Identifying Endogenous Peer Effects in the Spread of Obesity

Timothy J. Halliday[1]
Sally Kwak[2]
University of Hawaii- Manoa

October 2007

Abstract

Recent research in the New England Journal of Medicine (NEJM) purports to show the existence of peer effects in the spread of obesity. Using a dataset of 5124 residents from Framingham, Massachusetts spanning the years 1971 to 2003, the authors show correlations between own weight gain and friends' and relatives' weight gain over this period. They find, furthermore, that these results are strongest for males and weaker for females. We use the Adolescent Health Survey, a nationally representative dataset of seventh through twelfth graders in 1994 and 1996 to examine the effect of peers on weight gain. Despite the differences in the samples, we are able to replicate the pattern of results in the NEJM study. However the results are not robust to alternative definitions of the outcome variable. Furthermore, due to the various identification issues that are unresolved in both this and the NEJM paper, we conclude that the evidence for contagion effects in the spread of obesity is only suggestive at best.

Keywords: Peer effects, obesity, adolescent health
JEL: I10, I12, J01

[1] halliday@hawaii.edu
[2] kwaks@hawaii.edu

I. Introduction

A study released in the New England Journal of Medicine (NEJM) in July 2007 has many reporters, bloggers and researchers asking "Do your friends make you fat?"[3] As overweight and obesity rates climb and as medical researchers draw links between weight control and conditions such as diabetes, heart disease and others, the question takes on significance for public policy.

Using data from the Framingham Heart Study, researchers in the NEJM study examine 12,067 adults from 1971 to 2003 and study the correlations between individual weight gain and friends', neighbors', and family members' rates of weight gain. They find that a person's likelihood of becoming obese, defined as having body-mass index over 30, increases by 57 percent over 32 years if he or she has a friend who becomes obese over this same time period. This effect increases to 171 percent if the friendship is mutual. Furthermore, the effects of male peers on males are estimated to be 71 percent, while the effects of female peers on females are statistically insignificant. In this paper, we examine whether friends make friends fat in a national sample of adolescents.

Though our data covers changes over a relatively short two-year period, when we employ the outcome variable used by the NEJM study, we are able to replicate the results of the Framingham adults in our sample of adolescents. In keeping with the findings of the NEJM study, we find that peer groups are indeed correlated with own propensity to become fat. We find, too, that effects are strong for males but not statistically significant for females. However, we also find that these results are not robust to different specifications of outcome variables. Moreover, we do not find any evidence of a statistically significant relationship between own and peer weight gain once we employ instrumental variables estimation that helps to address the simultaneous

[3] See for example, New York Times 7/25/2007; http://blogs.ebay.com/allura-2005/entry/Can-Your-Friends-Make-You-Fat/_W0QQidZ299720017; slate.com 7/26/2007.

determination of weight gain between individuals and their peers. Our findings recommend caution in interpreting the NEJM estimates as evidence of endogenous peer effects in the spread of obesity.

In what follows, we will describe the data, our research methods and their relationship to those in the NEJM study and present our findings. The final section concludes.

II. Data

We use data from the National Longitudinal Adolescent Health Survey (Add Health). The Add Health survey was conducted by the Carolina Population Center and is available for a nationally representative sample of students who were in seventh through twelfth grades in 1994. Wave I, which was fielded in 1994-1995, consists of an In-School questionnaire that was filled out by 90,118 students in 145 schools in 80 communities. A subset of 20,745 students was then chosen for an in-depth In-Home survey. Wave II, which was fielded in 1996, includes an In-Home questionnaire that was completed by 14,738 students who were a subset of the original 20,745 pupils. The 145 schools in the Wave I survey consist of pairs of sister schools. That is, if a particular high school was included in the survey, the corresponding feeder junior high or middle school was also included. If a school spanning seventh through twelfth grades was chosen for the survey, no sister school was included. We use data from the two waves of the In-Home survey.

