Journal of Health Economics 27 (2008) 1382–1387

Contents lists available at ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic☆,☆☆

Ethan Cohen-Cole[a,1], Jason M. Fletcher[b,*]

[a] Federal Reserve Bank of Boston, 600 Atlantic Avenue, Boston, MA 02210, United States
[b] Yale University, School of Public Health, Division of Health Policy and Administration, 60 College Street, New Haven, CT 06510, United States

ARTICLE INFO

Article history:
Received 12 December 2007
Received in revised form 16 April 2008
Accepted 25 April 2008
Available online 9 May 2008

JEL classification:
D10
D71
I19
J11
Z13

Keywords:
Obesity
Peer effects
Social networks

ABSTRACT

This note's aim is to investigate the sensitivity of Christakis and Fowler's claim [Christakis, N., Fowler, J., 2007. The spread of obesity in a large social network over 32 years. The New England Journal of Medicine 357, 370–379] that obesity has spread through social networks. It is well known in the economics literature that failure to include contextual effects can lead to spurious inference on "social network effects." We replicate the NEJM results using their specification and a complementary dataset. We find that point estimates of the "social network effect" are reduced and become statistically indistinguishable from zero once standard econometric techniques are implemented. We further note the presence of estimation bias resulting from use of an incorrectly specified dynamic model.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

The United States has experienced a startling increase in average weight and in obesity over the past few decades (Flegal et al., 2002; Hedley et al., 2004). Though this phenomenon is by now well known and has been widely discussed and debated, there is still little consensus on its causes. One proposed explanation for the increase in obesity is long run technological changes that have impacted food prices as well as the propensity to exercise (Philipson and Posner, 2003; Cutler et al., 2003). Though some observers include genetic variation as a potential explanation for the rise of obesity because of the large estimates of heritability of obesity (Stunkard et al., 1990; Coady et al., 2002), most researchers acknowledge that genetic explanations are unlikely to explain the rapid increase in obesity over a relatively short period of time.

☆ The views in this paper are solely those of the authors and do not reflect official positions of the Federal Reserve Bank of Boston or the Federal Reserve System.

☆☆ Further exchanges between the authors may appear on their home pages. Jason Fletcher http://www.med.yale.edu/eph/faculty/fletcher.html; Ethan Cohen-Cole http://www.bos.frb.org/economic/econbios/cohen-cole.htm; James Fowler http://jhfowler.ucsd.edu/; Nicholas Christakis http://christakis.med.harvard.edu/.

* Corresponding author. Tel.: +1 203 785 5760; fax: +1 203 785 6287.
E-mail address: jason.fletcher@yale.edu (J.M. Fletcher).
[1] Tel.: +1 617 973 3294; fax: +1 617 573 5494.

0167-6296/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhealeco.2008.04.005

One particularly interesting hypothesis recently explored by Nicholas Christakis and James Fowler (henceforth, CF) in the *New England Journal of Medicine* is that obesity may spread through "social networks effects."[2] In fact, CF report that their findings suggest that social networks indeed facilitate the spread of obesity.[3] This provocative finding was detailed in many media sources, including the front page of the *New York Times*.[4] *USA Today* coverage indicated that "Obesity is contagious" and "...pick your friends carefully..." (Hellmich, 2007).[5] CF suggest some potential mechanisms by which this may occur, including that having obese peers may change a person's tolerance for being obese or may influence weight-related behaviors such as eating habits, smoking, or exercise. Additional mechanisms suggested by CF include infectious causes of obesity or physiological imitation.

However, as is well known in the economics literature, there are alternative hypotheses that also potentially explain the empirical finding that friends' weight is correlated across time that do not require the presence of social network effects. As CF identify in their study, there are at least three reasons why the weight status of individuals could be clustered within reference groups.[6] The first is that individuals could choose their friends based on factors associated with weight or weight trajectories. In economics, this is typically referred to as selection (CF as homophily). Thus, friendship selection could directly lead to the correlation between friends' weight or weight gain without an individual's weight causally affecting his friend's weight through a social network effect. Second, individuals may adjust behavior because of exposure to common influences. These effects are typically referred to as *contextual* influences (CF as confounding). For example, the opening of a fast food restaurant, convenience store, gym, etc. near a school could simultaneously affect the weight of all friends in a school's social network. Importantly, the presence of (often unmeasured) shared surroundings can lead to erroneously implicating social network effects in individual outcomes where none exist.[7] Finally, individuals may alter their behavior as others in their group change theirs. Economists are now generally labeling this an *endogenous* social effect (CF as social network effects).

