Nat Hum Behav. Author manuscript; available in PMC 2021 October 22.

Published in final edited form as:
Nat Hum Behav. 2021 October ; 5(10): 1349–1357. doi:10.1038/s41562-021-01103-x.

Social connections and the healthfulness of food choices in an employee population

Douglas E. Levy[*,‡,1,2], **Mark C. Pachucki**[‡,3], **A. James O'Malley**[4], **Bianca Porneala**[5], **Awesta Yaqubi**[6], **Anne N. Thorndike**[2,5]

[1.]Mongan Institute Health Policy Research Center, Massachusetts General Hospital; Boston, Massachusetts, U.S.A.

[2.]Harvard Medical School; Boston, Massachusetts, U.S.A.

[3.]Department of Sociology & Computational Social Science Institute, University of Massachusetts, Amherst; Amherst, Massachusetts, U.S.A.

[4.]Department of Biomedical Data Science and The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth; Hanover, New Hampshire, U.S.A.

[5.]Division of General Internal Medicine, Massachusetts General Hospital; Boston, Massachusetts, U.S.A.

[6.](current affiliation) Boston University School of Medicine; Boston, Massachusetts, U.S.A.

Abstract

Unhealthy food choice is an important driver of obesity, but research examining the relationship of food choices and social influence has been limited. We sought to assess associations in the healthfulness of workplace food choices among a large population of diverse employees whose food-related social connections were identified using passively-collected data in a validated model. Data were drawn from 3 million encounters where pairs of employees made purchases together in 2015–2016. The healthfulness of food items was defined by "traffic light" labels. Cross-sectional simultaneously autoregressive models revealed that proportions of both healthy and unhealthy items purchased were positively associated between connected employees. Longitudinal generalized estimating equation models also found positive associations between an employee's

[*]**Corresponding Author:** Douglas Levy; Mongan Institute Health Policy Research Center, Massachusetts General Hospital; 100 Cambridge St., Suite 1600, Boston MA 02114, tel 617.643.3595, fax 617.724.4738, dlevy3@mgh.harvard.edu;.
Author contributions: DEL, MCP, AJO, and ANT designed research; DEL, MCP, and AY performed research; DEL, MCP, and BP analysed data; and DEL, MCP, AJO, and ANT wrote the paper.
[‡]These authors contributed equally

Data Availability
Though the data are deidentified, combined demographic data could potentially identify individuals or small groups. As a result, interested parties may access the data by applying to the corresponding author and entering into an appropriate data use agreement.

Code Availability
The R code used to estimate the SAR, GEE, and instrumental variables models is illustrated in Supplementary Figures 10, 11, and 12. Full code is available upon request to the corresponding author.

Competing interest statement: The cafeterias providing data for this project are owned by the Massachusetts General Hospital, which employs Levy, Porneala, and Thorndike. The remaining authors declare no competing interests.

current food purchase and the most recent prior food purchase a co-worker made together with the employee. These data suggest that workplace interventions to promote healthy eating and reduce obesity should test peer-based strategies.

Unhealthy eating is a major and preventable risk factor for most chronic diseases. Obesity and overweight, which have reached epidemic levels in recent years, are largely attributable to dietary behaviours[1,2] and are counted among the leading preventable causes of morbidity and mortality, both in the United States and worldwide[3–6]. Diets high in sugar, sodium, and saturated fats lead to diabetes, hypertension, cardiovascular disease, and cancer. The rapid rise in prevalence of obesity and obesity-related diseases is contributing to a slow-down in improvements to U.S. life expectancy[7]. While it is well-established that the quality and quantity of food choices are affected by social and physical environments[8–12], there is little understanding of how interpersonal dynamics affect both individual and population food choices over extended periods of time.

Recent research has suggested that obesity spreads through social networks, but it is still not well-understood how this might occur[13–15]. Humans influence one another's food choices in systematic ways[16–18] and it is likely that food choices contributing to obesity and chronic disease could be transferred among socially connected individuals when a person imitates behaviours modelled by others. This might occur through descriptive norms which illustrate possible or prevalent behaviours, or through injunctive norms, which explicitly or implicitly convey expected behaviour[18]. The choices that one makes about food are partially based on physical need, yet also reflect some degree of elective choice. What humans eat is at least partially situationally and culturally patterned, and these eating patterns both shape, and are shaped by, one's social networks[19,20].

Most existing studies of social influence and food choice have been conducted in small-group experimental settings, often among young adults and college students, making it difficult to generalize findings to other age groups and to real-world environments. In addition, most studies focus on social influence during a single meal occasion[21]. Outside of these narrow settings, there has been limited research on food choice and social influence in observational studies from population samples of friends and family[22,23], but there is little if any research evaluating how an individual's food choices may be affected by the cumulative social influence of one or more non-intimate peers, such as in a workplace, over time.

Workplaces are artificially bounded communities (networks) in which employees may spend up to half their waking hours[24]. Workplace relationships are often distinct from other forms of personal affiliation, as social ties are primarily derived from common work tasks and goals. Employees often cannot choose their co-workers, and workplace relationships may be less intimate with different reciprocal obligations. There are also varying levels of hierarchy and power differentials[25]. Co-workers usually do not share genetics or home environments and may come from different cultural backgrounds. These factors, related to food choice, distinguish workplace relationships from more intimate relationships with family and friends. While studies of workplace relationships have focused on their role in productivity and friendships in these settings[26], this research has not investigated the role of workplace relationships in shaping health behaviours, including food choices, all of which can be quite personal.

The goal of the present study was to investigate whether there was evidence supporting social influence in the healthfulness of food choices among a diverse group of employees in a real-world workplace setting – 7 on-site cafeterias of a large hospital with approximately 26,000 employees – over a two-year period. Using a relational data analysis approach grounded in computational social science[27,28], we examined complicated pathways of influence by including data on employees ("egos" in the terminology of social network analysis), their co-workers ("alters"), and information on the pairings themselves (dyads). We used electronic food purchase data obtained from the hospital's cafeterias to examine 71,611,372 paired purchases, ultimately identifying 3,771,714 unique purchasing encounters during a two-year period between pairs of employees who were likely to be socially connected. Social connections were inferred based on a model examining patterns in the time and location of food purchases, as well as employee characteristics, that was validated using participant observation and a survey of employee eating behaviours. The healthfulness of purchases was determined using the hospital cafeteria's "traffic light" labelling system designating all food and beverages as green (healthy, "eat often"), yellow (less healthy, "eat less often"), or red (unhealthy, "there is a better choice in green or yellow"). The traffic light labelling system, informed largely by foods' calories, saturated fat, and healthy ingredients (fruit and vegetable, whole grain, and lean protein), was based on USDA dietary guidelines and has been in place in these cafeterias since 2010[12]. Prior research established that traffic light labels in the food environment were associated with population-level improvements in the healthfulness of food choices and that employees improved the healthfulness of their food choices when they were shown a comparison of the healthfulness of their own purchases relative to other employees' purchases[12,29,30]. In the present study, we conducted cross-sectional and longitudinal assessments of the degree to which the healthfulness of employees' purchases was associated with and potentially influenced by purchases among co-workers with (inferred) social connections to the employees. We hypothesized that the healthfulness of items purchased, based on the simple traffic light labelling system, would be correlated among socially connected employees.

We found that the healthfulness of individual employees' purchases was positively associated with the healthfulness of their co-workers' purchases in both the cross-sectional and longitudinal regression analyses. This was true for healthy and unhealthy food and beverage purchases, though point estimates for the associations were largest and most consistent for healthy foods. The robustness of these findings was bolstered by a range of sensitivity analyses. We acknowledge that our analyses cannot generally rule out the influence of latent homophily (e.g., co-workers with similar lifestyle and food preferences may be more likely to become friends and to eat with each other) that likely contributed at least partially to these associations or even detracted from the true associations (in the case of latent heterophily). However, several factors support the premise that social influence may have been an important contributor to the observed associations, including the strength of our research design, evidence from sensitivity analyses (including instrumental variables analyses), and consistency with prior research assessing social influence on food choices among narrower samples and in lab-based settings.