Students who were selected for the In-Home survey were asked for information on height and weight in Wave I and again in Wave II. Using this information, we construct a Body Mass Index (BMI) variable according to the formula: $\text{BMI} = 703 * \frac{\text{weight}}{\text{height}^2}$. Using BMI, we construct three outcome variables. First, "Weight Gain" is the level gain in BMI between Wave I (1994) and Wave II (1996). Second, "Percent Weight Gain" is the percentage change in BMI between Wave I and Wave II. Finally, "Change in Overweight" is the change between Wave I and Wave II in the

dummy variable "Overweight" which is defined to be one if BMI is greater than 25 and zero otherwise.[4]

Peer Networks

The Add Health survey is well-suited to our purposes because of the extensive data on friendship networks. In each of the surveys, students are asked to nominate five female friends and five male friends. In almost all cases, students report fewer than five male and five female friends indicating that they are not constrained in their choice of friends in their network by the ten-friend limit. These friend nominations include both friends in the same school as well as friends from outside of school. Because we do not have information on friends outside of the respondent's school, we are unable to include them in our measure of average peer group weight gain. However, the vast majority of friend nominations (approximately 85%) were to other students in the same school. There are a sizeable number of nominations to friends that were not found on the school rosters. This may be due to nicknames not matching official names, students who are new to the school, or errors in the school records. Excluding friends who are outside the respondent's school and friends who could not be located, we have a total of 1.27 friend nominations per pupil (with a standard deviation of 1.67).

Table I presents summary statistics. The average age in our sample is 15 and the average grade is between nine and ten. About half the sample is male and 54% of the sample is white. Over the two-year period covered in the data, students saw an average increase in BMI of .51 which translates to about a three percent increase. Approximately 22% of the sample had a BMI of greater than 25 in the base year and was thus considered overweight. An additional 3% of the sample became overweight over the study period. These statistics vary somewhat by gender: Boys were

[4] This definition of "Overweight" is the standard, accepted use of BMI.

somewhat more likely to be overweight in the base year and gained more weight over the study period.

III. Research Methods

The canonical structural model of endogenous peer effects is given by:

$$y = \alpha + \beta E[y \mid x] + E[w \mid x]'\gamma + w'\lambda + \varepsilon, \qquad (1)$$

where y is the outcome of interest (for example, obesity), x is a vector of group characteristics, w is a vector of individual characteristics and ε is an error term. $E[y \mid x]$ is the average behavior of the peer group and $E[w \mid x]$ is a vector of average characteristics of the peer group. Unfortunately, the effort to estimate peer effects has been stymied by a number of empirical issues. These issues are summarized in Manski (1995).

Following Manski, we assume that $E[\varepsilon \mid x, w] = x'\delta$. Then equation (1) can be rewritten as follows:

$$E[y \mid x, w] = \alpha + \beta E[y \mid x] + E[w \mid x]'\gamma + w'\lambda + x'\delta \qquad (2)$$

Equation (2) shows several ways in which own behavior and peer behavior are related. First, using Manski's terminology, if $\gamma \neq 0$ then the model shows contextual effects. These arise when "the propensity of an individual to behave in some way varies with the distribution of background characteristics in the group." Second, if $\delta \neq 0$ the model shows correlated effects. These describe "the propensity of individuals in the same group to behave similarly because they face similar institutional environments or have similar individual characteristics". In the NEJM study, the author's term for what we call correlated effects is "homophily."

A priori, there are many reasons to believe that correlated effects will be important in any obesity study. For example, in our context, students in the same schools may have similar

propensities to be overweight because they are served the same school lunches or because they can purchase candy, potato chips and carbonated beverages from the same vending machines. In addition, correlated effects would also be present if individuals with different propensities to gain or lose weight associated with one another. This might occur if either the wrestlers (who face similar pressure to maintain weight) or the Latin Club members (who presumably do not face these pressures) associated with each other.