We point to three problems with the CF method. First, CF do not include a sufficiently broad set of contextual effects to account for a range of hypothesized causes of the epidemic. Second, the CF method of controlling for selection is much too narrow in scope. Third, the CF dynamic model as estimated produced coefficients with large degrees of bias (Liu et al., 2006).

Once the first two errors are corrected, evidence for endogenous causes of obesity is thin. We find that the CF results are not robust. In fact, the econometric evidence points strongly to shared environmental factors as the principle operative social mechanism underlying the positive correlation in weight status within reference groups.[8] We find this remarkable given the preponderance of contexts in which endogenous effects appear present and the fact that this class of empirical models appears to generate the appearance of effects quite easily (Krauth, 2006). Our findings point to the difficulty in labeling the source of social effects, particularly in contexts with a direct policy reference. The public health implications given endogenous versus contextual drivers of obesity are quite different.

2. Data

We use the Add Health dataset to examine whether there are social network effects in weight outcomes for a national sample of adolescents who transition into early adulthood.[9] Importantly, we have information on friends for approximately 5000 individuals, nearly 2000 of whom are followed over time along with at least one same-sex friend.[10] This sample size gives us nearly 4000 person-year observations, slightly larger than the 3000 used by CF. Summary statistics for our sample are presented in Table 1.[11]

Though there are several important differences between the Add Health and the Framingham Heart Study (FHS) used in CF, the two data sets are sufficiently similar to use to evaluate the role of transmission mechanisms. One important advantage of the Add Health is that it is a national sample of 7–12th graders in 1994/1995 instead of being confined to a smaller population with a wider age range. Another advantage is that individuals in our data are in a setting (high school) that is, in principle, more social than the varied lives of individuals in the Framingham study. Finally, by focusing on a national sample

[2] In the language of the Christakis and Fowler article, social network effects are equivalent to *endogenous* social effects (Manski, 1993). We use the terms interchangeably.

[3] They state, "...our observations suggest an important role for a process involving the induction and person-to-person spread of obesity" (p. 377), "The spread of obesity in social networks appears to be a factor in the obesity epidemic" (p.378), and "...obesity appears to spread through social ties" in the summary of their conclusions (p. 370).

[4] See for example, Kolata (2007) for the *New York Times*, Hellmich (2007) for *USA Today*, and Dembner (2007) for the *Boston Globe*.

[5] An official with the National Institute on Aging is quoted in Dembner (2007) as calling the study, "one of the most exciting studies in medical sociology ... in decades".

[6] Manski (1993) provided the seminal contribution in economics in describing the alternative mechanisms that can lead to correlated outcomes within reference groups.

[7] See Fletcher (2007) and Lee (2007) on the importance of unobserved group-level characteristics in estimating peer effects.

[8] While many empirical studies label contextual effects as observable group-level phenomena, our main result is to show that the social network effects estimated by CF can be accounted for by unobserved group-level controls. Of course, in Manski's nomenclature, these are also contextual effects.

[9] A full description of the sample design, data, and documentation is available at: http://www.cpc.unc.edu/addhealth.

[10] For individuals for whom more than one friend's BMI information is available, we select the friend with the highest nomination (1st–5th). Nearly two-thirds of the individuals in our sample are only matched to one friend's BMI due to the sample design. We select only one friend to be consistent with CF.

[11] Table 1A in the appendix of our working paper (Cohen-Cole and Fletcher, 2008) shows that our analysis sample has similar characteristics as the full Add Health sample.