Our findings are important for several reasons. First, the results establish a specific, plausible behavioural mechanism (un/healthy food choice) through which social influence may affect obesity risk. Second, the study assessed interpersonal associations in food choice in a real-world setting among individuals across a range of ages and diverse socioeconomic positions, from unskilled service workers through highly educated physicians and scientists. Third, our focus on workplace social ties was notable given that these ties largely fall outside the boundaries of what scientists consider to be "social intimates", such as family members and close friends. Fourth, behaviour was assessed using objectively-collected continuous time secondary administrative data, thus avoiding the risk of a Hawthorne effect and the social desirability bias that may arise in experimental settings. And fifth, by examining purchases over a two-year period, we found that associations in food choice were not limited to a single interaction but existed across the span of multiple interactions.

To the extent that the associations we measured reflect social influence, this research has implications for the design of public health and policy interventions to prevent obesity that may be realized using ubiquitous and readily-available data linking individuals to one another. Social influence could be leveraged to target particular individuals or social connections to promote healthy eating or disrupt unhealthy eating. Furthermore, the physical environments where social interactions occur could be designed to target pairs of people (or larger groups of purchasers) making food choices. For example, healthfulness could be encouraged by offering two-for-one sales on salads (or other healthy foods) for pairs of purchasers. Lastly, quantifying the social transmission of food choice behaviour could allow policy-makers and researchers to develop interventions and policies that efficiently target specific groups of people for increased population effect[31,32]. Eating together is an example of a public health scenario that likely involves a large "network multiplier" effect[33]. If food choice is "socially contagious" and an intervention improves healthy eating in a particular group, the benefit of that intervention will accrue not only to that group, but to individuals socially connected to group members, as well. This more complete capture of benefits will provide greater incentives for stakeholders to adopt health-promoting interventions. Though our findings do not prove the existence of social influence in food choice, they are consistent with such an explanation and together with these implications highlight the value of additional work to test peer-based strategies for promoting healthy eating.

Results

Estimating tie probabilities in transaction data

For our analyses, we considered a social connection present when we could infer that two employees knew one another, met at a cafeteria, and made a purchase at the same time. In the cafeteria transaction data, we were only able to observe purchasing behaviour recorded by computerized cash register records, including the purchase time and location. Following similar strategies utilized in other settings, we hypothesized that if we observed two employees making purchases at the same cafeteria within +/− 2 minutes of one another, they might know one another[34–38]. Further, if two employees made purchases within +/− 2 minutes on multiple occasions and/or at multiple locations, that would be stronger evidence of a real social connection. To test this hypothesis, we surveyed employees who used the hospital cafeterias regularly and presented each with an individualized list containing the names of co-workers who we thought they might have eaten with based on purchasing times and locations, as well as the names of co-workers who did not meet those criteria. We asked the employees to confirm whether or not they ate with the co-workers listed. Of 1,946 employees surveyed, 1,054 (54%) responded. We used respondent-confirmed social connections as a "gold standard" and generated a predictive model of "true social ties" using cash register data on the time and location of purchases, as well as linked information on the demographic and work characteristics of the employees from human resources data as predictors. Using split-sample validation, we identified a model with an area under the receiver operating characteristic curve of 0.85 when the model was applied to the validation data set (the portion of the sample not used to train the model). The strongest predictors of "true" matches were being from the same department, having made ≥5 purchases together within an 8-week period, and having met at 3 or more different cafeterias over 8 weeks. Additional important predictors were having the same job type, same gender, and similar age (difference <10 years) (see Supplementary Table 1). The model was highly specific even at low probability thresholds (Figure 1, panel A), and had a positive predictive value above 75% for thresholds above 0.6 (Figure 1, panel B). As expected, when applied to our 2015–2016 purchasing data, most employee-co-worker dyads had low probability of being "true" (Figure 1, panel C).

Cross-sectional associations in healthful purchasing

Figure 2 presents results from our cross-sectional analyses examining the associations between the percent of all items an employee purchased during 2015 that were labeled green (or red) and all their co-workers' green (or red) purchases over that same time period. Associations were estimated using simultaneously autoregressive (SAR) models which were adapted from spatial regression techniques. These models allow an employee's purchases to appear as both an outcome (where the person is the "ego") and as an independent variable (where the person is an "alter" to another employee), reflecting the potential bidirectional nature of social influence. SAR models also allow the inclusion of data on purchases employees and co-workers made at the same time and place[39,40]. Ties between employees and co-workers were incorporated into the models through an "adjacency matrix" that not only accounted for the presence and frequency of a social tie but also incorporated tie probabilities estimated from our predictive model, thus more heavily weighting alters most likely to be involved in true social interactions with egos in the computation of average alter food purchasing variables. Models adjusted for employee characteristics. We estimated models for 2015 and 2016 separately to assess our hypothesis that the associations were consistent over time. In SAR models, our parameter of interest was a covariance coefficient termed rho (ρ), interpreted similarly to a regression coefficient. We found that employees whose co-workers purchased a larger proportion of green-labeled items themselves purchased a higher proportion of green-labeled items than is typical for an otherwise similar employee (Figure 2). For example, in 2015, compared to employees whose co-workers made average purchases of green-labeled items, those whose co-workers purchased a 10 percentage point higher proportion of green-labeled food items would be expected to themselves purchase 4.0 percentage points (99% CI 3.5 to 4.6, $p<0.0001$; Supplementary Table 2) more green-labeled food items. We observed positive, statistically significant associations for all four types of purchases assessed (green-/red-labeled foods/beverages), though there was variation in their magnitude, with point estimates for green-labeled items at least 85% higher than those for red-labeled items (non-overlapping 99% confidence intervals indicate these differences are at minimum statistically significant at the p<0.01 level; Supplementary Tables 2 and 3). For context, labels on food items purchased by members of the study cohort in 2015 were on average 34% green and 28% red, and for beverages were 35% green and 20% red (Supplementary Table 4). These associations and overall purchasing patterns were stable over time; models estimated using 2016 data yielded very similar results.

Longitudinal associations in healthful purchases

We then used generalized estimating equations (GEEs) to examine transactions over time at the pairwise employee (dyadic) level. This allowed for assessment of how the proportion of an employee's purchases that was labeled green (or red) at transaction t related to the proportion of items labeled green (or red) purchased by a co-worker at the previous transaction the two made together (t-1), adjusting for both employee and co-worker characteristics[41–43], dyad characteristics, as well as the food environment (purchase time, location). Dyads were defined based on networks estimated using 8-week intervals of data, mirroring the window in the employee survey that we used to validate tie definitions, while the time interval between t and t-1 varied by dyad/encounter. We conjectured that controlling for dyad characteristics and time of purchase may have helped minimize unmeasured homophily, as these elements may reflect shared characteristics involved in tie formation. To assess the extent to which our models might be unduly influenced by poorly measured or weak ties, we estimated a series of models (Figure 3). The first included all ties (based on coincident purchasing), despite our strong assumption that many were not, in reality, social ties (Supplementary Table 5). Subsequently, we estimated models using increasing tie probability thresholds, beginning at ≥0.1 and increasing to ≥0.6. The latter was the preferred model as this cut-off yielded a high positive predictive value for true ties based on the employee survey – Figure 1b). It is clear in Figure 3 that including all ties provided narrow confidence intervals, but with potentially misleading results. Restricting models to dyads more likely to represent strong or accurately measured ties showed much larger associations at the cost of some reduction in precision. In sensitivity analyses explicitly measuring the relationship between the association and tie probability, we found a statistically significant positive relationship (Supplementary Table 5a).