Finally, if $\beta \neq 0$ then the model shows endogenous peer effects. These occur when weight gain or loss among a pupil's peers *causes* the pupil to behave in a similar fashion. This might occur if weight loss (gain) among a pupil's peers alters the pupil's perception of what an acceptable weight is and, thus creates incentives for the pupil to also lose (gain) weight. In such a scenario, $E[y|x]$ is an equilibrium outcome and is therefore endogenous. This phenomenon is called the "reflection problem." Both our study and the NEJM study are primarily interested in the identification of β. In the NEJM study, the author's term for what we call endogenous peer effects is "induction."

Empirical Approach

Our empirical counterpart to equation (1) relates own BMI to peer-group average BMI and is given by:

$$y_{ist} = \beta \cdot \bar{y}_{ist} + \delta_s + \alpha_i + \varepsilon_{ist}, \qquad (3)$$

where y_{ist} is one of three constructed outcome variables as described in section II: BMI, relative BMI (where BMI in 1994 is normalized to 1), or a dummy indicating whether BMI is greater than 25 or not. The parameters δ_s and α_i are school-level and individual-level fixed effects, respectively and ε_{ist} is a time-variant unobserved component to individual behavior. The subscript *t* indexes one of two survey years, 1994 or 1996. Finally, $\bar{y}_{ist}$ is the peer-group average of BMI,

relative BMI, or a dummy for BMI greater than 25. The averages are taken over all of the peers in the pupil's friend nominations.

The elements of equation (3) correspond to the contextual, correlated and endogenous effects in the following way. First, the inclusion of average peer obesity captures the endogenous effects. Because own and peer obesity are jointly determined, OLS will not consistently identify β. Second, the inclusion of school and individual fixed effects captures the correlated effects. Failure to address these in the estimation will bias our estimate of β since we expect the individual and school fixed effects to be correlated with both peer and own behavior. Third, we assume that average peer background characteristics do not directly affect own weight gain. Thus, contextual effects are omitted from equation (3). This assumption is the exclusion restriction that enables us to identify instrumental variables estimates of β. Gaviria and Raphael (2001) make a similar assumption.

To address the correlated effects in equation (3), we difference the data and are left with the following specification:

$$\Delta y_{is} = \Delta \bar{y}_{is}\beta + \Delta \varepsilon_{is} \qquad (4)$$

where Δy_{is} denotes change in the outcome between Wave I and Wave II survey years, $\Delta \bar{y}_{is}$ denotes change in average outcome of the individual's peer-group between survey years, and $\Delta \varepsilon_{is}$ is the change in the unobserved component to individual behavior between survey years. Since an individual's level of obesity is partly a function of genetic endowment and other factors which cannot be controlled by the individual, it is important to account for these in estimating the endogenous effects. First-differencing rids our estimates of these and any other time-invariant individual or school fixed effects. However, OLS estimation of equation (4) will still be biased by the reflection problem.

We also estimate a slight modification of equation (4):

$$\Delta y_{is} = \Delta \bar{y}_{is}\beta + x_i\lambda + bmi_i\delta + \Delta\varepsilon_{is}, \qquad (5)$$

where bmi_i denotes individual i's BMI in the base year and x_i denotes a vector of individual i's observable characteristics measured in the base year. Controlling for baseline BMI and individual characteristics addresses any possible omitted factors that may also be correlated with weight fluctuations. For example, certain ethnicities may be more or less prone to weight gain.

Finally, we estimate equations (4) and (5) by instrumental variables methods which enables us to address both the correlated effects and the reflection problem under suitable exclusion restrictions. We use average education of the father and mother in the peer group as an instrument for the change in obesity inside the peer network. Our measure of parental education is a dummy variable for having obtained a college degree. This instrument will be valid if there are no contextual effects and if average parental education in the network is not correlated with any omitted time-varying correlated effects.