Table 1
Summary statistics Add Health, analysis sample $N = 1988$; two observations per person

Variable	Wave	Mean	Std. Dev.
BMI	2	22.66	4.38
BMI	3	26.29	5.97
Friend BMI	2	22.50	4.10
Friend BMI	3	26.03	5.58
Obese	2	0.07	0.26
Obese	3	0.21	0.41
Friend obese	2	0.06	0.24
Friend obese	3	0.20	0.40
Male	All	0.44	0.50
Age	2	16.24	1.51
Age	3	21.70	1.50
Black	All	0.16	0.37
Hispanic	All	0.15	0.36
Maternal education[a]	All	13.34	2.34
Family income[a]	All	49.72	46.98
Missing family information	All	0.28	0.45

[a] Imputed missing data.

of individuals in an active social setting,[12] we are largely able to rule out that our findings may be due to the nature of our dataset. Thus, while one may believe that the finding of network effects in the FHS suggests an age or social setting difference in effects, it would be particularly difficult to make a generalized claim using the CF results in the absence of similar results from a national sample.

Similarities between the data sets include the time lags between interviews (approximately 3 years on average) and the type of information that was collected (see Table 4A in Cohen-Cole and Fletcher, 2008). It will be important for our baseline results to closely mirror those found by CF in order to be able to compare the results from our extended analyses. For brevity and to maximize the overlapping information contained in each dataset, we focus on one of the most interesting and robust results in CF—the positive relationship between the weight status of individuals and their same-sex friends.

3. Social network effects vs. shared experience

Central to our discussion is the distinction between *endogenous* effects, also labeled "induction" or social network effects by CF, and *contextual* effects. In the case of obesity, one can think of endogenous effects as describing the propensity to become obese because of the direct interaction with another individual. One may decide to eat more (or higher caloric foods) because their friend, spouse, neighbor does so. Because the two individuals are directly connected, they may influence each other for a variety of reasons. Contextual influences, on the other hand, reflect the shared surroundings of the members of a group that could lead to similar weight outcomes.[13] Without detailed information on an individual's characteristics, choices, preferences, and environment, it is difficult to discern whether two friends' simultaneous weight gain is attributable to their friendship or to an exposure of a common environmental factor (Manski, 1993; Durlauf, 2004).

4. Empirical methods

4.1. CF specification and replication

CF use data on obesity status for an individual (in their terminology, an "Ego") at a given point in time and estimate its relationship to the obesity status of a friend, spouse or relative (an "Alter") as well as its relationship to the Ego's age, gender, educational level, and past obesity status. The CF specification uses the BMI of an Ego (i) who lives in community (c) at time ($t+1$)[14] as a function of the individual's previous BMI, the individual's measured characteristics (X), Alter's (j) current and previous BMI, and an unobserved error term:

$$\mathrm{BMI}^{\mathrm{ego}}_{ic,t+1} = \theta_1 \mathrm{BMI}^{\mathrm{alter}}_{ic,t+1} + \theta_2 \mathrm{BMI}^{\mathrm{alter}}_{jc,t} + \theta_3 \mathrm{BMI}^{\mathrm{ego}}_{ic,t} + \beta X^{\mathrm{ego}}_{ic,t+1} + \varepsilon^{\mathrm{ego}}_{ic,t+1} \quad (1)$$

This type of model has three features that can, and regularly do, impact inference. First, if unmeasured community-level variables, which we will denote $c_{c,t+1}$, are positively correlated with individuals' BMIs, this will bias the estimate of the social network effect. Critically, the absence of a relevant contextual variable can lead to spurious inference on the endogenous

[12] Studies of peer effects often focus on school and college settings because of the frequent and repeated contact amongst individuals.
[13] Consider the case of two friends: each friend may simultaneously become obese due to the recent introduction of a high caloric restaurant near his place of residence. Though both soon have a higher BMI, this would not have been attributable to their friendship. The difficulty, of course, is distinguishing between the two potential mechanisms.
[14] Recall that in the Add Health data, there are three waves of surveys for which we have BMI data.

