy on Diet, Physical Activity and Health. (2004). at <http://www.who.int/dietphysicalactivity/strategy/eb11344/strategy_english_web.pdf?ua=1>

53. Royal College of Physicians. *Action on Obesity: Comprehensive care for all*. (RCP, 2013).

54. Gortmaker, S. L. *et al.* Changing the future of obesity: science, policy, and action. *The Lancet* **378,** 838–847 (2011).

55. Powell, K. *et al.* The role of social networks in the development of overweight and obesity among adults: a scoping review. *BMC Public Health* **15,** (2015).

56. Cunningham, S. A., Vaquera, E., Maturo, C. C. & Narayan, K. M. V. Is There Evidence that Friends Influence Body Weight? A Systematic Review of Empirical Research. *Soc. Sci. Med. 1982* **75,** 1175–1183 (2012).

57. Fletcher, A., Bonell, C. & Sorhaindo, A. You are what your friends eat: systematic review of social network analyses of young people's eating behaviours and bodyweight. *J. Epidemiol. Community Health* **65,** 548–555 (2011).

58. Nam, S., Redeker, N. & Whittemore, R. Social networks and future direction for obesity research: A scoping review. *Nurs. Outlook* **63,** 299–317 (2015).

59. Smith, K. P. & Christakis, N. A. Social networks and health. *Annu. Rev. Sociol.* **34,** 405–429 (2008).

60. Marsden, P. Memetics and social contagion: Two sides of the same coin. *J. Memet.-Evol. Models Inf. Transm.* **2,** 171–185 (1998).

61. Crandall, C. S. Social contagion of binge eating. *J. Pers. Soc. Psychol.* **55,** 588 (1988).

62. D, W. & Re, B. The relationship of loneliness, social isolation, and physical health to dietary adequacy of independently living elderly. *J. Am. Diet. Assoc.* **91,** 300–304 (1991).

63. Epstein, L. H., Valoski, A., Wing, R. R. & McCurley, J. Ten-year outcomes of behavioral family-based treatment for childhood obesity. *Health Psychol.* **13,** 373 (1994).

64. Wing, R. R. & Jeffery, R. W. Benefits of recruiting participants with friends and increasing social support for weight loss and maintenance. *J. Consult. Clin. Psychol.* **67,** 132–138 (1999).

65. Christakis, N. A. & Fowler, J. H. The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* **357,** 370–379 (2007).

66. Cohen-Cole, E. & Fletcher, J. M. Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. *J. Health Econ.* **27,** 1382–1387 (2008).

67. Lyons, R. The spread of evidence-poor medicine via flawed social-network analysis. *Stat. Polit. Policy* **2,** (2011).

68. De la Haye, K., Robins, G., Mohr, P. & Wilson, C. Obesity-related behaviors in adolescent friendship networks. *Soc. Netw.* **32,** 161–167 (2010).

69. Hill, A. L., Rand, D. G., Nowak, M. A. & Christakis, N. A. Infectious disease modeling of social contagion in networks. *PLoS Comput. Biol.* **6,** e1000968 (2010).

70. Giabbanelli, P. J., Alimadad, A., Dabbaghian, V. & Finegood, D. T. Modeling the influence of social networks and environment on energy balance and obesity. *J. Comput. Sci.* **3,** 17–27 (2012).

71. Bahr, D. B., Browning, R. C., Wyatt, H. R. & Hill, J. O. Exploiting Social Networks to Mitigate the Obesity Epidemic. *Obesity* **17,** 723–728 (2009).

72. El-Sayed, A. M., Seemann, L., Scarborough, P. & Galea, S. Are network-based interventions a useful antiobesity strategy? An application of simulation models for causal inference in epidemiology. *Am. J. Epidemiol.* **178,** 287–295 (2013).

73. Konchak, C. & Prasad, K. Incorporating social network effects into cost-effectiveness analysis: a methodological contribution with application to obesity prevention. *BMJ Open* **2,** (2012).

74. Leroux, J. S., Moore, S. & Dube, L. Beyond the ‘I’ in the obesity epidemic: a review of social

relational and network interventions on obesity. *J. Obes.* (2013). at <http://go.galegroup.com/ps/i.do?id=GALE %7CA368282559&v=2.1&u=mlin_m_tufts&it=r&p=&sw=w&asid=4443dce1772ac358f6718f34f6053091>

75. Stanley Wasserman & Katherine Faust. *Social Network Analysis*. (Cambridge University Press, 1994).

76. Eric D. Kolaczyk. *Statistical Analysis of Network Data*. (Springer, 2009).

77. Kolaczyk, E. D. & Csárdi, G. *Statistical analysis of network data with R*. (Springer, 2014).

78. Robins, G., Pattison, P., Kalish, Y. & Lusher, D. An introduction to exponential random graph (p*) models for social networks. *Soc. Netw.* **29,** 173–191 (2007).