IV. Contrasting Our Approach with the Existing Literature

We contrast our paper with the NEJM study. In the NEJM study, "the basic statistical analysis involved the specification of longitudinal logistic-regression models in which the ego's obesity status at any given examination or time point (t+1) was a function of various attributes, such as the ego's age, sex, and educational level; the ego's obesity status at the previous time point (t); and most pertinent, the alter's obesity status at times t and t+1." Note that the NEJM study refers to own obesity as the ego's obesity and refers to the peer's obesity as the alter's obesity. The authors of the study claim that the use of lagged own (or ego) obesity addresses "genetic endowments" as well as any other predispositions towards obesity and that the use of lagged peer (or alter) obesity

addresses "homophily." They further claim that in this specification, the coefficient on peer (or alter) obesity in the contemporaneous period can be interpreted as an endogenous peer effect.

There are two problems with this assertion. First, given the lack of a formal estimation equation, it is not clear how this empirical strategy addresses unobserved individual-level heterogeneity. Their discussion suggests that they should estimate a rich dynamic panel data model with a rigorous treatment of unobserved heterogeneity as in Honoré and Kyriazidou (2000). In the absence of such modeling there is no clear reason why any correlation between unobserved heterogeneity and contemporaneous alter obesity status is eliminated by the inclusion of lagged ego and alter obesity levels. Moreover, because the estimation is in levels and not differences, we expect these biases to be quite large. Second, even if the NEJM study did adequately address bias due to correlated effects or "homophily," it does not address the reflection problem.

V. Results

Before we discuss our empirical findings, it is important to point out that the sample sizes in our estimations will be substantially smaller than the 14,738 pupils who are present in both waves of the In-Home survey. There are several reasons for this. First, there is a lot of missing data. This is especially true in the case of BMI since the construction of this variable requires two variables, height and weight. Second, our estimations require that at least one friend could be found in both waves of the survey. Third, our estimations require that we have BMI information for at least one of the friends in the peer network. Finally, because we are working in first-differences, our estimations require that the all other explanatory data for the individual are present in both waves of the survey.

Table 2 displays estimates of equations (3) and (4) on the pooled sample of males and females. In columns 1, 4, and 7, we see the simple correlations between changes in own weight and

changes in peer-group weight. The estimates are not significantly different from zero when the outcome variable is defined as changes in level of BMI or percent change in BMI. However, the relationship is significant when the outcome variable is defined as in the NEJM study, a binary variable for whether the individual is overweight or not (defined as BMI>25). In Columns 2, 5, and 8, we include baseline BMI and we see a similar pattern, namely, that the level change and percent change variables show no significant relationship between peer-group outcomes and individual outcomes, but the binary variable does. Finally, in columns 3, 6 and 9, we include additional controls and still see that the estimate on changes in peer obesity is only significant when the outcome is binary.

Table 3 presents these same results, restricting the sample to boys. We find, as the authors of the NEJM study find, that the results are quite strong when the outcome variable is defined as a binary variable for whether the subject is overweight or not. However, once again these results are not robust to other definitions of weight change. The coefficient on baseline BMI is strongly significant in all six columns indicating the strong effect of an individual's ingrained habits, genetic endowment and biology on weight gain. The significant race variables may be a result of varying racial propensities to gain weight, but may also indicate that adolescents choose their peer groups and that weight gain and loss is endogenous to the peer group, not determined by it.

Table 4 presents results for the girls in the sample. We find, as the authors in the NEJM study do, gender differences in peer group effects. The peer-group estimates in columns 3, 6, and 9 are no longer significant as they were in the larger sample and in the sample of boys. And as in Tables 2 and 3, the peer-group estimates for the BMI level and BMI percent change variables are not significant.

We note that there are two opposing sources of bias operating in the estimations of Tables 2 through 4. First, because we estimate the models in these tables using OLS, these estimates will be

upwards biased by the reflection problem. So the true value of β should be even smaller than our estimates suggest. Second, there may be a large degree of misreporting of BMI which will create errors in our measures of average peer obesity. Provided that these are classical measurement errors, this should bias our estimates towards zero. Moreover, as pointed out by Deaton (1995), because we estimate the model in first-differences, any attenuation bias from measurement error will be exacerbated.