Focusing on the strictest definition of a social tie (tie probability ≥0.6; n=1,441 foods, n=1,138 beverages), we estimated that when the proportion of green-labeled items in co-workers' purchases increased by 10 percentage points, there was an associated 4.5 (99% CI 3.2 to 5.8, p<0.001) percentage point increase in the expected proportion of green-labeled items in employees' subsequent purchases (Figure 3, Green-labeled Foods; Supplementary Table 6a). A statistically significant association was also observed for red-labeled food items: if the proportion of red-labeled items in co-workers' purchases increased by 10 percentage points, there was an associated 2.9 (99% CI 1.8 to 4.0, p<0.001) percentage point increase in the expected proportion of red-labeled items in employees' subsequent purchases (Figure 3, Red-labeled Foods; Supplementary Table 6a). Co-workers' green- and red-labeled beverage purchases also had positive associations with employees' purchases, but they were not statistically significant at the (Bonferroni-corrected) $p<0.05$ level for the strictest tie probability thresholds (Supplementary Table 6b).

To evaluate an alternative tie definition that did not rely on demographic data, we estimated models restricted to dyads that made purchases at the same location and time on at least 5 occasions over the course of a given 8-week time interval (Supplementary Table 7). This threshold was chosen because of the sharply right-skewed distribution of repeated dyads (Supplementary Figure 3b). In addition, this inferred tie definition based on purely repeated co-location data may be easier to replicate in other settings. Both this threshold and the strictest social tie definition (including demographic data) comport with an approximately 2% sample of the all-ties dyad definition. Estimated associations from this model were generally in between those estimated when the tie probability was restricted to ≥0.6 and those estimated with no restrictions on tie probabilities (Supplementary Figure 7).

We conducted further sensitivity analyses to investigate the properties of the observed associations and to evaluate evidence regarding potential homophily. We explored whether there was evidence to suggest that the magnitude of associations differed depending on shared characteristics of dyads, for example whether dyads consisted of male-male, male-female, or female-female pairs (Supplementary Table 8a); same-race pairs (Supplementary Table 8b); same-department pairs (Supplementary Table 8c); or same-job type pairs (Supplementary Table 8d). Although there was some evidence of variation by same-department and same-gender dyads, results of these analyses were not consistent across outcomes, and sample sizes for some groups may have been too small to detect an effect. We also tested whether the strength of the associations diminished with elapsed time since the prior joint purchase (i.e., the elapsed time between t and t-1, which varied by dyad purchase encounter), and we found no difference in any of the four food or beverage models at the corrected $p<0.05$ significance level (Supplementary Table 6c). The presence of an interaction would have provided evidence against homophily in the sense that latent homophily effects are assumed to be unrelated to elapsed time, though the absence of an interaction does not imply there was homophily. We performed a similar test to adjudicate whether time of day moderated associations and found no evidence for such an interaction (Supplementary Table 6d). We also evaluated whether removing race (one of the covariates with the largest coefficients in our models and one associated with both the healthfulness of food purchases and social tie formation) affected the estimated associations in our models (Supplementary Table 9)[44,45]. Large changes in the estimated associations would have indicated that the associations reported in Figure 3 (we focus on the model with tie probability ≥0.6, Supplementary Tables 6a and 6b) would be likely to absorb the effects of homophily. We did not find that to be the case, suggesting our association estimates may be minimally biased by homophily. In another analysis, we estimated "*e-values*"[46] to assess how great confounding due to homophily would need to be to render our estimates statistically insignificant. We found the effect of an omitted confounder would need to have a combined relationship with the outcome (employees' purchases at transaction t) and predictor (co-workers' purchase with the employee at t-1) larger than any measured for the observed covariates in our models before our statistically significant measures were rendered non-significant (Supplementary Table 10). Finally, we conducted an instrumental variables (IV) analysis using purchases by a second degree coworker (the coworker's coworker) as a putatively exogenous source of variation in the coworker's purchasing, thus eliminating concerns of homophily to the extent that it manifests as unmeasured confounding (Supplementary Figure 9)[47]. For these analyses, we began with the same subset of data used above in Figure 3 involving the strictest definition of a social tie (tie probability ≥0.6), but to justify the IV exclusion restriction, we excluded any 2nd-degree alters of a given ego that were also 1st-degree alters of that same ego during the study period, as well 2nd-degree alters that were egos in the analysis themselves. Because of this, the sample size of egos in the IV models is approximately half that in the models reported in Figure 3. These IV analyses estimated significant and positive, though imprecisely-measured, peer effects for purchases of green-labeled foods (coefficient 0.40; 95% CI 0.06 to 0.73, p=0.02; Supplementary Table 11), and the magnitude of the effect was similar to that estimated in the GEE analysis. There was insufficient evidence for or against peer effects among red-labeled foods due to the inherent loss of precision in instrumental variables analyses (Supplementary Table 12). It is important to note that the validity of effects estimated using these instrumental variables models rests on the strong assumption that the 2nd-degree coworker's purchase was unrelated to the employee's purchase through any causal pathways other than through the coworker. These findings, while consistent with the SAR and GEE findings, should be interpreted with caution as violations of instrumental variables assumptions can lead to significant bias.

Discussion

Results of this research indicate that the healthfulness of food choices is correlated in a large workplace social network and reinforce the plausibility of food choice as an interpersonal mechanism for the transmission of obesity in a non-intimate social network. We observed interpersonal associations in food choice for both healthy and unhealthy foods. Because there is a middle ground in yellow-labeled purchases (we do not present analyses of yellow-labeled items here), green- and red-labeled purchases are largely independent of one another. Associations with beverage purchases were less robust, appearing in year-long cross-sectional analyses, but not in the strictest longitudinal analyses assessing associations from one transaction to the next. Weaker interpersonal associations for beverage purchases may reflect that beverage preferences are more routinized and fixed, and therefore less susceptible to social influence. Alternatively, these associations may be related to the smaller choice set available for beverages, including choosing freely available tap water. Our consistent findings for food purchases, which form the vast majority of calories purchased[29], are particularly important to the broader argument that socially transmitted food choice may be a mechanism by which obesity spreads through a social network[13,22,48].

We found that associations tended to be larger for green-labeled foods than for red-labeled foods. This was similar to findings from most one-time interventions tested in university-based naturalistic settings and in lab-based experimental settings, where it has been observed that healthy social norms have a positive effect on healthy food choice[49–52]. One study observed greater social influence on food choice for unhealthy foods than for healthier foods[53].

Our data and methods extend prior research investigating social influence and food choice behaviour that has typically been conducted in small experimental settings. Earlier studies have focused on young study subjects, generally under the age of 25 (for reviews, see[10,17,54]). The mean age of our study subjects was 45, ranging from 19 to 83. In addition, prior studies were usually lab-based or conducted in otherwise tightly controlled settings that were restricted in the number of interactions, the nature of social contacts, or the range of food choices available to participants. Thus, existing studies were not generalizable to more diverse populations in real-world settings. In contrast, we investigated social influence on food choice in the workplace, where most people spend a great deal of their day in the company of consistent, but non-intimate, acquaintances. This social setting has been given far less scrutiny, particularly with respect to health behaviours, than ties among close social intimates. Our study included a multi-ethnic sample of 6,665 socioeconomically diverse men and women engaging in nearly 3 million dyadic interactions overall and who had a median of two interactions with one another over a two-year period (first quartile=2, third quartile=3, maximum=262, see Supplementary Figure 2). We observed behaviour in a food environment spanning 7 cafeterias that offered a wide range of food items; over 2,300 unique items are available, including entrees, snacks, desserts, soft drinks, and coffee drinks. This real-world setting allowed us to observe social and behavioural dynamics as they occurred naturally.

A number of additional elements of this work increased the robustness of the findings. The SAR models yielded near identical results from one year to the next, demonstrating the consistency of our findings in the context of evolving employee populations, pairings, and food choices. In addition, for food purchases, we saw qualitatively similar results in both cross-sectional SAR analyses, which have the advantage of including contemporaneous purchases and a pattern of behaviour exhibited over a year, and in longitudinal GEE analyses that accounted for the temporal relationship between one person's food choices and another's, and also accounted for individuals' personal attributes and their shared dyad characteristics.

