Obesity: transition from adolescence to adulthood and feedback partial gmm logistic model with time-dependent covariates

Di Fang [1], Kyle M. Irimata [2], Rachael N. Rhodes [3], Jeffrey R. Wilson [4]

(1) Di Fang PhD AgriBusiness, Department of Agricultural Economics and Agribusiness University of Arkansas
(2) PhD Statistics, School of Mathematical and Statistical Sciences, Arizona State University
(3) MS Statistics, School of Mathematical and Statistical Sciences, Arizona State University
(4) PhD Statistics, CPCOM 465D, Department of Economics, Arizona State University Tempe, AZ 85287

CORRESPONDING AUTHOR: *Jeffrey R. Wilson PhD Statistics, CPCOM 465D, Department of Economics, Arizona State University Tempe, AZ 85287, Email: jeffrey.wilson@asu.edu*

DOI: ***10.2427/13022***
Accepted on January 16, 2019

ABSTRACT

Background: The aim of this study is to investigate the impact of certain covariates on obesity. More importantly, we seek to determine the feedback of obesity on depression, and physical activity as they transition from adolescence to young adulthood.
Methods: Using 15 years of nationally representative data from 6560 adolescents (Add health data), we estimate feedback and associations between depression, and activity scale on obesity while we adjusted for gender, age, race/ethnicity, socioeconomic status through a GMM logistic regression model with time-dependent covariates.
Results: Activity ($p<0.001$) and depression ($p<0.001$) have significant impact on Obesity. In early years, alcohol had no impact ($p=0.895$ and $p=0.476$) on obesity but in later years it did ($p<0.001$). In the early years, television hours had an impact but as they got older, it did not.
Conclusion: Our findings suggest that public health researchers can target obesity simultaneously with depression, and activity scale. These findings contribute new insights into the feedback of obesity on depression, and activity. This unique model allows segments of associations to be addressed rather than assuming all associations remain the same over 15 years.

Key words: Obesity, depression, activity scale, time-dependent covariates

INTRODUCTION

The prevalence of obesity is still high in the USA. Thus, obesity and childhood obesity, in particular, are the focus of many public health efforts in the United States, with one-third of adults and 17% of child obese (Let's move 2014, NCCOR2014). Obesity is associated with several risk factors for later heart disease and other chronic diseases including hyperlipidemia, hyperinsulinemia, hypertension, and early atherosclerosis (WHO 1998). Yet, little is known about the feedback of obesity on depression, and on physical activity. Studies showed that obesity during adolescence carries with it important consequences into adulthood [1,2,3]. In an effort to identify the impact of childhood obesity on adulthood ou,tcomes, researchers have recently focused on health disparities early in life and on the evolution of these gradients as age increases [1,4]. Although critical periods appear to exist for the onset of obesity in childhood, the relative contribution of obesity that begins in the prenatal period, the period of adiposity rebound or in adolescence to the prevalence of adult obesity and its associated complications remains unclear.

In this paper, we study how certain factors are associated with childhood obesity and their progression into adulthood through the National Longitudinal Study of Adolescent to Adult Health (Add Health) database. The analysis was done with a unique statistical model for feedback and related changes in association. In addition, we looked at the impact of these factors on adult obesity, as well as the impact of childhood obesity outcomes on the focal factors in adulthood.

METHODS

Existing models

We are unaware of any studies that have examined the long-term effects of obesity, though several studies in various disciplines have examined the short-run consequences of obesity. Typically, these studies estimated cross- sectional associations between obesity and a variety of short-term, contemporaneous health and social outcomes, such as general health status, perceived wellbeing, physical activity, emotional problems, life satisfaction, and behavioral problems [1,2,3,6,7]. Even though previous literature is suggestive of negative consequences of obesity on health and socioeconomic outcomes, several limitations hinder statistical inference. Prominent among these are, error in the measurement of obesity conditions, a focus on short-run outcomes, and the likelihood of environmental or family-level confounding in the estimated relationships.

Stockwell [6] showed that the beneficial effects of childhood advantage that translates into future outcomes highlight the importance of educational attainment and race/ethnicity. Similar research found the impact of physical activity on obesity to be conditional on age and gender. Physical activity, varying in its intensity, benefits overweight children of age 9 to 12 when it is moderate; and reduces abdominal fat in male adolescents when it's vigorous [2,3,8]. On the other hand, the effects of behavioral characteristics such as smoking and watching TV on obesity status are not as clear. For example, [9] identified an increased obesity rate with the cigarettes price, whereas [5] found no evidence to support this conclusion. Finally, depression has been found to have a prolonged effect on obesity from adolescence to adulthood with a heavier influence on teenage girls, [10].