Finally, Table 5 presents results from the instrumental variables regressions which use average parental education in the peer group as an instrument for changes in peer obesity. The table provides no evidence of any relationship between own and peer obesity. However, we concede that many of the estimates in this table are very imprecise and that the poor performance of the instrumental variables estimation may be a result of weak instruments.

VI. Conclusion

We conclude that many of the key findings of the recent NEJM article on contagion in obesity can be replicated in our nationally representative sample of adolescents. However, we present two caveats to these results. First, these findings are not robust to the specification of the dependent variable. Second and more importantly, due to the plethora of identification issues that go unresolved in our paper and especially in the recent NEJM paper, we do not believe that the results in either paper can be reasonably construed as conclusive evidence of contagion in obesity. Both papers are only suggestive at best.

References

Deaton, Angus (1995): "Data and Econometric Tools for Development Analysis," in *Handbook of Development Economics*, vol. 3A, ed. Jere Behrman and T.N. Srinivasan. Amsterdam: North Holland.

Christakis, Nicholas A. and James H. Fowler, (2007): "The Spread of Obesity in a Large Social Network over 32 Years," *New England Journal of Medicine* 357: 370-9.

Gaviria, Alejandro and Steven Raphael (2001): "School-Based Peer Effects and Juvenile Behavior," *Review of Economics and Statistics*, 83: 257-268.

Honoré, Bo and Ekaterina Kyriaidou (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, 68: 839-874.

Manski, Charles F., (1995): *Identification Problems in the Social Sciences*. Cambridge, MA: Harvard University Press.

Table 1. Summary Statistics

Background Variables		
Age	15.05	N=14,910
	(1.71)	
Male	.49	N=14,855
	(.50)	
Grade	9.66	N=14,826
	(1.62)	
White	.54	N=14,957
	(.50)	
Black	.23	N=14,957
	(.42)	
Outcome Variables-- All		
Increase in BMI	.51	N=13,687
	(2.17)	
% Increase in BMI	.03	N=13,687
	(.09)	
Overweight	.22	N=19,667
in Base Year	(.42)	
Change in	.03	N=13,687
Overweight	(.30)	
Outcome Variables-- Boys		
Increase in BMI	.57	N=4815
	(2.01)	
% Increase in BMI	.03	N=4815
	(.09)	
Overweight	.24	N=7125
in Base Year	(.43)	
Change in	.03	N=4815
Overweight	(.30)	
Outcome Variables-- Girls		
Increase in BMI	.41	N=5093
	(2.14)	
% Increase in BMI	.02	N=5093
	(.09)	
Overweight	.20	N=7367
in Base Year	(.40)	
Change in	.03	N=5093
Overweight	(.28)	

Notes: Standard deviations in parentheses.

Table 2. Correlations between individual outcome and peer-group outcome

	Own weight gain			Own percent weight gain			Own change in overweight		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Weight gain of peers	.02	.02	.01						
	(.02)	(.02)	(.04)						
Percent weight gain of peers				.02	.02	.01			
				(.02)	(.02)	(.02)			
Change in overweight of peers							.04 *	.04 *	.07 ***
							(.02)	(.02)	(.03)
BMI in base year		.07 ***	.14 ***		-.00	.00		.01 ***	.01 ***
		(.01)	(.02)		(-1.00)	(.00)		(.00)	(.00)
Age			.01			-.00			-.00
			(.10)			(.00)			(.01)
Male			-.35 **			.01 *			.02
			(.15)			(.00)			(.01)
Grade			-.14			-.01 **			.00
			(.10)			(.00)			(.01)
White			.02			-.01 *			-.01
			(.17)			(.00)			(.01)
Black			.75 ***			.01			.04 **
			(.29)			(.00)			(.02)
N	2708	2708	1892	2708	2708	1892	2708	2708	1892

Notes: Standard errors in parentheses.