Table 2
Association between own-obesity status and friend's obesity status comparing results across studies and methods

Outcome: ego currently obese?			
Logit specification	CF	Baseline replication	School trends
Alter currently obese?	0.62***	0.588***	0.411**
	(0.24)	(0.160)	(0.162)
	[1.71]	[1.80]	[1.51]
Ego previously obese	4.38***	4.617***	4.634***
	(0.19)	(0.24)	(0.27)

Note: ***$p < 0.01$, **$p < 0.05$. Robust standard errors in parentheses. Odds ratios in brackets. CF results in Table S1 of web appendix, "same-sex friend" column. We do not include fixed effects in this table as results for short panel logistic specifications are known to be biased.

variable. CF appear to approach confounding by examining the type and direction of the friendship networks. For example, if individual A declared himself a friend of individual B but not vice-versa, then a social network effect should appear for A but not B. While network structures can be useful for identification of social network effects, their presence does not rule out the possibility that confounding environmental effects overlap and influence the decisions of network members.[15] Since CF never directly control for environmental factors, we view their results using directionality of friendship nominations as suggestive rather than conclusive.

Second, similarly incorrect inference can result if an individual's error term ($\varepsilon^{\text{ego}}_{ic,t+1}$) is correlated with his friend's BMI. For example, the (unmeasured) propensity to exercise or be involved in sports might increase the chances that two thin individuals are friends. Selection issues appear in this context. CF also provide an ad hoc solution for this issue. Their claim is that "the use of a lagged independent variable for an Alter's weight status controlled for homophily." Unless selection is conditioned *only* on this variable, this statement is spurious. As well, in the presence of social interactions, the use of lagged variables can lead to bias in estimation, a topic we turn to now.

Third, the presence of a lagged dependent (or independent) variable in a social interactions model can lead to substantial biases in estimation. Liu et al. (2006) find, using simulation evidence, that a model very similar to (1) above leads to significant bias in estimation relating to the dynamic role of social interactions.[16] Most importantly for our note is that misspecification of the model or error structure can lead to very large biases and thus incorrect inference. This essentially highlights the well known fact in this literature that inference errors are particularly hard to avoid. Inclusion of a lagged dependent variable appears to magnify the problem.[17]

Acknowledging these concerns, we take the CF model as is and replicate it using the Add Health data. Using Eq. (1) above, a logistic regression model as proposed by CF, we find the odds that an Ego becomes obese increase by 80% if his or her Alter is also obese. Broadly, this is in line with the CF finding of 71% for same-sex friends. Our basic results are presented in Table 2 as logit coefficients in order to compare with the table in CF.

4.2. Extension

We extend the CF model as follows. Assume the model specification now appears as

$$\text{BMI}^{\text{ego}}_{ic,t+1} = \theta_1 \text{BMI}^{\text{alter}}_{ic,t+1} + \theta_2 \text{BMI}^{\text{alter}}_{jc,t} + \theta_3 \text{BMI}^{\text{ego}}_{ic,t} + \beta X^{\text{ego}}_{ic,t+1} + c_{c,t+1} + \varepsilon^{\text{ego}}_{ic,t+1} \quad (2)$$

where we have added a set of environmental confounders ($c_{c,t+1}$).[18] This formalizes the notion of contextual effects. We use either a time-invariant measure as above or a time-dependent set of location-specific (in our case, schools) covariates. Econometrically, this can be expressed as using the following for the environmental confounders: $c_{c,t+1} = tc_{c,t+1}$, where t is a time variable. These represent a much richer set of controls to absorb average changes in social context experienced by all individuals in the sample. To explain further, these school-specific trends account for any environmental factors shared by individuals at the same school. CF control for year effects, but their specification does not capture any shared confounders that also vary across geographic space. For example, CF can control for the fact that the density of fast food restaurants has increased over time but not the fact that the number of fast food restaurants has grown faster in some areas than other areas. For example, suppose that the number of fast food restaurants has grown faster in Boston, Massachusetts than in western Massachusetts. Controlling for year effects (which controls for the growth in the number of fast food restaurants across the states in a given year) is not as appealing as controlling for the number of fast food restaurant in an individual's local area.

[15] The argument for identification of social network effects using network architecture has been formalized in Bramoullé et al. (2007). The methodology is intriguing and may be sufficient to control for confounding in the CF case, however, CF do not employ it, and it is not clear whether the single-Alter structure of most of their data permit identification in this setup in any case.