In this study, we overcome many of these limitations by using a longitudinal dataset with information on the same individual at multiple points in time to explore the longer run. One major advantage of our longitudinal study is its capacity to separate change over time within student and differences among students (cohort effects). However, when dealing with longitudinal data not only does the response variable change over time, but the predictors or covariates can also change over time. Such covariates are referred to as time-dependent covariates. Thus, the treatment of time-dependent-covariates in the analysis of longitudinal data necessitates strong statistical inferences about dynamic relationships and provides more efficient estimators than can be obtained using cross-sectional data [11].

The generalized linear models (GLMs) are inappropriate for analyzing longitudinal data due to the clustering, thereby violating the necessary independence assumption for the observations. An appropriate model should be capable of addressing the clustering. The presence of clustering when fitting marginal regression models with time-dependent covariates has shown to be effectively modeled through the use of generalized methods of moments (GMM) [5,11,12,13,14].

The fit of binary marginal models to data with time-independent covariates is well established, [15,16,17]. However, the application of such models to data with time-dependent covariates is still developing. While it is common to fit generalized estimating equation (GEE) models with certain working correlation matrix when confronted with repeated measures data, doing so with time–dependent covariate is not the best approach. In particular, when time-dependent covariates are present, [18,19] noted that the consistency of estimators is not assured with arbitrary working correlation structures and suggested fitting generalized estimating equations (GEE) with independent working correlation matrix. However, such a model-fit approach does not use all the information available as they omit certain moment conditions that should be included in estimating the regression coefficients. Lalonde, Wilson and Yin [11] in expanding on the methods of [12] used a GMM approach to remedy this. This approach makes optimal use of the information provided by time-dependent covariates, when obtaining regression coefficients estimates. However, the associations over time are linked and not parsed out.

In longitudinal studies, the feedback of a response on a covariate is of great importance at times, especially with health and health related data. Some researchers have pointed out that inter-dependencies should not be ignored in any longitudinal data modeling [20]. Liang and Zeger [21] indicated how GEE can be used to study feedback. As an example, [22] demonstrated the impact of feedback and long-term effects through the analysis of data collected from Indonesian pre-school children who were medically examined quarterly for eighteen months. The aim was to assess the role of vitamin A deficiency in children's morbidity. At each visit, it was noted whether a child had xerophthalmia, (an ocular condition due to vitamin A deficiency, respiratory infection or diarrheal infection), as well as age, weight and height. They were interested in whether there exists a feedback mechanism whereby children with vitamin A deficient are more likely to suffer respiratory and diarrheal infections, which in turn depleted stores of vitamin A and increase their risk of subsequent infections. Similarly, we want to identify covariates which have an impact on obesity and whether obesity in turn has an impact on certain covariates, *physical activity*, and *depression scale*. We find this to be of importance since it has public health importance as obesity is among the leading causes of diabetes and depression. To investigate whether such a feedback exists, we must simultaneously model or characterize the expectation of a subject's response to obesity at certain times as a function of the subject's covariate at particular time. However, we chose to fit marginal models as opposed to transition models that seem to characterize the expectation of a subject's response to obesity at time t as a function of the subject's covariates at other times, and the subject's past response at times. As researchers have pointed out and we concur, marginal models are appropriate when inferences about the population average are the primary focus or when future applications of the results require the expectation of the response as a function of current covariates, [12]. We expand on these and use a GMM logistic regression model with valid moments and allow different associations across waves or periods.

Add Health Data

The National Longitudinal Study of Adolescent to Adult Health (Add Health) is a school-based, longitudinal study of the health-related behaviors of adolescents and their outcomes in young adulthood. Beginning with an in-school questionnaire administered to a nationally representative sample of 6504 students in grades 7 through 12 in 1994–95 (wave 1), the study followed up with a series of in-home interviews of students approximately one year (wave 2), then six years later (wave 3), and finally in 2008 (wave 4). We focus our analysis on respondents who reported their obesity status across all four waves. Using these data, we explored the impact of feedback between obesity and a set of time-dependent covariates (physical activity, and depression) overtime and adjusted for certain covariates, age, gender, smoking and alcohol.