Robustness was also strengthened by our data sources. Though the study was observational in nature, we were able to capture objective data on food purchases rather than relying on subjective self-reports of consumption. By analysing secondary administrative data captured over the span of two years, our assessments reflected changes of behaviour over time rather than at a single point in time under experimental circumstances. Furthermore, we identified social connections using passively-collected purchasing data rather than directly observed or self-reported information. This allowed inference of social connections on a large scale and in an objective fashion that was not influenced by social desirability or other reporting biases that may arise in primary data collection. While the practice of inferring social connections based on objectively-assessed temporal and spatial proximity is not new, it has not generally been applied to study food-related behaviours, in a relational context, using a validated measure of tie strength derived from participant observation and survey data. The stronger associations observed when the longitudinal models were restricted to dyads with a higher probability of being "true" helps to support the validity of both our conclusion that we are observing prospective associations consistent with a social influence process and our evidence that ties are "true." Our estimated tie probability may reflect the probability that an observed tie is present in reality, or alternatively, it may be a measure of the strength of the tie between two people. In either event, we would expect higher values to be associated with stronger estimated associations in purchasing behaviour. Similarly, if associations are real, we would expect that they would be attenuated in situations where social ties are weaker or poorly measured, which is what we found.

Prior network research has used GEEs to estimate marginal behavioural associations in network settings that exploited tie directionality to identify a peer effect[13,41]. Due to the non-independence of network data, using a GEE framework to separately account for homophily, confounding, and simultaneity mechanisms has proven to be challenging[13,55,56], though these models have been shown to be fairly robust[43,57]. By using alter data from a prior time period to predict ego purchase behaviour at a current time period, our longitudinal GEE models avoid violating temporal dependence assumptions that have complicated earlier work. We sought to reduce the extent to which the measured associations captured homophily by controlling for dyad characteristics and for time of day when the ego's purchases were made. Additional sensitivity analyses suggested that unmeasured confounding due to homophily was not likely to be a major source of bias (see Supplementary Tables 9, 10, and 13). Nevertheless, we cannot rule out the likelihood that at least some portion of the measured association in this observational study reflects homophily. The exception to that caveat comes from our instrumental variables analysis which, if its assumptions hold, would remove all bias due to homophily, confounding, or simultaneity. We reiterate our caution in accepting that our available instrument satisfies the exclusion restriction in the presence of homophily, confounding, or simultaneity and the independence assumption required for valid instrumental variables estimation[47]. Future work could consider using different instrumental variables (including the special case of randomized assignment to peers/networks) in a completely identified, closed network, though these are challenging methods to execute in real-world scenarios.

Along with the strengths noted above, our study has certain limitations. While based on objective data and a validated model, information on social ties was inferred and contains some degree of error. Approximately 5% of the hospital's 26,000 employees had observed connections with a 60% or greater probability of being true social ties according to the tie prediction model. This could limit the reach of interventions that rely on contacting known employee/co-worker dyads. Another limitation is the fact that the objectively inferred ties did not reveal whether ties or influence were one-sided or bidirectional. While we were able to objectively determine what food individuals purchased, the information in these purchases was incomplete. We could not observe what employees actually consumed, nor could we know about foods purchased by or for others, foods purchased with cash, or foods acquired outside of the cafeteria system. Lastly, our analyses focused on pair-wise relationships only, and did not assess whether peer influence is amplified when groups collectively exert behavioural influence on individuals.

While the present research relies on specialized, site-specific data sources, its insights can be constructively applied more broadly. Society has experienced a rapid expansion of systems gathering data to track human activity, whether it be in the context of consumer behaviour (for example, customer loyalty programs), online social networks (both general and those focused on health and wellbeing[58]), “smart” devices such as phones, watches, and fitness trackers, and individually identifiable cards and keys (e.g. RFID technology). As a result, tools to explore and apply insights on social interactions and human health are ubiquitous and still evolving. Given the high proportion of time many adults spend at work, many eating decisions are made in the presence of co-workers[59]. Our findings highlight the potential for using food purchasing data to test targeted messages or incentives to effect positive changes in health behaviours, particularly in a workplace environment. Furthermore, recent work in this same employee population has shown that the healthfulness of workplace food purchases is associated with employees’ overall diet quality, and with their risk of obesity, prediabetes, and hypertension[60]. It is reasonable to assume that social correlations in food choice may influence a range of chronic conditions. Insights on social connections and food choice have the potential to yield new tools for improving public health.

Materials and Methods

Ethics and human subjects’ approval

This research was approved by the Institutional Review Board (IRB) of Partners HealthCare (parent institution of Massachusetts General Hospital) on April 14, 2016. Analyses of these deidentified secondary (administrative) data conducted at the University of Massachusetts, Amherst relied on Partners’ IRB via inter-institutional agreement. This was an observational study of deidentified secondary administrative data so the IRB determined that the need for explicit subject consent could be waived for our main analyses examining associations in cafeteria purchases across inferred employee dyads. It is worth stating explicitly that although this study was determined to be “low-risk” to subjects, these secondary data were aggregated for the purposes of this study in ways that cafeteria patrons likely did not envision when they used their employee IDs to purchase items[61]. However, small cell sizes in covariate descriptions are not reported, socio-demographic covariates are reported using common generic categories, and informationally risky network visualizations based upon inferred network data are not included here[62]. Separately, participants in the tie-validation study did provide explicit consent for their participation in the survey and for linking their survey responses to cafeteria purchasing data. As is common in low-risk survey studies, completion of the survey constituted implied consent, and no signature was solicited. Linked survey/purchasing data were used to construct the subsequently deidentified data file used in the tie validation analyses. Our analyses of purchasing associations across dyads used deidentified data exclusively.

Setting and Participants

The research took place at Massachusetts General Hospital (MGH), a large teaching hospital with over 26,000 employees. MGH is the largest non-government employer in eastern Massachusetts. Study data were collected from cash registers in the 7 cafeterias located on the MGH main campus that were in business during the study period. Though hours varied from one venue to another, the cafeterias were open for breakfast lunch, and dinner (one cafeteria is also open late for the night shift) and served hot meals, salad bar options, grab-and-go sandwiches and similar items, snacks, and beverages, including prepared-to-order coffee drinks. The cash register purchasing database included information on the time, location, and items purchased for all transactions.

Employees at the hospital have the option of using their ID cards to make purchases that are debited directly from their pay checks. This allowed us to link transactions to individuals and their demographic characteristics using human resources data. Approximately 7,000 employees are enrolled in the program at any point in time. The cafeterias conduct approximately 71,000 transactions in an average week; 19,800 of those transactions are conducted by employees using their ID cards. Our study population consisted of 6,665 MGH employees who used their IDs at least once to make either a food or beverage purchase at the same time and place as a co-worker at MGH cafeterias during the period January 1, 2015-December 31, 2016. A description of employee characteristics can be found in the Supplementary Information Appendix (Supplementary Table 4); these employees were closely representative of MGH employees overall.

