American Economic Journal: Economic Policy 2014, 6(2): 258–290
http://dx.doi.org/10.1257/pol.6.2.258

Giving Mom a Break: The Impact of Higher EITC Payments on Maternal Health†

By William N. Evans and Craig L. Garthwaite*

The 1993 expansions of the Earned Income Tax Credit created the first meaningful separation in benefits between families containing two or more children and those with only one child. If income is protective of health, we should see improvements over time in the health for mothers eligible for these higher EITC benefits. Using data from the Behavioral Risk Factors Surveillance Survey, we find improvements in self-reported health for affected mothers. Using data from the National Health and Nutrition Examination Survey, we find reductions in the probability of having risky levels of biomarkers for these same women. (*JEL* H24, I12, I14, J16)

The Earned Income Tax Credit (EITC) is a refundable tax credit that provides cash payments to poor families and individuals, with the most generous payments for families with children. In 2008, the program distributed $49 billion in payments to 24 million people,[1] roughly the same level of spending for Temporary Assistance for Needy Families (TANF) and the Supplemental Nutritional Assistance Program (SNAP) programs combined.[2] Families earning the maximum credit could see their adjusted gross income increased by as much as 15 percent.

The 1993 expansions of the EITC created the first meaningful separation in benefit levels for families based on the number of children, with families containing two or more children receiving substantially more in payments. We exploit this change to examine the impact of income on health for low income mothers. If income is protective of health, we should see improvements over time in the health for EITC eligible mothers with two or more children compared to those with only one child. This empirical methodology has been used by Hotz, Mullin, and Scholz (2006) and Adireksombat (2010) in their analysis of the 1993 expansions on employment.

Using data from the Behavioral Risk Factor Surveillance System (BRFSS), we find evidence that the higher EITC payments increased self-reported health and

* Evans: Department of Economics, University of Notre Dame, 437 Flanner Hall, Notre Dame, IN 46556 (e-mail: wevans1@nd.edu); Garthwaite: Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208 (e-mail: c-garthwaite@kellogg.northwestern.edu). The authors thank the Russell Sage Foundation for financial support, Ron Mariutto and Andrew Kenney for excellent research assistance, and seminar participants at Dartmouth College, the University of Illinois-Chicago, and the University of Oregon for their helpful comments.

† Go to http://dx.doi.org/10.1257/pol.6.2.258 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

[1] http://www.eitc.irs.gov/central/press/.

[2] SNAP payments total $34.6 billion in calendar year 2008 (http://www.fns.usda.gov/FSP/faqs.htm) while assistance payments under TANF totaled $14.5 billion in fiscal year 2008 (http://www.acf.hhs.gov/programs/ofs/data/2008/overview.html).

reduced the number of poor mental health days reported by eligible women with two children compared to similar women with only one child. We also use data from the National Health and Nutrition Examination Survey (NHANES) to estimate the effect of the EITC expansion on health indicators that are measured by blood and medical tests. We utilize data on eight biomarkers that indicate whether the respondent has problems associated with cardiovascular diseases, metabolic disorders, and inflammation. These biomarkers are in most cases associated with higher stress levels and are predictive of a wide range of conditions. The expansion of the EITC is associated with a large and statistically significant decrease in the counts of risky biomarkers, especially those measuring inflammation.

The results and methods in this paper contribute to two distinct literatures. The first is the literature examining the economic consequences of the EITC. As we outline below, analysts have examined the impact of the EITC on outcomes as diverse as labor supply, fertility, marriage, living arrangements, poverty, educational attainment, and spending patterns. Little attention has been paid to the effect of the EITC on health, despite the fact that improving the living conditions of low-income families was an explicit objective of the EITC.

Research on the health effects of government programs has traditionally centered on those programs that directly affect the provision of medical services such as Medicaid (Currie and Gruber 1996a, 1996b), and Medicare (Card, Dobkin, and Maestas 2009), policies that influence the ability to obtain health insurance coverage or care (Bitler, Gelbach, and Hoynes 2005), or increase access to food and nutrition through the Women, Infants, and Children program (Hoynes, Page, and Stevens 2009). Adler and Newman (2002, 63) noted that there is "… little research in the United States examining how redistributive policies or other income distributions changes affect health outcomes." Only recently have authors begun to consider the role of income support and transfer programs on health. Milligan and Stabile (2011) examined the impacts of the Canada Child Tax Benefit and the National Child Benefit Program on a variety of outcomes for the child and mother. The authors found that higher tax benefits generated statistically significant higher test scores, height, and self-reported health among children. They also found that higher tax benefits reduced a mother's depression score but there is no impact on the mother's self-reported health. In contrast, we find improvements in the self-reported health of affected mothers. Furthermore, Milligan and Stabile (2011) were unable to examine the effect of these Canadian programs on the medically measured biomarkers we utilize.

