Sample Proposal Updated 04.01.09

April 1, 2009

Peter Meyer, Director
NCHS Research Data Center
3311 Toledo Road, Suite 4113
Hyattsville, MD 20782

Peter Meyer:

This letter is to request permission to access restricted use data from the 2005 National Health Interview Survey (NHIS) through the Research Data Center. Our research project seeks to study the effects of the female labor force participation on adult and childhood obesity. In order to merge vital neighborhood indicators from an external source to the 2005 NHIS data, we need access to restricted geographic variables, such as county and state.

Attached you will find the research proposal that follows the Proposal Outline found on the RDC website. It includes the data requested, methods, proposed output with sample table shells, and data dictionaries for all public use, restricted, and non-NCHS data.

Research Topic: Female Labor Force and Obesity
Survey and Survey Years: NHIS 1997-2005
Type of Access: On Site (Hyattsville, MD) and Remote Access
Restricted Variables Requested: County and State, Height and Weight (for children)
Primary Contact:
Name
Phone Number
Email Address

Sincerely,

Ima Researcher

B. PROJECT TITLE

"The Effects of Female Labor Force Participation on Adult and Childhood Obesity"

C. ABSTRACT

Obesity is a major public health problem in the U.S. Obesity rates in the last 20 years have more than doubled for adults, and more than tripled for children. Understanding the causes behind the increase in obesity rates is fundamental for devising policies aimed at stopping the increase (and eventually decreasing) those rates.

The objective of this project is to assess whether there is a relationship between in female labor force participation and adult and childhood obesity. We exploit a "natural experiment", the expansion of the Earned Income Tax Credit. This change in benefits affected differentially the labor supply decisions of families with children versus childless families and families with two or more children versus one child families. This allows us to identify the effects of increased female labor supply on the obesity rates of both mothers and children.

D. PERSONAL IDENTIFICATION

Ima Researcher
Primary Investigator
Department of Economics
State University
P.O. Box 1234567
City, State 99999-1234
Phone: 1-333-123-4567
Fax: 1-333-123-5678
Email: iresearcher@state.edu

Project Coordinator
(contact information)

Doctoral Student
(contact information)

Programmer
(contact information)

E. DATES OF PROPOSED TENURE AT RDC

We would like to come on site to the NCHS RDC located in Hyattsville, MD for a period of one week. After that time, we would like to begin using the remote access system for a period of 6 months. We hope to come on site in May 2009 and use remote access from June-January 2009.

F. SOURCE OF FUNDING

A grant from the Institute for Research on Poverty (IRP) under the IRP-USDA RIDGE Program will provide the funding for all the costs associated with using the services at the RDC as well as the travel expenses associated with the project.

G. BACKGROUND STUDY

1. Key Study Questions

- What relationship exists between increases in female labor force participation and increases in adult and childhood obesity?
- How does this differ in mothers with low education and their children?

There are several potential explanations for the rapid increase in the weight of the American population observed over the past three decades. One hypothesis is that increasing female labor force participation is related to rising obesity through changes in time allocation and food consumption. Chou, Grossman and Saffer (2003) suggest that women devote more time to work and less to food preparation, increasing their reliance on convenient food and fast food (which is high in caloric content).

Also, as shown by Anderson, Butcher and Levine (2003), who find that a child is more likely to be overweight the higher the hours per week worked by her mother over the child's life. Furthermore, Anderson, Butcher and Schanzenbach (2007) document that the correlation in weight outcomes between parents and children has increased since the early 1970s, so that as much as 40% of the increase in children's weight can be explained by changes in parent's weight.

An alternative hypothesis is that technological change caused food prices to decrease, and also transformed the type of work people perform, from physically demanding to sedentary jobs. Lower food prices increase the consumption of food, which translates into higher calorie consumption, while the increase in sedentary jobs implies a lowering in calories spent (Lakdawalla, Philipson and Bhattacharya, 2005; Philipson and Posner, 2003). Technological change could also operate through the decline in the time cost of prepared foods. Cutler, Glaeser, and Shapiro (2003), suggest that the lower time cost of prepared foods is behind the decline in cooking times and home meals, and behind the increase in the consumption of prepared food observed in the data.