We used the Add Health data to obtain estimates of regression coefficients based on feedback that relates obesity to certain covariates. These longitudinal data were also used to assess the health and socioeconomic effects of obesity. *Obesity* was measured using each subject's BMI, based on self-reported height in feet and inches and weight in pounds in all waves. Following the CDC growth chart for children and teens and the growth chart for adults [23] obesity was defined as a BMI (kg/m^2) greater than or equal to the 95th percentile for age and gender. We categorized subjects with BMI over the critical value as obese, and the remaining as non-obese. We focused on the following covariates, which were collected over the four waves. *Physical activities*, computed based on the responses to several questions. Physical activities may impact obesity by impacting exercise, which is an essential ingredient for maintaining a healthy body weight. Such effects may accumulate over time if lifestyle choices are habit-forming. Public health guidelines suggest that adolescents should engage in at moderate to vigorous physical activity at least three times per week. Depression Scale was measured with the Center for Epidemiologic Studies *Depression Scale* based on a series of questions covering the child's emotional well-being. *Gender*: Female was the reference category as evidence suggests obesity is more common among adolescent males [24,25]. *Race/ethnicity* as a covariate was categorized as white and non-white.

GMM logistic regression Model

As a feedback model to these *Obesity* data, we begin to fit a population-averaged logistic regression model with time-dependent covariates Z_{kit} (depression, television hours, activity scale, and alcohol) as,

$$\text{logit}[P(O_{it} = 1|\mathbf{Z}_{it})] = \beta_0 + \beta_1 Z_{depression\_t} + \cdots + \beta_p Z_{alcohol\_t}$$
$$Z_{kit} = \rho Z_{kit-1} + \tau_{it}$$

where O_{it} is a binary response for student i measuring obesity in time t the Z_{kit} is the i^{th} individual measured at the t^{th} time period for the k^{th} covariate, p is a measure of the correlation between the t^{th} and $t+1^{th}$ observation, is an error term β_k and are the regression coefficients associated with Z_{kit}. We fit a model that accounts for the correlation among the responses as well as feedback with the response and the covariate. This approach involves identifying valid moment conditions to determine when these correlations are measurable. We use GMM estimation to make optimal use of the valid moment conditions. We categorize and summarize these correlations and feedback in Figure 1. Therefore, obesity in time may impact the covariates in later time points *t+s*. The fact is that if one is interested in

providing information about future responses, it is necessary for the expectation of the response to be a function of the current covariates. Thus, we fit a flexible class of feedback logistic regression models, by specifying only the form of the conditional mean and variance of each response given a subset of the other responses at the same time as well as prior times. However, these models (LWY and LS) do not allow the associations between the covariate and the response to be differentiated in different waves. Thus, we utilize a unique model in the Partitioned GMM logistic regression model as an alternative approach to existing GMM logistic regression models [11,12] for time-dependent covariates.

Partitioned GMM Model

The Partitioned GMM model accounts for the relationships between the outcomes and covariates within the same time-period, as well as in different time-periods using extra regression parameters. Instead of grouping all valid moment conditions, this approach partitions the moment conditions to separate the effects of the covariates on the responses across time. The valid moment conditions are grouped based on the time lag between the covariate and the response. This approach is best applied to data without many repeated observations as compared to the number of subjects, [5].

The Partitioned GMM model estimates the relationships between the outcomes observed at time t, Y_t, and the j^{th} time-dependent covariate observed at time s, X_{js} for $s \leq t$. In fitting this model, for each time-dependent covariate X_j measured at time s=1,2,..., T; for the i^{th} subject, the data matrix is reconfigured as a lower triangular matrix,