Food Labelling

In 2010, members of our research team (DEL and ANT) worked with the MGH Nutrition and Food Services to implement the “Choose Well, Eat Well” program in the main MGH cafeteria, where all foods and beverages were given traffic light labels – green (healthy), yellow (less healthy), or red (unhealthy); the program was designed to provide patrons with simple information regarding the healthfulness of food items sold in the cafeteria. Since 2014, the labelling system has been implemented in all MGH cafeterias, and it is well-known and understood by most employees[12,63,64]. The labelling system was originally based on 2005 USDA dietary guidelines[65] and has been updated with the 2010 and 2015 revisions[66,67]. Green-labeled items have a main ingredient of whole grains, lean protein, and/or fruits/vegetables, and are low in calories. Red-labeled items are high in calories and/or saturated fat and have little nutritional value. Prominent messaging in the cafeterias explains the labelling system. These color designations are included for all items in the cash register database and were used in our analyses to differentiate healthy (green) versus unhealthy (red) choices. We do not separately analyse purchases of yellow-labeled items because it is difficult to interpret whether changes in these purchases reflect shifts towards or away from healthy foods. We also note that, consistent with prior research on this employee sample and the availability of data, we made the decision to only analyse cold beverage purchases (simply termed “beverages” throughout the manuscript) from all 7 cafeterias on campus[12,64]. These cold beverages consist of items such as bottles, cans, and cartons of soft drinks, juices, milk, and water. It was not possible to assign traffic light labels to hot beverages (coffee, tea) prepared by customers because information on milk, cream, and sugar additions was not available. In the Supplementary Information, we explicitly label cold and hot beverages as necessary, and in Supplementary Table 14, we offer a sensitivity analysis of staff-prepared hot beverages from data at the 2 eating locations that were coffee shops. This sensitivity analysis reveals only a modest association between ego and alter green hot beverage purchases, consistent with longitudinal analysis of cold beverages.

Estimating Social Tie Probabilities

Based on ethnographic observation over 3 weeks in the MGH cafeterias and a separate validation survey of employees to discern the frequency and likelihood of eating with particular types of co-workers, we limited potentially legitimate social ties to purchases made within +/− 2 minutes of one another. We linked the survey data containing the self-reported information on social ties with demographic and purchasing data to calculate a logistic regression and subsequently the predicted probability that a social tie was likely to be real as a function of similarity in age (within a 5-year age difference), same gender, same race, same workplace department, same job type, number of joint purchasing events observed within discrete 8-week periods within a year, and number of separate cafeterias at which dyads were observed making purchases together. Additional details of the tie prediction survey and validation methods are described in the Supplementary Information Appendix (Part 2. Employee survey of commensal purchasing behaviours, Supplementary Figure 4, Supplementary Table 1).

Procedure and Design.

We first sought to test cross-sectional associations between employee food purchases during each calendar year by estimating simultaneously autoregressive (SAR) models. Because we found significant associations between ego and alter food purchases, we then sought to test for longitudinal associations between an ego's purchases and an alter's purchases during their most recent prior transaction together using marginal regression models estimated using generalized estimating equations (GEE).

Supplementary Figures 1 and 2 describe the data processing workflow for SAR and GEE models, and Supplementary Figure 5 illustrates conceptual models. All analyses relied upon the same initial number of dyads, but within-year cross-sectional SAR and longitudinal purchase event-lagged GEE models diverged due to different model specification requirements. Stata 15/MP[68] and R 3.5.2-4.0[69] were used for data processing and analyses. SAR models extended previously-published R code[70], GEE models were estimated using the geepack package v.1.2.1 in R[71]. Instrumental variables models used the ivreg package v0.5 in R[72] and fracivp in Stata[73]. Stata 15.1 was used for receiver operation characteristic (ROC) curve analyses to test specificity and sensitivity of tie prediction. All statistical tests are Wald tests for coefficients (null hypothesis: coefficient is equal to 0) and are reported with two-sided p-values.

Simultaneously Autoregressive (SAR) Models.

We estimated two yearly cross-sectional SAR models as a first step towards understanding whether employees' food purchases were, over the course of a year, associated with one another. These models account for the direct effect of the adjusted weighted average of all alters' yearly purchases on the average of an ego's yearly purchases, while taking into consideration that people may have had repeated meals with a variety of different people (i.e. autocorrelation among a network of employees). This family of spatial models was originally developed to understand geographic dependence between adjacent areas, but has been adopted for use in network research[39,40,70].

$$Pct\,Purchase_{i,\,color} = \beta_0 + \beta_1 EC_i + \varepsilon_i \text{ where } \varepsilon \sim \left[\mathbf{0}, \sigma^2 \left\{ \left(\mathbf{I} - \rho W^T\right)\left(\mathbf{I} - \rho W\right) \right\}^{-1}\right] \quad \text{Eq. 1.}$$

Our model is described in Equation 1, where the dependent variable is the percent of items purchased over a 12-month period that are labeled green (or red) for employee i. SAR models employ an adjacency matrix, W, to incorporate information on the frequency of ties between subjects. To this matrix we applied weights defined using the predicted tie probabilities from our tie validation model to increase the numeric importance of relationships with a high probability of being real and reduce the impact of ties that were potentially spurious. Because the tie probability model was based on survey responses which asked participants using an 8-week recall frame, tie probabilities applied in the adjacency matrices are based on 8-week blocks of time. In the adjacency matrix, any repeated tie observations were collapsed into a single tie that had a weight equal to the sum of the repeated observations. A conceptual illustration of this model is given in the Supplementary Figure 5, Panel B.

The measure of association estimated in SAR models is a parameter, ρ, which is based on the model's residuals, and therefore captures the influences of co-workers' purchases on egos' purchases based on how different employees' purchases are from the expected (average), conditional on employee characteristics. Specifically, ρ measures the conditional association between a 1-unit increase in the ego's alters' weighted residual (their observed purchases less their predicted purchases based on their covariates) and the ego's purchases. This quantity allowed us to evaluate whether the combined impact of exposure to alters' purchases (weighting alters according to how frequently and strongly they were connected to the ego) was associated with a change in ego's purchases. The term EC_i contains a vector of ego-level covariates including indicators for female sex, age categories (17–29, 30–39, 40–49, 50–60+), race/ethnicity (White, Black, Asian, Hispanic, other – categories as listed in our human resources data source), job type (administrative support, management/ clinician, professionals, service, technicians), and educational attainment (high school or less, some college, college, advanced degree, missing). For context, employees included in these analyses had a mean (sd) age of 42 (12) and 78% were female. Full model results are reported in the Supplementary Information Appendix (Supplementary Tables 2–3)

Generalized Estimating Equations.

Population-average (or marginal) regression models are useful for estimating outcomes of interest under circumstances where it is assumed that an individual's behaviours are correlated with one another[42,74]. As these models are also appropriate to model dyadic influence[41,55], we estimated GEE regression models that clustered on ego-level observations over time. To crudely control for time trends and seasonality and to match employees' 8-week retrospective survey recall when establishing employee/co-worker dyads, we divided each calendar year into roughly 8-week periods (because of an uneven number of weeks, the last period of the year had 10 weeks). We excluded purchases during the last two weeks in December because of the significant drop-off in workplace attendance inducing aberrant purchasing patterns, and then pooled dyadic data from each period for analysis (see Supplementary Figure 5, Panel C). The longitudinal models estimated in this study are described by Equation 2.

$$f\left(E\left(PropPurchase_{i,\,t}\right)\right) = \beta_{0t} + \beta_1\left(A_{i,\,t-1}\right) + \beta_2 EC_i + \beta_3 AC_i + \beta_4 S_i + \beta_5 N_i \qquad \text{Eq. 2.}$$

In the GEE model described in Equation 2, $PctPurchase_{i,t}$ is the proportion of the ego's purchase at time t that is labeled green (or, in a separate model, red) as a function of the proportion of food or beverages labeled green (or red) purchased by the alter at the prior occasion where the ego and alter made a purchase together ($A_{i,t\text{-}1}$). (Note that here t-1 does not refer to a fixed time interval, but rather refers to the occasion of the prior purchase the dyad made together, regardless of the elapsed time.) The peer association was captured in coefficient b_1. The proportion is specified using a logistic link function because the data-generating process constrains the outcome to be between 0.0 and 1.0 and is heavily skewed towards the tails. The term EC_i represents a vector of ego socio-demographic characteristics (gender, age, race/ethnicity, job type, educational attainment). For context, the mean age (sd) of employees included in the food models was 42 (13) and 78% were female. The term AC_i represents the alter's same socio-demographic characteristics (omitting alter gender because of a dyadic gender covariate described below). The term S_i represents a vector of dyadic similarity measures (including ego and alter having the same race, job type, department, and one of three gender combinations (MF, FF, MM); and the term N_i represents an ego's number of unique alters during the entire 24-month period, the time of day purchases were made (before lunch vs. during/after lunch – see Supplementary Figure 3c), a set of indicators for twelve 8-week periods during the 2015–16 period (omitting two-week end-of-year holiday periods in each year), and the estimated tie probability for the dyads. All models were estimated using robust standard errors and were specified with an independent correlation structure. Sample characteristics (Supplementary Table 15) and fully adjusted estimates for GEE models (Supplementary Tables 5–10, 14) are reported in the Supplementary Information Appendix.