[16] The size and direction of bias depends on the type of variable, sample size, panel length and type of misspecification. However, bias appears in essentially every case. That is, correct inference on this type of model requires some type of bias correction.

[17] Of course, simply omitting a lagged dependent variable where it should belong does not alleviate the inference problem!

[18] As mentioned previously, we measure an individual's community by the high school they attend. This equation represents our capture of school-level confounders as a fixed effect.

Table 3
Association between own-body mass index and friend's body mass index comparing results across studies and methods

Outcome: ego current BMI				
OLS specification	CF	Baseline replication	School trends	Trends and FE
Alter current BMI	0.05**	0.054**	0.037	0.033
	(0.02)	(0.023)	(0.023)	(0.039)
CI	(.01, .09)	(.01, .10)	(−.01, .08)	(−.043, .109)

Note: $^{**}p < 0.05$. Robust standard errors in parentheses. Baseline replication are results that are identical in specification to CF with the exception of data used. School Trends column updates CF specification by adding school-specific time trends. Trends and FE columns includes both school-specific trends and individual level fixed effects.

Without accounting for the trends, clustering of obesity in social networks that changed over time would incorrectly be absorbed in estimation by the endogenous variable. Though one, in principle, would want many more controls to account for additional contextual effects, we will note shortly that the endogenous effect vanishes even with this relatively simple characterization.

We add to our evaluation by accounting for self-selection of friends (homophily). This is accomplished by looking only at the change in BMI from the time of declaration of friendship until the subsequent weight measurement. Note the distinction between this method and the lagged independent variable used in CF. Our method allows us to distinguish between the desire to become friends based on similarity in weight, which would appear based on the simultaneous measurement of friendship and weight, and the friendship effect of weight gain.

We find that inclusion of a more complete set of controls for school-level environmental confounders leads to a large drop in the coefficient of interest. The odds fall to 50% in column 3 of Table 2 (logit coefficient reported). As well, Liu et al. (2006) report that an increase in the number of contextual variables reduces bias in most cases.

Since linear specifications more easily accommodate fixed effects and time trends, in Table 3 we estimate models predicting BMI. In our data the association between own-BMI and friend-BMI is 0.05, which is identical to the CF results.[19] After controlling for environmental confounding (through school-specific time trends), our estimate falls by over 30% to 0.037. Finally, we control for friendship selection by controlling for individual fixed effects.[20] With this inclusion, the coefficient of interest declines further and becomes statistically indistinguishable from zero. These results, using similar sample sizes and data as CF, suggest that the "social network effects" implicated in CF are not robust to standard economic critiques found in the peer effects literature. In fact, our results suggest that omitted group-level characteristics are the most likely mechanism responsible for correlated body weight within peer groups.[21]

5. Conclusion

Our evaluation suggests that the spread of obesity is related to the environment in which individuals live. Though we do not completely rule out the possibility of induction and person-to-person spread of obesity, our results suggest that shared environmental factors can cause the appearance of social network effects. While comparing results across datasets that are quite different in design and focus is usually fraught with difficultly, we were encouraged to be able to closely replicate findings of CF using similar methods; this provides suggestive evidence that the structure of obesity decisions is similar across the two samples. It is only when we incorporate more extensive controls than those found in CF that our results diverge. This divergence in results suggests caution in interpreting correlations in linked individuals' BMI as social network effects. In fact, we find evidence that community-level factors are able to explain a large share of the "social network effect" in our data.

We find results that suggest difficulty in distinguishing social network effects and environmental confounds of weight gain. And though we advise caution in interpreting the available evidence of a social contagion in weight, we concur with CF on the use of network phenomena, broadly writ, to help to ameliorate the epidemic. Some of the encouraging evidence mentioned in CF, in particular the fact that alcohol and smoking cessation programs are more effective when coupled with peer support, can be exploited independent of the degree of *induction* present. That is, one need not find evidence that obesity is spread via networks to draw on the logic of peer support to address important issues in public health such as smoking and obesity.[22] Our data show tightly interconnected networks of friends at the high school level and a high degree of coincidence of obesity within these networks; this can be exploited for interventions even if the *cause* of the obesity came from a joint

[19] OLS results are discussed in the CF article but not presented. The authors provided supplementary material to us upon request.