Table 3. Correlations between male outcomes and peer-group outcomes

	Own weight gain			Own percent weight gain			Own change in overweight		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Weight gain of peers	.04	.04	.03						
	(.03)	(.03)	(.03)						
Percent weight gain of peers				.04	.03	.02			
				(.03)	(.03)	(.03)			
Change in overweight of peers							.14 ***	.14 ***	.14 ***
							(.03)	(.03)	(.03)
BMI in base year		-.04 ***	-.03 ***		-.00 ***	-.00 ***		-.00 *	-.00 **
		(.01)	(.01)		(.00)	(.00)		(.00)	(.00)
Age			-.01			-.00			.01
			(.08)			(.00)			(.01)
Grade			-.10			-.00			-.00
			(.08)			(.00)			(.01)
White			-.36 ***			-.02 **			-.03
			(.12)			(.01)			(.02)
Black			.01			.00			.05 **
			(.16)			(.01)			(.02)
N	1384	1384	1375	1384	1384	1375	1384	1384	1375

Notes: Standard errors in parentheses.

Table 4. Correlations between female outcomes and peer-group outcomes

	Own weight gain			Own percent weight gain			Own change in overweight		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Weight gain of peers	.01	.01	-.00						
	(.03)	(.03)	(.03)						
Percent weight gain of peers				.04	.04	.02			
				(.03)	(.03)	(.03)			
Change in overweight of peers							.01	.01	.01
							(.03)	(.03)	(.03)
BMI in base year		-.06 ***	-.05 ***		-.00 ***	-.00 ***		-.00 **	-.01
		(.01)	(.01)		(.00)	(.00)		(.00)	(.01)
Age			-.04			-.00			.01
			(.09)			(.00)			(.01)
Grade			-.04			-.00			-.00
			(.09)			(.00)			(.02)
White			.16			.00			.02
			(.13)			(.01)			(.02)
Black			.39 **			.01 **			-.00
			(.16)			(.01)			(.00)
N	1500	1500	1496	1500	1500	1496	1500	1500	1496

Notes: Standard errors in parentheses

Table 5. Instrumental variables estimates of peer effects

	Own weight gain			Own percent weight gain			Own change in overweight		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Weight gain of peers	1.71	1.73	-1.44						
	(2.76)	(8.93)	(1.34)						
Percent weight gain of peers				.74	-2.82	-.99			
				(1.76)	(11.33)	(.96)			
Change in overweight of peers							-1.11	-3.06	.76
							(1.78)	(6.57)	(1.53)
BMI in base year	.08 **	-.01	-.02	.00	-.00	-.00	.00	-.01	-.00
	(.04)	(.10)	(.04)	(.00)	(.00)	(.00)	(.01)	(.01)	(.00)
Age	.29	.33	-.27	.00	-.04	-.01	.01	.09	-.02
	(.71)	(2.40)	(.27)	(.02)	(.12)	(.01)	(.03)	(.15)	(.03)
Grade	-.14	-.19	.08	-.00	.01	.01	-.02	-.09	.02
	(.28)	(1.17)	(.20)	(.01)	(.04)	(.01)	(.03)	(.18)	(.03)
White	-.01	-.29	.25	-.00	.01	.01	-.03	-.02	.02
	(.24)	(1.08)	(.27)	(.01)	(.06)	(.01)	(.07)	(.08)	(.02)
Black	.64	.99	.42	.02	-.05	.01	-.01	-.13	.05
	(.92)	(4.11)	(.48)	(.02)	(.23)	(.02)	(.08)	(.40)	(.04)
Boys only	No	Yes	No	No	Yes	No	No	Yes	No
Girls only	No	No	Yes	No	No	Yes	No	No	Yes
N	724	489	596	724	489	596	724	489	596

Notes: Standard errors in parentheses.