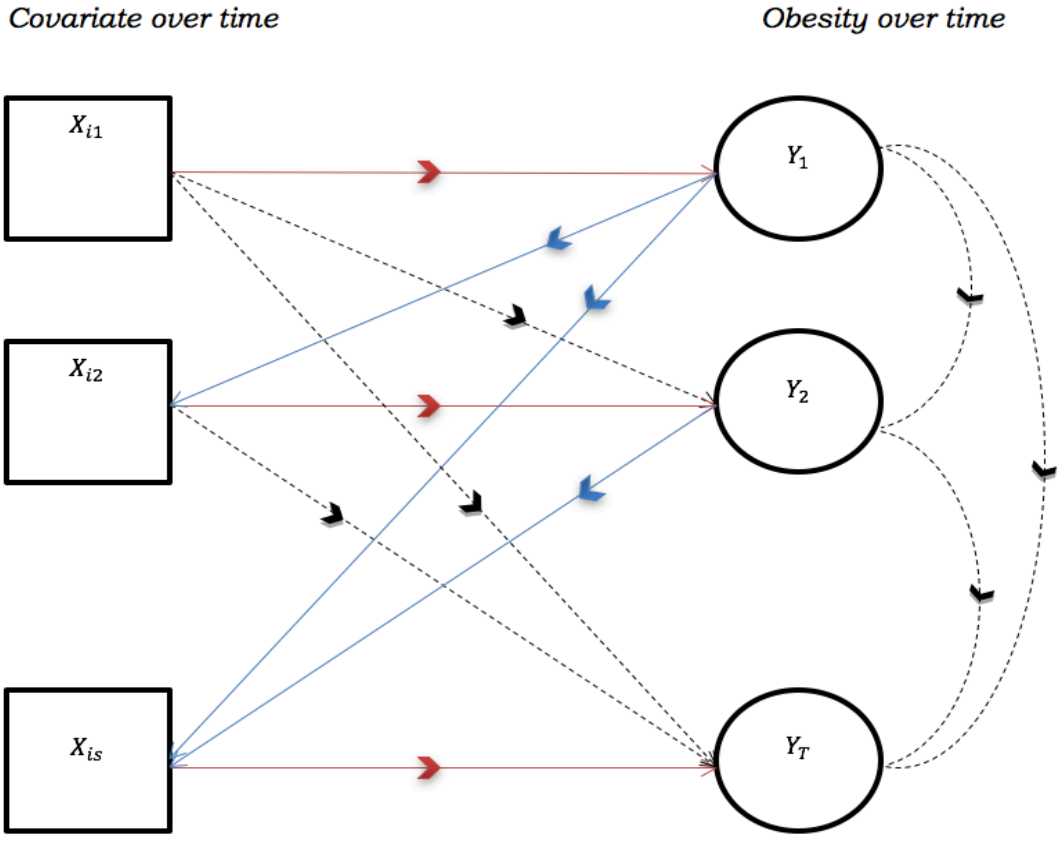
$$\mathbf{X_{ij}} = \begin{bmatrix} 1 & X_{ij1} & 0 & \dots & 0 \\ 1 & X_{ij2} & X_{ij1} & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{ijT} & X_{ij(T-1)} & \dots & X_{ij1} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & X_{ij}^{[0]} & X_{ij}^{[1]} & \dots & X_{ij}^{[T-1]} \end{bmatrix}$$

where the bracketed superscript denotes the difference t-s, in time-periods between the response time and the covariate time s. Thus, the model is

$$g(\mu_{it}) = \beta_0 + \beta_j^{tt} X_{ij}^{[0]} + \beta_j^{[1]} X_{ij}^{[1]} + \beta_j^{[2]} X_{ij}^{[2]} \dots + \beta_j^{[T-1]} X_{ij}^{[T-1]} \quad (3.1)$$

and in matrix notation $g(\mu_i)=X_{ij}\beta_j$, where the X_{ij} matrix denotes the systematic component and the mean $\mu_i = (\mu_{i1},\dots, \mu_{iT})$ depends on the regression coefficients $\beta_i=(\beta_0, \beta_i^{tt}, \beta_i^{[1]}, \beta_i^{[2]},\dots, \beta_i^{[T-1]})$. The coefficient β_i^{tt} represents the effect of the covariate X_{jt} on the response when observed in the same tth time-period s<t. When the covariate is observed in an earlier time-period , we denote