We conducted several sensitivity analyses for the GEE models. Our initial model simply included tie probability, estimated from the predictive model, as a covariate (this is the "all-tie" model in Fig. 3; model estimates are in Supplementary Table 5). In a subsequent model, we tested the interaction between alter's purchases ($A_{i,t\text{-}1}$) and the tie probability to assess whether the association was stronger when the estimated tie probability was higher (Supplementary Table 5a). Upon determining that the interaction was strong, we defined our preferred analyses as models limited to dyads with inferred tie probabilities ≥0.6 (based on a cut-off where the positive predictive value was above 75%, Figure 1c), while also presenting intermediate models restricted to tie probabilities ≥0.1 - ≥0.6 for illustration (Figure 3). We also estimated an alternative model where ties were inferred to exist when dyads made coincident purchases ≥5 times during the 24-month study period (Supplementary Table 7). This co-location threshold-based tie definition was estimated with purchasing data only and no demographic information. Lastly, we conducted a series of analyses described in the Supplementary Information Appendix that helped minimize concerns of bias due to homophily. We reassigned purchasing data randomly across the alters (Supplementary Table 13). We assessed whether the association diminished in magnitude as the duration between *t* and *t*-1 increased by testing the interaction between this duration and the alter's purchases ($A_{i,t\text{-}1}$) (Supplementary Table 6c). We removed race from the models, as it was the confounder with the largest coefficient and is known to be related to food choice and tie formation (Supplementary Table 9). And we calculated "e-values" to determine how sensitive our findings might have been to omitted variables, including those related to tie formation (Supplementary Table 10).

For all models, we adjusted p-values for multiple hypothesis tests using a Bonferroni correction when reporting separate models for ego's green food/beverage purchases and red food/beverage purchases.

Instrumental Variables Analysis

In the sensitivity analyses employing instrumental variables methods, we used two-stage least squares estimation including only dyads meeting the ≥0.6 tie probability threshold. The data setup for each stage (i.e., the timing of purchases included as dependent and independent variables) was identical to that used in the GEE analyses. In the first stage regression, purchases by 2^{nd}-degree alters (the primary alters' alters – see Supplementary Figure 9) were used as an instrument for primary alters' purchases[47]. To strengthen the plausibility of the exclusion restriction, only purchases by secondary alters who did not appear elsewhere in the data as either first-degree alters or egos were included as instruments. This reduced the number of unique egos in the data by approximately half, when compared to models reported in Figure 3 and Supplementary Tables 6a and 6b. It also required a more parsimonious regression model, so the instrumental variables models controlled for ego and dyad characteristics, but not alter characteristics. Standard errors in all analyses were adjusted for clustering by ego, and p-values adjust for multiple hypothesis tests as in the GEE analyses above.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported in part by a National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) R21-DK109548 grant to DEL and a National Heart Lung and Blood Institute (NHLBI) R01-HL125486 grant to ANT. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

References

1. Cutler DM, Glaeser EL & Shapiro JM Why have Americans become more obese? Journal of Economic perspectives 17, 93–118, (2003).
2. Gortmaker SL et al. Changing the future of obesity: science, policy, and action. The Lancet 378, 838–847, (2011).
3. Calle EE, Rodriguez C, Walker-Thurmond K & Thun MJ Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. N Engl J Med 348, 1625–1638, (2003). [PubMed: 12711737]

4. Finucane MM et al. National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9· 1 million participants. The Lancet 377, 557–567, (2011).
5. Mokdad AH et al. The State of US Health, 1990–2016 Burden of Diseases, Injuries, and Risk Factors Among US States. Jama-J Am Med Assoc 319, 1444–1472, (2018).
6. Wang YC, McPherson K, Marsh T, Gortmaker SL & Brown M Health and economic burden of the projected obesity trends in the USA and the UK. Lancet 378, 815–825, (2011). [PubMed: 21872750]
7. Preston SH, Vierboom YC & Stokes A The role of obesity in exceptionally slow US mortality improvement. Proceedings of the National Academy of Sciences 115, 957–961, (2018).
8. Ball K, Jeffery RW, Abbott G, McNaughton SA & Crawford D Is healthy behavior contagious: associations of social norms with physical activity and healthy eating. International Journal of Behavioral Nutrition and Physical Activity 7, 86, (2010).
9. Chevalier M Increase in sales due to in-store display. J Marketing Res 12, 426–431, (1975).
10. Cruwys T, Beyelander KE & Hermans RCJ Social modeling of eating: A review of when and why social influence affects food intake and choice. Appetite 86, 3–18, (2015). [PubMed: 25174571]
11. Downs JS, Loewenstein G & Wisdom J Strategies for promoting healthier food choices. American Economic Review 99, 159–164, (2009).
12. Thorndike AN, Sonnenberg L, Riis J, Barraclough S & Levy DEA 2-phase labeling and choice architecture intervention to improve healthy food and beverage choices. American journal of public health 102, 527–533, (2012). [PubMed: 22390518]
13. Christakis NA & Fowler JH The spread of obesity in a large social network over 32 years. New Engl J Med 357, 370–379, (2007). [PubMed: 17652652]
14. de la Haye K, Robins G, Mohr P & Wilson C Homophily and Contagion as Explanations for Weight Similarities Among Adolescent Friends. J Adolescent Health 49, 421–427, (2011).
15. Valente TW, Fujimoto K, Chou C-P & Spruijt-Metz D Adolescent affiliations and adiposity: A social network analysis of friendships and obesity. J Adolescent Health 45, 202–204, (2009).
16. Fischler C Commensality, society and culture. Social Science Information 50, 528–548, (2011).
17. Herman CP The social facilitation of eating. A review. Appetite 86, 61–73, (2015). [PubMed: 25265153]
18. Higgs S Social norms and their influence on eating behaviours. Appetite 86, 38–44, (2015). [PubMed: 25451578]
19. McLean PD Culture in networks. (Polity, 2017).
20. Pachucki MA & Breiger RL Cultural Holes: Beyond Relationality in Social Networks and Culture. Annual Review of Sociology, Vol 36 36, 205–224, (2010).
21. Pachucki MC et al. Eating with others and meal location are differentially associated with nutrient intake by sex: The Diabetes Study of Northern California (DISTANCE). Appetite 127, 203–213, (2018). [PubMed: 29601920]
22. Pachucki MA, Jacques PF & Christakis NA Social Network Concordance in Food Choice Among Spouses, Friends, and Siblings. American journal of public health 101, 2170–2177, (2011). [PubMed: 21940920]
23. Rosenquist JN, Murabito J, Fowler JH & Christakis NA The spread of alcohol consumption behavior in a large social network. Ann Intern Med 152, 426–433, W141, (2010). [PubMed: 20368648]
24. U.S. Bureau of Labor Statistics. Average hours per day spent in selected activities on days worked by employment status and sex. (2019).
25. Morrison RL & Cooper-Thomas HD in The psychology of friendship (eds Hojjat M & Moyer A) 123–140 (Oxford University Press, 2016).
26. Rumens N Researching workplace friendships: Drawing insights from the sociology of friendship. J Soc Pers Relat 34, 1149–1167, (2017).
27. Lazer D et al. Computational Social Science. Science 323, 721–723, (2009). [PubMed: 19197046]
28. Edelmann A, Wolff T, Montagne D & Bail CA Computational Social Science and Sociology. Annu Rev Sociol 46, (2020).