Using U.S. aggregate data Gomis-Porqueras and Peralta-Alva (2007) study the implications of the decline in both the monetary and the time (relative) cost of prepared foods on adult calorie intake. In their framework, the time channel operates by declines in income taxes and the gender wage gap which increase female labor supply and the opportunity cost of cooking at home, thus decreasing the time spent cooking at home. Their results suggest that up to two thirds of the increase in the consumption of calories of the average adult can be explained by this channel.

In this project we will study whether the aggregate relationship between female labor force participation and obesity is confirmed by the individual-level data. The challenge is to identify the relationship of labor force participation on obesity. Previous studies have illustrated a correlation between the two variables for adults (see for example Bleich, Cutler, Murray, and Adams, 2007), and for small children of high socioeconomic status mothers Anderson, Butcher and Levine (2003) established a causal effect of hours worked on childhood obesity. In our project we will rely on changes in the Earned Income Tax Credit (EITC) in the 1980s and 1990s as a source of exogenous variation in female labor participation of low education single (and married) mothers, to identify the effects of participation on both adult obesity and childhood obesity.

Changes in the EITC implied that the maximum benefit increased in real dollars from 1986 to 1987 by 50%, from 1990 to 1991 by 25%, from 1993 to 1994 by 63%, and from 1994 to 1995 by 20%. More importantly, the changes affected differentially the incentives of taxpayers with different number of children.

Starting in 1991 the EITC credit has been different for one-child taxpayers versus two-or-more-children taxpayers. The difference in the credit was very small up to 1994, when the difference increased 25% in favor of taxpayers with more than one child. Starting in the same year, childless taxpayers became eligible for a small credit (in the order of $300 maximum).

To understand, and quantify, the relationship between changes in female labor force participation on obesity we will exploit expansions in the Earned Income Tax Credit (EITC) in 1987 and 1993, which have been credited with increasing the labor force participation of low education single mothers. This provides us with a credible empirical strategy to study the effects of labor force participation on obesity, because by making comparisons between groups and across time we will be able to control for other confounding factors that might be related to changes in obesity. The empirical analyses will use more than two decades of information from the National Health Interview Survey (NHIS).

Because the EITC can be received only by working taxpayers (primarily with children), it generates changes in the incentives to work for low income individuals. In particular for single parents (mostly women), there should be unambiguous incentives to increase labor force participation. Working mothers must trade off the advantages of greater income against the disadvantages of less time for home food production and supervision of children's activities. The main hypothesis of this project is that the increased labor force participation by single mothers caused the increase in weight problems for both the working mothers and their children. These women devote more time to work and less to food preparation, increasing their reliance on convenient food and fast food (which is high in caloric content), thus consuming more calories.

2. Public Health Benefits

Given the sharp increase in adult and childhood obesity, research like the one in this project is needed to better understand why increases in obesity rates have occurred. This is an important first step for finding solutions to the nation's health problem. In particular, in order to promote specific behavioral changes, we need to understand the motivation behind these behaviors. This is key for developing policies that promote behavioral changes to counteract the sharp increases in adult and childhood obesity of recent decades. In particular, obesity rates among low income women and their children are higher than for any other group. If labor force participation can explain an important part of the changes in weight problems for those groups, then the results from this research can help tailor intervention campaigns as well as determine how to efficiently allocate funds to reduce obesity rates. For example, if women are more likely to eat foods prepared away from home in order to be able to work, then it is important to provide information and guidance on how to eat healthily.

H. DATA REQUIREMENTS

1. Data System(s) NHIS

Data Year(s) 1997-2005

Data File(s) Sample Adult, Sample Child, Household

Unit of Analysis sample adults 20-50 years old, sample children 12-18 years old (two different analyses)

2. Cases to be Included

Our analyses will be different for adults and children (where children are defined as those individuals with age less or equal to 18 in the calendar year of the survey).

For adults we will consider only individuals in the ages 20 to 50 years old. This gives us a total of 1,092,150 adults. In addition, we will impose the condition that all adults considered are not disabled and are not students. Furthermore, we will drop from our analysis file all families for which the family structure cannot be determined

appropriately, in particular the number of children (and who is likely to be the parent or guarding of those children), given that the number of children is the key variable associated to the amount of EITC benefits. These constraints imply that our final analysis file will have in the order of 1,000,000 observations for adults (once we account for missing values for several of the key variables in our analyses the final number may be a lit bit less than that).