FIGURE 1. Types of correlation structuresls



the lagged or carry-over effect of X_{is} on Y_t the by the set of regression coefficients. These added coefficients allow the effect of the time-dependent covariate on the response to vary across time and to be identified separately, rather than using a single linear combination of all valid moment conditions. For instance, $\beta_j^{[1]}$ denotes the lagged effect of X_{is} on Y_t across a one time-period lag. Each of the J time-dependent covariates produces a maximum of T partitioned coefficients of β_j. Let β denote vector composed of the concatenation of the regression parameters associated with the J time-dependent covariates. Then, the data matrix X will be of maximum dimension N by T, and β is of maximum length Jx(T+1). Despite the addition of these extra regression parameters, the estimates produced by this approach remain reliable when the number of clusters is large relative to the number of time-periods. In selecting valid moment conditions, we utilize the test of validity discussed [11] as well as the type II covariate discussed [12], though the approach can also utilize alternative techniques for selecting moment conditions, [5]. These models are fitted for the partitioned GMM: https://github.com/kirimata/Partitioned-GMM

RESULTS

The response and time-dependent covariates under consideration at each wave, number of hours spent watching television, amount of physical activities; depression scale, alcohol and age are described in Table 1. There are 5.8% obese in wave 1, increasing to 7.8% in wave 2, 22.9% in wave 3, and to 35.7% in wave 4. There is a decline in physical activity and an increase in depression scale.

From an unconditional logistic regression model with random intercept in wave, we obtain an intraclass correlation of $\frac{3.8853}{3.8853+3.29}$=0.5415. The denominator is the sum of the variance of the latent continuous variable and a level-1 residual that follows a logistic distribution with a mean of zero and a variance of 3.29 [26]. Thus, 54.15% of the total variation in adolescent obesity status is attributed to random effects over time. The other 45.85% of variation is explained by other covariate. This large measure of intraclass correlation necessitates a model that accounts for such, [5].

If one ignores the large intraclass correlation and fit a standard logistic regression model to each wave of the data, one will ignore any feedback. This approach does not account for the possibility that the students' obesity statuses can impact their depression or their physical activities or vice versa. As such, fitting such models only provides a snapshot or a cross-sectional view, Table 2.

Considering the correlation, the feedback, and the changing impact across waves, we fit a Partitioned GMM logistic regression model with time-dependent covariates. It adjusts for all these factors including allowing the association to vary between waves. These results are provided in Table 3. The Partitioned GMM model was fitted using the Lalonde, Wilson and Yin [11] approach to testing valid moment conditions (Partitioned-LWY), as well as using the Lai and Small [12] type II covariate (Partitioned-LS). Both these Partitioned GMM models allow us to adjust for time-dependent covariates and to evaluate the varying effects over time, though the Partitioned-LS model may include moment conditions which are not valid. We provide these results, as well as the results for the LWY-GMM and LS-GMM in Table 3, though the latter do not partition out the impact of the covariates. There are four waves; thus, these models (Partitioned LWY and Partitioned LS) produce results for cross-sectional, lagged one period, lagged two-period, and lagged three-period parameters. In comparison, the LWY-GMM and LS-GMM are cross-sectional models and as such produce only one parameter estimate per covariate. For these cross-sectional models, we found that activity ($p<0.001$) and depression ($p<0.001$), amongst other covariates, had significant impacts on *Obesity*. Conversely, when using the Partitioned-GMM approaches, we found that in the early years, alcohol had no impact ($p=0.895$ and $p=0.476$) but that in later years it did ($p<0.001$). In the early years, television hours had a significant impact on obesity status, but that as the children got older, it did not, Table 3.

Had we ignored the change of association across the waves and run the LWY-GMM model or the LS-GMM model we would have concluded that depression, television hours, activity, and alcohol have an impact on obesity. However, when using the Partitioned GMM to analyze these data by wave, we concluded that activity and depression had no impact on obesity, though alcohol showed an early impact.

TABLE 1. Mean or Percentage of Event at each wave

Variable	Wave 1	Wave 2	Wave 3	Wave 4
Obese	5.8%	7.8%	22.9%	35.7%
Television hours	15.863	14.031	12.058	12.775
Activity Scale	2.565	2.341	1.247	0.899
Depression Scale	0.794	0.806	1.101	0.984
Alcohol	53.1%	49.4%	80.3%	83.0%
Age	15.600	15.987	21.447	28.600

TABLE 2. Factors Affecting Obesity through Logistic regression models for Waves 1-4