29. Thorndike AN, Gelsomin ED, McCurley JL & Levy DE Calories Purchased by Hospital Employees After Implementation of a Cafeteria Traffic Light–Labeling and Choice Architecture Program. JAMA network open 2, e196789–e196789, (2019). [PubMed: 31290988]
30. Thorndike AN, Riis J & Levy DE Social norms and financial incentives to promote employees' healthy food choices: A randomized controlled trial. Prev Med 86, 12–18, (2016). [PubMed: 26827617]
31. Smith KP & Christakis NA Social networks and health. Annu Rev Sociol 34, 405–429, (2008).
32. Zhang JW & Centola D Social Networks and Health: New Developments in Diffusion, Online and Offline. Annual Review of Sociology, Vol 45 45, 91–109, (2019).
33. VanderWeele TJ & Christakis NA Network multipliers and public health. International journal of epidemiology, (2019).
34. Barnett ML, Landon BE, O'Malley AJ, Keating NL & Christakis NA Mapping physician networks with self-reported and administrative data. Health Serv Res 46, 1592–1609, (2011). [PubMed: 21521213]
35. Crandall DJ et al. Inferring social ties from geographic coincidences. P Natl Acad Sci USA 107, 22436–22441, (2010).
36. Hunter RF et al. "Hidden" Social Networks in Behavior Change Interventions. American journal of public health 105, 513–516, (2015). [PubMed: 25602895]
37. Lauw HW, Lim E-P, Pang H & Tan T-T Social network discovery by mining spatio-temporal events. Computational & Mathematical Organization Theory 11, 97–118, (2005).
38. Matthews LJ, DeWan P & Rula EY Methods for inferring health-related social networks among coworkers from online communication patterns. Plos One 8, e55234, (2013). [PubMed: 23418436]
39. Paez A, Scott DM & Volz E Weight matrices for social influence analysis: An investigation of measurement errors and their effect on model identification and estimation quality. Soc Networks 30, 309–317, (2008).
40. Wang W, Neuman EJ & Newman DA Statistical power of the social network autocorrelation model. Soc Networks 38, 88–99, (2014).
41. Christakis NA & Fowler JH Social contagion theory: examining dynamic social networks and human behavior. Statistics in Medicine 32, 556–577, (2013). [PubMed: 22711416]
42. Hubbard AE et al. To GEE or Not to GEE Comparing Population Average and Mixed Models for Estimating the Associations Between Neighborhood Risk Factors and Health. Epidemiology 21, 467–474, (2010). [PubMed: 20220526]
43. Zachrison KS, Iwashyna TJ, Gebremariam A, Hutchins M & Lee JM Can longitudinal generalized estimating equation models distinguish network influence and homophily? An agent-based modeling approach to measurement characteristics. BMC medical research methodology 16, 174, (2016). [PubMed: 28031023]
44. Levy DE, Riis J, Sonnenberg LM, Barraclough SJ & Thorndike AN Food choices of minority and low-income employees: a cafeteria intervention. American journal of preventive medicine 43, 240–248, (2012). [PubMed: 22898116]
45. Smith JA, McPherson M & Smith-Lovin L Social Distance in the United States Sex, Race, Religion, Age, and Education Homophily among Confidants, 1985 to 2004. Am Sociol Rev 79, 432–456, (2014).
46. VanderWeele TJ & Ding P Sensitivity analysis in observational research: introducing the E-value. Annals of internal medicine, (2017).
47. O'Malley AJ, Elwert F, Rosenquist JN, Zaslavsky AM & Christakis NA Estimating Peer Effects in Longitudinal Dyadic Data Using Instrumental Variables. Biometrics 70, 506–515, (2014). [PubMed: 24779654]
48. De la Haye K, Robins G, Mohr P & Wilson C Obesity-related behaviors in adolescent friendship networks. Soc Networks 32, 161–167, (2010).
49. Burger JM et al. Nutritious or delicious? The effect of descriptive norm information on food choice. Journal of Social and Clinical Psychology 29, 228–242, (2010).
50. Mollen S, Rimal RN, Ruiter RA & Kok G Healthy and unhealthy social norms and food selection. Findings from a field-experiment. Appetite 65, 83–89, (2013). [PubMed: 23402712]

51. Prinsen S, de Ridder DT & de Vet E Eating by example. Effects of environmental cues on dietary decisions. Appetite 70, 1–5, (2013). [PubMed: 23791633]

52. Salmon SJ, Fennis BM, de Ridder DT, Adriaanse MA & De Vet E Health on impulse: When low self-control promotes healthy food choices. Health Psychology 33, 103, (2014). [PubMed: 23477580]

53. Robinson E & Higgs S Food choices in the presence of 'healthy'and 'unhealthy'eating partners. British journal of nutrition 109, 765–771, (2013).

54. 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).

55. O'Malley AJ The analysis of social network data: an exciting frontier for statisticians. Stat Med 32, 539–555, (2013). [PubMed: 23023735]

56. Shalizi CR & Thomas AC Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. Sociol Method Res 40, 211–239, (2011).

57. VanderWeele TJ Inference for influence over multiple degrees of separation on a social network. Statistics in medicine 32, 591–596, (2013). [PubMed: 23341081]

58. Aral S & Nicolaides C Exercise contagion in a global social network. Nat Commun 8, (2017).

59. Linnan L, Fisher EB & Hood S The Power and Potential of Peer Support in Workplace Interventions. Am J Health Promot 28, Tahp2–Tahp10, (2013).

60. McCurley JL et al. Association of Worksite Food Purchases and Employees' Overall Dietary Quality and Health. American journal of preventive medicine, (2019).

61. Salganik MJ Bit by bit : social research in the digital age. (Princeton University Press, 2018).

62. Adams J Gathering social network data. 1 Edition. edn, (SAGE Publications, 2019).

63. Sonnenberg L et al. A traffic light food labeling intervention increases consumer awareness of health and healthy choices at the point-of-purchase. Preventive medicine 57, 253–257, (2013). [PubMed: 23859926]

64. Thorndike AN, Riis J, Sonnenberg LM & Levy DE Traffic-light labels and choice architecture: promoting healthy food choices. American journal of preventive medicine 46, 143–149, (2014). [PubMed: 24439347]

65. US Department of Health & Human Services. United States Department of Agriculture: Dietary Guidelines for Americans, 2005. Washington, DC: US Government Printing Office, (2004).

66. Dietary Guidelines Advisory Committee. Dietary guidelines for Americans 2015–2020. (Government Printing Office, 2015).

67. United States Dietary Guidelines Advisory Committee. Dietary guidelines for Americans, 2010. (US Department of Health and Human Services, US Department of Agriculture, 2010).

68. Stata Statistical Software: Release 15 (StataCorp LP, College Station, TX, 2017).

69. R: A language and environment for statistical computing. R Foundation for Statistical Computing (Vienna, Austria, 2018).

70. O'Malley AJ & Marsden PV The Analysis of Social Networks. Health Serv Outcomes Res Methodol 8, 222–269, (2008). [PubMed: 20046802]

71. Højsgaard S, Halekoh U & Yan J The R Package geepack for Generalized Estimating Equations Journal of Statistical Software. Journal of Statistical Software 15, 1–11, (2006).

72. Fox J, Kleiber C & Zeileis A ivreg: Instrumental-Variable Regression by 2SLS for R. (2020).

73. Williams R fracivp -- fractional response ivprobit models for Stata (beta). (2020).

74. Hardin JW & Hilbe JM Generalized estimating equations. (Chapman and Hall/CRC, 2002).

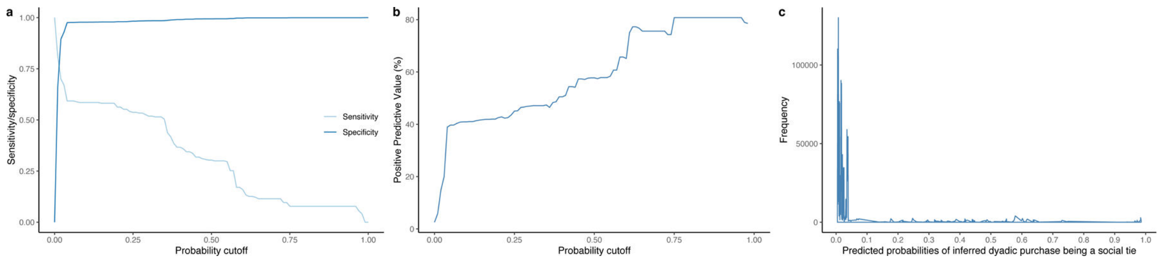

Figure 1. Properties of the predictive model for identifying social ties using cafeteria transaction and human resources data.