For certain regressions we will use data on children of ages 12 to 18. We have identified approximately 200,000 observations in the NHIS for those ages, with valid information on our variables of interest.

3. Public Use Variable Summary

Demographic variables
Education variables
Employment variables
Height, Weight, Body Mass Index (BMI)

4. Restricted Use Variable Summary

Geographic indicators: state and county FIPS
Height and weight for children

5. Non-NCHS Variable Summary

From Census Bureau's Covered Employment and Wages (CEW):

Local labor market conditions
Welfare policy variables
Income tax data
Density of full-service and fast-food establishments

6. Explanation of Need for Restricted Data

For all our analyses, we think it is necessary to control as well as possible for differences in labor market conditions, local regulations, local services, local availability of prepared-food establishments, etc. This can only be done by accessing the restricted-use version of the NHIS because we will need to link using geographic indicators. We plan to include measures of local labor market conditions (like average employment and earnings in different sectors, from the Census Bureau's Covered Employment and Wages, CEW, data), measures of the average generosity of the welfare state program and tax rates at the state level.

In addition to being able to run county-fixed effects regressions and to add county and state level time-varying variables in the analyses, our last reason to request access to the restricted-use version of the NHIS is that we want to analyze data on obesity for children, not only for adults. However, weight and height (and thus BMI) information for children is not available in the public-use version of the NHIS.

I. METHODS

1. Analytic Strategy and Statistical Methods

This research project follows the strategy of Hotz, Mullin and Scholz (2005) of comparing the changes in labor force participation and obesity rates of single mothers with one child, versus single mothers with two or more children, before and after the 1994 EITC expansion for families with more than one child. In addition, we follow the strategy of Meyer and Rosenbaum (2001) of comparing single women with children versus single women without kids, before and after the 1987 EITC expansion. If the results show that the effects on labor supply are similar to the ones found in the literature, it will validate the application of this methodology to the study of changes in obesity rates of adults and children.

Similar analyses can be performed for the evolution of Body Mass Index (BMI), rates of overweight and obesity, for adults and children. So, differences-in-differences regressions of the following form will be examined:

$$Yi = ß0 + ß1\ Xi + ß2\ Dkids + S.t\ Dyear\_t + Sdt\ (Dyear\_t * Dkids) + .c + ui, tt$$

where i refers to an individual; the variable Yi will represent alternatively an indicator variable for whether the person is employed, the BMI of the person, an indicator for whether the person is overweight, or an indicator for whether the person is obese. The vector Xi will include a variety of individual covariates like age, race and ethnicity, number of children of different ages, education levels, plus a series of county- and state-level variables. The variable Dkids will be an indicator for whether the family of person i is a two-or-more children family (for the specifications where the comparison group will be one-child families), or it will be an indicator for whether the family of person i has any children (for the cases where the comparison group will be women with no children). The t variables Dyear_t represent indicator variables for the year in which observation i is observed (of course there will one year omitted from the regressions), and .c represent the county fixed effects. The coefficients of interest here are the dt coefficients associated to the interaction terms between the kids dummy and the year dummies. These coefficients identify the differences in the average dependent variable for the two groups identified by the variable Dkids.

With our proposed methodology we can to assess whether a relationship exists between increases in female labor force participation and increases in adult and childhood obesity. In particular, we exploit a “natural experiment”, the expansion of the Earned Income Tax Credit. An important advantage of our approach is that by comparing types of families at the same point in time, we are able to hold constant confounding factors (like price changes and technological changes) that have been proposed as explanatory for the increase in obesity. In that sense, unless these factors differentially affect families with different number of children, our strategy presents a clean way of understanding the effects of female labor force participation on obesity. Finally, the key assumption for the differences-in-differences method to be valid is that the “treatment” and “comparison” groups have not changed in composition over time. Comparisons of one-child versus two-or-more-children families should be less subject to this concern.

2. Software Requirements

During our on site visit we will use both Stata and SAS. We recognize that analyses completed using the remote access system must use SAS exclusively.

3. Sampling Weights and Standard Errors

All summary tables and regressions will use sampling weights that take into account the probability of being selected into the sample for each observation. The proper weight variable to use is the final annual weight WFTA_SA for the sample adult and WFTA_SC for the sample child files of the NHIS.