	WAVE 1		WAVE 2		WAVE 3		WAVE 4	
Parameter	**Estimate**	**P-value**	**Estimate**	**P-value**	**Estimate**	**P-value**	**Estimate**	**P-value**
Intercept	-1.906	0.001	-2.582	<.0001	-1.924	<.0001	0.814	0.083
Age	0.141	<.0001	0.167	<.0001	0.109	<.0001	0.041	0.007
Race_	0.249	0.010	0.258	0.004	0.232	<.0001	0.130	0.014
Gender	-0.271	0.005	-0.093	0.293	0.061	0.292	-0.089	0.082
Television hours	0.013	<.0001	0.012	<.0001	0.013	<.0001	0.008	<.0001
Activity scale	-0.060	0.169	0.013	0.723	0.042	0.252	-0.031	0.334
Depression scale	-0.023	0.885	-0.113	0.458	0.096	0.317	0.103	0.203
Alcohol	-0.246	0.020	-0.177	0.057	-0.232	0.001	-0.215	0.001

TABLE 3. Time-Dependent Covariate models

			PARTITIONED-LWY		PARTITIONED-LS	
	Parameter	**Odds**	**Estimate**	**p-Value**	**Estimate**	**p-Value**
	Intercept	0.049	-3.025	<.001	-3.076	<.001
	Race	1.249	0.222	0.020	0.074	0.433
Cross-sectional	Depression	1.650	0.501	<.001	0.384	<.001
	TV Hrs	1.015	0.015	<.001	0.015	<.001
	Activity	0.848	-0.165	<.001	-0.059	0.057
	Alcohol	1.010	0.010	0.895	-0.060	0.414
Lagged one period	Depression	1.790	0.582	<.001	0.315	<.001
	TV Hrs	1.004	0.004	0.089	0.002	0.216
	Activity	0.909	-0.095	<.001	-0.028	0.197
	Alcohol	1.047	0.046	0.476	0.078	0.189
Lagged two periods	Depression		–	–	0.661	<.001
	TV Hrs				0.013	<.001
	Activity	1.197	0.180	<.001	0.069	0.002
	Alcohol	1.343	0.295	<.001	0.068	0.283
Lagged three periods	Depression		–	–	0.417	<.001
	TV Hrs				0.012	<.001
	Activity	1.171	0.158	<.001	0.019	0.493
	Alcohol				0.017	0.822
			LWY GMM		LS GMM	
	Parameter	**Odds**	**Estimate**	**P-Value**	**Estimate**	**P-Value**
Cross-Sectional	Intercept	0.128	-2.059	<.0001	-2.369	<.0001
	Race	1.192	0.176	0.0563	0.310	0.0031
	Depression	2.319	0.841	<.0001	1.019	<.0001
	TV Hrs	1.016	0.016	<.0001	0.017	<.0001
	Activity	0.505	-0.683	<.0001	-0.854	<.0001
	Alcohol	1.132	0.124	0.0683	0.244	0.0004

DISCUSSION

There are many factors working in tandem contributing to the rising rates of obesity, including social demographics, participation in physical activities or sedentary activities, emotional behaviors, and social drinking and smoking

habits. These factors will have serious impact on an adolescent becoming obese, as they get older. While we do not see feedback with social demographics, we do not feel the same with participation in physical activities or sedentary activities as well as social drinking habits.

Physical and sedentary activities affect the probability that an adolescent will become obese. Adolescents who watch more television are 1.01 times more likely to be obese than adolescents who do not watch television. The activity score shows that adolescents who are more active are 1.09 to 1.19 likely to obese that people who do not participate in physical activities. These factors support the literature that more time spent on activities that do not require movement can increase the chances of being obese, while more time spent with activities that require more movement can decrease the chances of becoming obese.

Emotional factors, such as depression, can impact their chances of being obese when they get older. For example, an adolescent who is depressed is 1.65 times less likely to be obese than someone who is not depressed. Depression's negative impact on obesity can due to an adolescent's poor appetite or lack of motivation, which are common side effects to depression. An adolescent's drinking habits can also impact their obesity. Adolescents who are social drinkers are 1.34 times less likely to be obese than adolescents who consume more than three to four drinks or do not drink at all.

Given the complexity of data such as those from the Add Health study, an appropriate model such as the Partitioned GMM is important in identifying the effect of time-dependent covariates with feedback, correlation, and changing associations across periods using valid moment conditions.

Acknowledgements

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

Funding Sources

RR was supported by Add Health Users Co from 2016

JRW was supported by National Institute of Health Alzheimer's Consortium Fellowship Grant, Grant No. NHS0007.