(a) Sensitivity/specificity trade-off; (b) Positive predictive value by probability cut-off; (c) Distribution of estimated tie probabilities. n = 1,054 for panels a and b. n = 2,974,388 dyads for panel c.

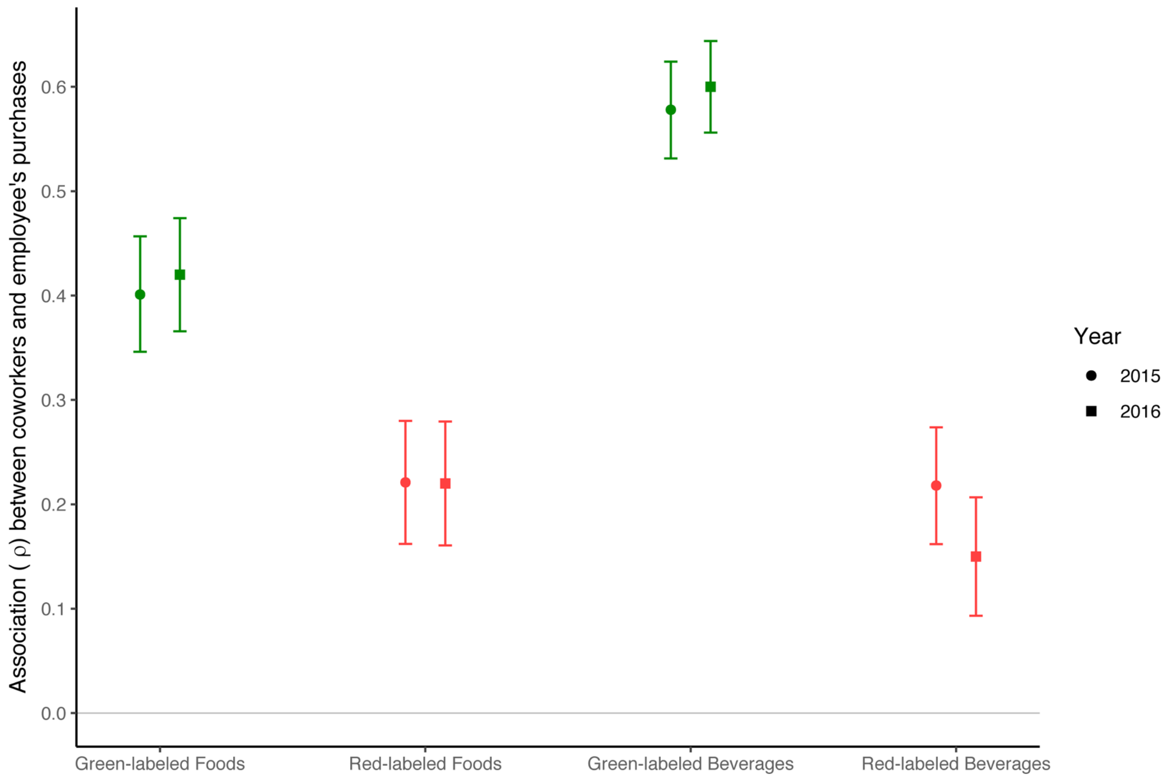

Figure 2. Yearly associations between employee and co-worker purchases.

Coefficients (ρ) and 99% confidence intervals from simultaneously autoregressive models assessing associations between the healthfulness of employee and co-workers' purchases of healthy and unhealthy foods and beverages over one-year cross-sections of time, adjusting for employee characteristics. Green- and red-labeled foods, 2015, n=5,934; green- and red-labeled foods, 2016, n=5,929; green- and red-labeled beverages, 2015, n=5,550; green- and red-labeled beverages, 2016, n=5,492 (Supplementary Tables 2 and 3).

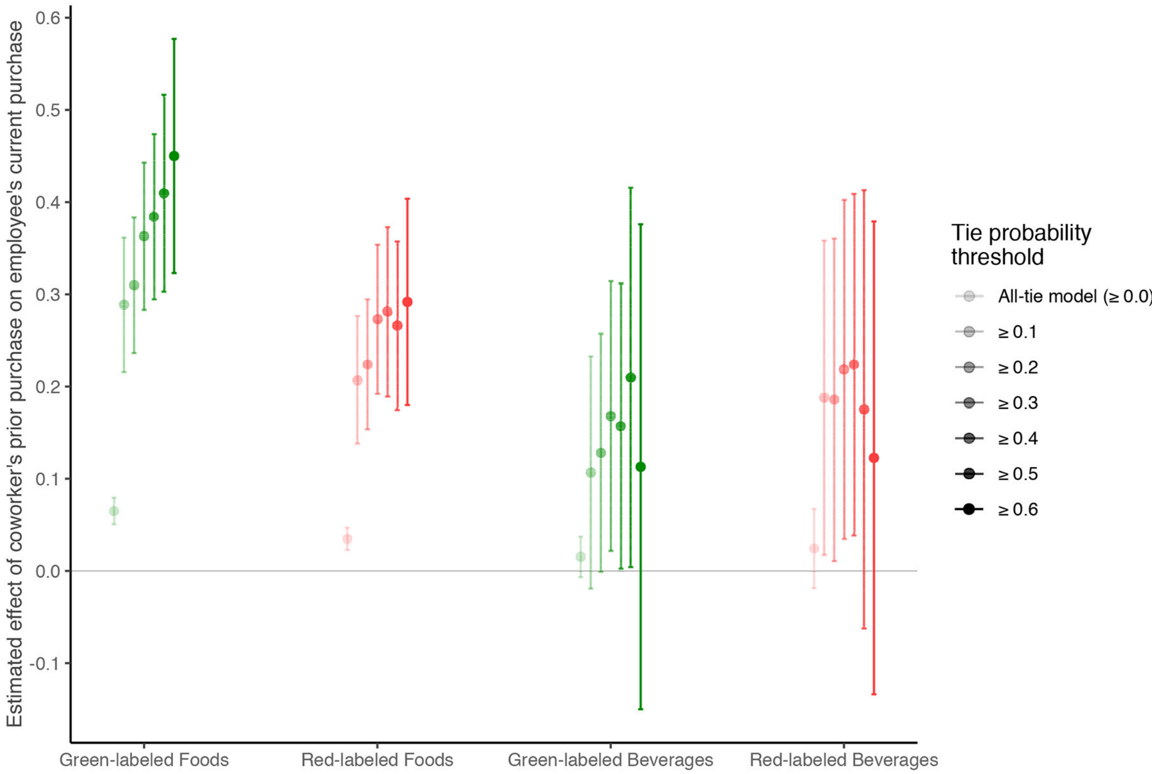

Figure 3. Prospective associations between employee' current and co-worker's prior purchase.
Coefficients and 99% confidence intervals of generalized estimating equations assessing the relationship between the healthfulness of co-workers' purchases and employees' subsequent purchases over 2015–2016, adjusting for employee, co-worker, and dyad characteristics, stratified by tie probability thresholds defining dyads. Higher tie probabilities indicate greater confidence in inferred social ties. Sample sizes decreased with increasingly restrictive tie probability thresholds (all-tie model through ≥0.6). They were n=6,382 and n=1,441 for foods and n=5,642 and n=1,138 for beverages, respectively (Supplementary Tables 5 and 6).