References:

(Use any standard bibliographic or reference format)

J. DESCRIPTION OF OUTPUT:

We plan to produce some summary statistics tables (where each cell will be big enough), and regression results, where we believe there will not be any disclosure risks. We have done some preliminary analyses, and even when looking at the smallest groups in which we are interested in, by year, the smallest cell size is close to 100 observations. The desired output will consist of the following four tables.

Note to Researchers: These are very brief examples of table shells. Your examples should be as complete as possible. It is also appropriate to include examples of graphs or other forms of output in this section (only table shells are shown). Please make sure that your output matches your research questions and methods.

Table 1: Summary statistics providing the mean and standard deviations of all the variables used in the regressions.

	Mean	Standard Deviation
Education Less than high school High school Some college Completed college		
Number of Children 1 2 3 4 5+		
Marital Status Single Married		
Employment Status Unemployed Full-Time Part-Time		
BMI Overweight Obese		

Table 2: Summary statistics describing the employment rates for women ages 19-50 by education, number of children and marital status.

	% Unemployed	% Full-Time Employed	% Part-Time Employed
Education Less than high school High school Some college Completed college			
Number of Children 1 2 3			

4 5+			
Marital Status Single Married			

Table 3: Summary statistics describing the average BMI, overweight and obesity rates for women ages 19-50 by education, number of children and marital status.
Table 4: The same as Table 3 but for the children ages 12-18 of the women included in Table 3.

	Average BMI	% Overweight	% Obesity
Education Less than high school High school Some college Completed college			
Number of Children 1 2 3 4 5+			
Marital Status Single Married			

Table 5: Odds Ratio of overweight and obesity for women 19-50.
Table 6: Odds Ratio of overweight and obesity for children by maternal characteristics.

	Overweight			Obese		
	OR (95% CI)	Beta Coefficient	P value	OR (95% CI)	Beta Coefficient	P value
Education Less than high school High school Some college Completed college						
Number of Children 1 2 3 4 5+						
Marital Status Single Married						

Appendix 1.

Curricula Vitae for each person who will participate in the research team

Curriculum Vitae for the Principal Investigator.

Curriculum Vitae for Second Researcher.

Curriculum Vitae for Third Researcher.

Curriculum Vitae for the Programmer.

Appendix 2: Letter for Student Researcher: N/A

Appendix 3: Restricted Data Dictionary (variables requested from the RDC)

Note to Researchers: These are very brief examples of a data dictionary. It is important to begin each dictionary with the variable that will be used to merge the datasets. If you are using multiple years of data, you may have a different data dictionary for each year. If you are using multiple files from the same survey (as shown here), please separate the variables. Your data dictionaries must include all the variables you would like to be in your final dataset (they are often many pages long).

VARIABLE NAME	VARIABLE LABEL
ID	Identification Number (merge to public use data files)
STFIPS	State FIPS Code (merge to non-NCHS data)
CNTYFIPS	County FIPS Code (merge to non-NCHS data)
CHLDHT	Height of Child
CHLDWT	Weight of Child

Appendix 4: Researcher Supplied Data Dictionaries

Public Use Variables: Household File

VARIABLE NAME	VARIABLE LABEL
ID	Identification Number (merge to public use data files)
HHNUM	Number of People Living in Household
INCOME	Average Household Income

Public Use Variables: Sample Adult File

VARIABLE NAME	VARIABLE LABEL
ID	Identification Number (merge to restricted use data files)
GENDER	Gender
AGE	Age

Public Use Variables: Sample Child File

VARIABLE NAME	VARIABLE LABEL
ID	Identification Number (merge to restricted use data files)
GENDER	Gender
AGE	Age

Data from non-NCHS Source: Census Bureau's Covered Employment and Wages

VARIABLE NAME	VARIABLE LABEL
STFIPS	State FIPS (merge to restricted use data file)
CNTYFIPS	County FIPS (merge to restricted use data file)
EMPOP	Employment/Population Ratio
EARN	Average Real Earnings
UNEMPL	Unemployment Rate

5. Description of Email System

All output will be sent to imaresearcher@university.edu. This email client used is Microsoft Outlook and is secure and maintained by University Information Technology personnel.