References

1. Baum CL and Ruhm CJ. Age, socioeconomic status and obesity growth. J Health Econ 2019;28 (3):635-648
2. Gutin B, Litaker M, et al. Body-composition measurement in 9 –11-y-old children by dual-energy X-ray absorptiometry, skinfold-thickness measurements, and bioimpedance analysis. Am J Clin Nutr 1996;63:287
3. Dionne I, Almeras N, et al. The association between vigorous physical activities and fat deposition in male adolescents. Med Sci Sports Exerc 2000;32:392
4. Fletcher, JM , Green, JC Neidell MJ - Long term effects of childhood asthma on adult health. J Health Econ 2010 – Elsevier Pages 377-387
5. Irimata, K.M. Broatch, J. and Wilson JR. (2019). Partitioned GMM Logistic Regression Models for Longitudinal Data Statistics in Medicine accepted
6. Stockwell T et al., eds (2005). Preventing harmful substance use: the evidence base for policy and practice. London, John Wiley and Sons.
7. Gruber J and Frakes M. Does falling smoking lead to rising obesity? J Health Econ 2006;25(2):183-97
8. Pu J, Fang D, Wilson JR [2017] Impact of communities, health, and emotional-related factors on smoking use: comparison of joint modeling of mean, dispersion, and Bayes hierarchical models on ADD Health Survey. BMC Med Res Methodol 2017;17(1):20
9. Chou SY, Grossman M, Saffer H. An Economic Analysis of Adult Obesity: Results from The Behavioral Risk Factor Surveillance System. J Health Econ 2004; 23(3):565-87
10. Richardson LP, Davis R, Poulton R, McCauley E, Moffitt TE, Caspi A, Connell F. A longitudinal evaluation of adolescent depression and adult obesity. Arch Pediatr Adolesc Med 2003;157(8):739-745.
11. Lalonde T, Wilson J, Yin J. (2014). GMM logistic regression models for longitudinal data with time-dependent covariates and extended classifications. Stat Med 2014; 33(27), 4756-4769.
12. Lai TL, Small D. Marginal Regression Analysis of Longitudinal Data. J R Stat Soc Series B Stat Methodol 2007;69(1):79-99.
13. Zhou Y, Lefante J, Rice J, Chen S. Using modified approaches on marginal regression analysis of longitudinal data with time-dependent covariates. Stat Med. 2014;33(19):3354-3364.
14. Cai K, Wilson JR. SAS Macro for Generalized Method of Moments Estimation for Longitudinal Data with Time-Dependent Covariates. Paper presented at: SAS Global Forum2016; Las Vegas, NV.
15. Wilson JR and Lorenz K (2015) Modeling binary correlated responses using SAS, SPSS and R Springer
16. Allison PD (2012) Logistic regression using SAS: Theory and application
17. Agresti A (2002). Categorical Data Analysis (2nd Edition).
18. Pepe M, Anderson G. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. Commun Stat Simul Comput 1994;23(4), 939–951.
19. Hu F.-C. (1993). A statistical methodology for analyzing the causal health effect of a time dependent exposure from longitudinal data.

Unpublished Sc.D. dissertation, Harvard School of Public Health, Department of Biostatistics, Boston, MA.
20. Zeger S, Liang K. Feedback Models for Discrete and Continuous Time Series. Stat Sin 1991; 1:51-64.
21. Zeger S, Liang K. Feedback Models for Discrete and Continuous Time Series. Stat Sin 1991; 1:51-64.
22. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73:13-22
23. Sommer A, Kat, J, Tarwoto I. Increased risk of respiratory disease and diarrhea in children with pre-existing mild vitamin A deficiency. Am J Clin Nutr 1984;40:1090-1095
24. Kuczmarski RJ, Ogden CL, Guo SS, et al. 2000 CDC growth charts for the United States: Methods and development. National Center for Health Statistics. Vital Health Stat 2000;11(246): 2002
25. Troiano RP, Flegal KM. Overweight children and adolescents: description, epidemiology, and demographics. Pediatrics. 1998;101:497–505
26. Goodman E. The role of socioeconomic status gradients in explaining differences in US adolescents' health. Am J Public Health 1999;89:1522–1528
27. Snijders TAB, Bosker RJ (1999) Multilevel analysis: An Introduction to basic and advanced multilevel modeling. Thousand Oaks, CA: Sage.

*