[20] In a supplemental appendix available upon request, we also show that our results here are robust to removing the lagged individual obesity status. Including the lagged individual status is akin to assuming that individuals follow an autoregressive process in obesity. Not only is this very difficult to verify, its incorporation makes inference on the object of interest (friend's influence) very difficult.

[21] The Liu et al. (2006) simulations also find a *decrease* in estimation error for an increase in the number of contextual effects suggesting that while our extensions may have bias, the finding that the endogenous effects is not significant is robust.

[22] In fact, none of the cited articles in CF illustrate a link between success in cessation programs and induction. This is a very interesting question and one that we encourage as a topic of future research. Until that point, the key relationship is simply the presence of a peer group for support—which can and does occur independent of the obesity transfer mechanism.

external source. Similar arguments can be made to support the notion that obesity can be approached from a public-health perspective rather than a clinical one.

Acknowledgements

The authors thank Elizabeth Bradley, Paul Cleary, John Mullahy, David Paltiel, and Jody Sindelar for very helpful comments and Jonathan Morse for research assistance. This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu).

References

Bramoullé, Y., Djebbari, H., Fortin, B., 2007. Identification of Peer Effects through Social Networks. University of Laval Working Paper.
Christakis, N., Fowler, J., 2007. The spread of obesity in a large social network over 32 years. The New England Journal of Medicine 357, 370–379.
Coady, S.A., Jaquish, C.E., Fabsitz, R.R., Larson, M.G., Cupples, L.A., Myers, R.H., 2002. Genetic variability of adult body mass index: a longitudinal assessment in Framingham families. Obesity 10, 675–681.
Cohen-Cole, Ethan, Fletcher, J.M., 2008. Is Obesity Contagious? Social Networks vs. Environmental Factors in the Obesity Epidemic. SSRN Working Paper.
Cutler, D.M., Glaeser, E.L., Shapiro, J.M., 2003. Why have Americans become more obese? Journal of Economic Perspectives 17-3, 93–118.
Dembner A., 2007. Obesity spreads to friends, study concludes. Boston Globe; July 26.
Durlauf, S.N., 2004. Neighborhood effects. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics, Vol. 4, 1 ed. Elsevier, Amsterdam, pp. 2173–2242.
Flegal, K.M., Carroll, M.D., Ogden, C.L., Johnson, C.L., 2002. Prevalence and trends in obesity among US adults, 1999–2000. Journal of the American Medical Association 288, 1723–1727.
Fletcher, J.M., 2007. Social Interactions and Smoking: An IV/FE Approach. Yale University Working Paper.
Hedley, A.A., Ogden, C.L., Johnson, C.L., Carroll, M.D., Curtin, L.R., Flegal, K.M., 2004. Prevalence of overweight and obesity among US children, adolescents, and adults, 1999–2002. Journal of the American Medical Association 291, 2847–2850.
Hellmich, N., 2007. Obesity Can Run in Social Circles. USA Today; July 26.
Kolata, G., 2007. Find Yourself Packing It On? Blame Friends. New York Times; July 26.
Krauth, B., 2006. Simulation based evidence of peer effects. Journal of Econometrics 133, 243–271.
Lee, L.F., 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. Journal of Econometrics 140, 333–374.
Liu, X., Lee, L.F., Kagel, J., 2006. Dynamic Discrete Choice Models with Lagged Social Interactions: With an Application to a Signaling Game Experiment. Ohio State; Mimeo.
Manski, C.F., 1993. Identification of endogenous social effects: the reflection problem. Review of Economic Studies 60-3, 531–542.
Philipson, T., Posner, R., 2003. The long-run growth in obesity as a function of technological change. Perspectives in Biology and Medicine 46-3 (Supplement), S87–S107.
Stunkard, A.J., Harris, J.R., Pedersen, N.L., McClearn, G.E., 1990. The body-mass index of twins who have been reared apart. New England Journal of Medicine 322 (21), 1483–1487.