79. GOODREAU, S. M., KITTS, J. A. & MORRIS, M. Birds of a Feather, or Friend of a Friend? Using Exponential Random Graph Models to Investigate Adolescent Social Networks. *Demography* **46,** 103–125 (2009).

80. Rue, H., Martino, S. & Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71,** 319–392 (2009).

81. Fonseca, M. de J. M. da, Faerstein, E., Chor, D., Lopes, C. S. & Andreozzi, V. L. Associações entre escolaridade, renda e Índice de Massa Corporal em funcionários de uma universidade no Rio de Janeiro, Brasil: estudo Pró-Saúde. *Cad Saúde Pública* **22,** 2359–67 (2006).

82. Hruschka, D. J., Brewis, A. A., Wutich, A. & Morin, B. Shared Norms and Their Explanation for the Social Clustering of Obesity. *Am. J. Public Health* **101,** S295–S300 (2011).

83. Strauss, R. S. & Pollack, H. A. Social marginalization of overweight children. *Arch. Pediatr. Adolesc. Med.* **157,** 746–752 (2003).

84. Schaefer, D. R. & Simpkins, S. D. Using Social Network Analysis to Clarify the Role of Obesity in Selection of Adolescent Friends. *Am. J. Public Health* **104,** 1223–1229 (2014).

85. Valente, T. W., Fujimoto, K., Chou, C.-P. & Spruijt-Metz, D. Adolescent Affiliations and Adiposity: A Social Network Analysis of Friendships and Obesity. *J. Adolesc. Health* **45,** 202–204

(2009).

86. Gigante, D. P., Moura, E. C. de, Sardinha, L. M. V. & others. Prevalência de excesso de peso e obesidade e fatores associados, Brasil, 2006. *Rev Saúde Pública* **43,** 83–9 (2009).

87. Lino, M. Z. R., Muniz, P. T. & Siqueira, K. S. Prevalência e fatores associados ao excesso de peso em adultos: inquérito populacional em Rio Branco, Acre, Brasil, 2007-2008 Prevalence of overweight and associated factors in adults: a population survey in Rio Branco. *Cad Saúde Pública* **27,** 797–810 (2011).

88. Oliveira, L. P. *et al.* Fatores associados a excesso de peso e concentração de gordura abdominal em adultos na cidade de Salvador, Bahia, Brasil Factors associated with overweight and abdominal fat in adults in Salvador, Bahia State, Brazil. *Cad Saúde Pública* **25,** 570–582 (2009).

89. De la Haye, K., Robins, G., Mohr, P. & Wilson, C. Homophily and Contagion as Explanations for Weight Similarities Among Adolescent Friends. *J. Adolesc. Health* **49,** 421–427 (2011).

90. Fowler, J. H. & Christakis, N. A. Estimating peer effects on health in social networks: A response to Cohen-Cole and Fletcher; and Trogdon, Nonnemaker, and Pais. *J. Health Econ.* **27,** 1400–1405 (2008).

91. Trogdon, J. G., Nonnemaker, J. & Pais, J. Peer effects in adolescent overweight. *J. Health Econ.* **27,** 1388–1399 (2008).

92. Halliday, T. J. & Kwak, S. Weight gain in adolescents and their peers. *Econ. Hum. Biol.* **7,** 181–190 (2009).

93. Aggarwal, B., Liao, M., Allegrante, J. P. & Mosca, L. Low Social Support Level is Associated with Non-Adherence to Diet at 1 Year in the Family Intervention Trial for Heart Health (FIT Heart). *J. Nutr. Educ. Behav.* **42,** 380–388 (2010).

94. Kayman, S., Bruvold, W. & Stern, J. S. Maintenance and relapse after weight loss in women: behavioral aspects. *Am. J. Clin. Nutr.* **52,** 800–807 (1990).

95. ELSA Brasil. Questionário ELSA Onda 2. (2012).

96. Cláudia Medina Coeli, Ana Luiza Alfaya Gallego Soares & Marilia Sá Carvalho. Redes Sociais.

97. Aquino, E. M. L. *et al.* Brazilian Longitudinal Study of Adult Health (ELSA-Brasil): Objectives

and Design. *Am. J. Epidemiol.* **175,** 315–324 (2012).

98. Camargo Jr., K. R. de & Coeli, C. M. Going open source: some lessons learned from the development of OpenRecLink. *Cad. Saúde Pública* **31,** 257–263 (2015).

99. Kenneth Rochel de Camargo Jr. & Cláudia Medina Coeli. OpenRecLink: Guia do usuário. (2012).

100. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal Complex Syst.* **1695,** 1–9 (2006).

101. Qual a faixa de renda familiar das classes? at <http://cps.fgv.br/node/3999>

102. Lemon, S. C. *et al.* Perceptions of Worksite Support and Employee Obesity, Activity and Diet. *Am. J. Health Behav.* **33,** 299–308 (2009).