This current work also advances the understanding of the link between income and health. A large literature with contributions from a variety of disciplines has established that health outcomes are much better among individuals with higher socioeconomic status. Despite the robust correlations, the literature has failed to definitively answer whether the income/health gradient represents a causal relationship. Those with more income are not a random sample of people and the factors that lead one to have higher socioeconomic status (patience, persistence, parents with resources, etc.) may also play a role in improving health outcomes. Likewise, health shocks reduce both health status and income so poor health may cause lower income rather than the other way around (Smith 1999). Given this possibility of reverse causation and the lack of an obvious causal pathway from income to health, Deaton (2003, 118) notes that "… much of the economics literature has been skeptical about *any* causal link from income to health, and instead tends to emphasize causality in the opposite direction … ."

Economists have attempted to identify whether income impacts health by exploiting exogenous variation in income such as winning the lottery (Lindahl 2005), German reunification (Frijters, Haisken-DeNew, and Shields 2005), receipt of an inheritance (Meer, Miller, and Rosen 2003), a drop in income due to crop damage (Banerjee et al. 2007), a rise in South African pensions (Case 2004), changes in Social Security payments (Snyder and Evans 2005), and permanent changes in cohort earnings brought about by technological shocks (Adda, Banks, and von Gaudecker 2009). However, the results from these papers are incredibly varied.

The conflicting evidence from previous studies is due to at least two factors. The first is that many of the papers listed above exploit unusual shocks to income that are not replicable and the results might therefore have limited external validity. In contrast, the source of variation in this paper is a change in income affecting tens of millions of low income Americans every year. The second is the primary focus on self-reported health and mortality as the outcome of interest. Mortality is rare among many demographic groups (including the one we consider here) and, therefore, failing to detect a causal effect of income on mortality could be a Type-II error. Self-reported health outcomes are subject to a variety of well-known biases. This paper represents one of the first efforts in the economics literature to utilize medically documented biomarkers as the outcomes of interest. As we document below, these indicators of health are well measured, objective, and predictive of future health events. Our results suggest that biomarkers are a promising avenue for answering many questions. Given the decreasing costs of collecting these data, they are now being incorporated into many national and cross-national datasets.

To the extent that the results of this analysis of the health effects of the 1993 EITC expansion can be generalized to individuals on similar income support programs, they could provide valuable information regarding optimal policy decisions regarding redistributive programs. As Lindahl (2005) stated "if income causally determines health, an evaluation of a policy affecting people's income should take into account its effect on their health."

I. The Earned Income Tax Credit and the Omnibus Reconciliation Act of 1993

The EITC is a refundable tax credit available only to individuals with positive earnings. Since its creation in 1975, there have been several large expansions of the EITC, including the Omnibus Reconciliation Act of 1993 (OBRA93),[3] which increased the typical benefit and dramatically increased the differences between the maximum benefit available to families with two or more children as compared to families with only one child.[4]

[3] Public Law 103-66. See http://thomas.loc.gov/cgi-bin/query/z?c103:H.R.2264.ENR:.
[4] Prior to 1993, there were only small differences in the size of the benefit by family size.

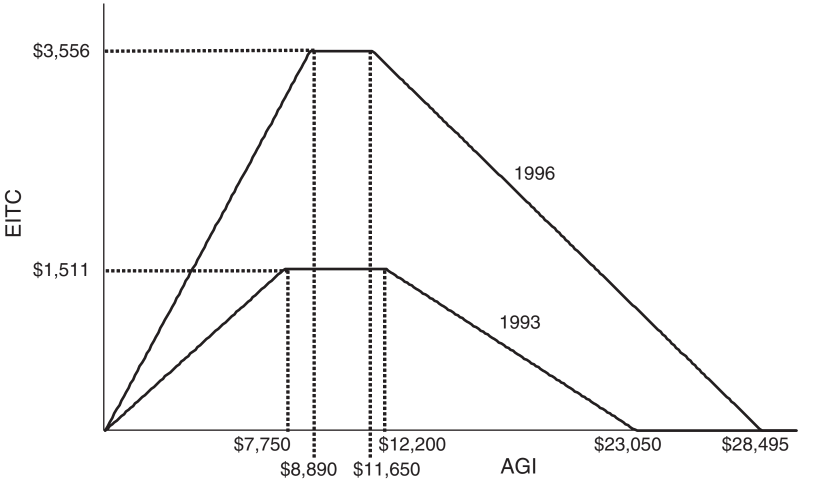

Figure 1. EITC Payments for Families with Two or More Children, 1993 and 1996

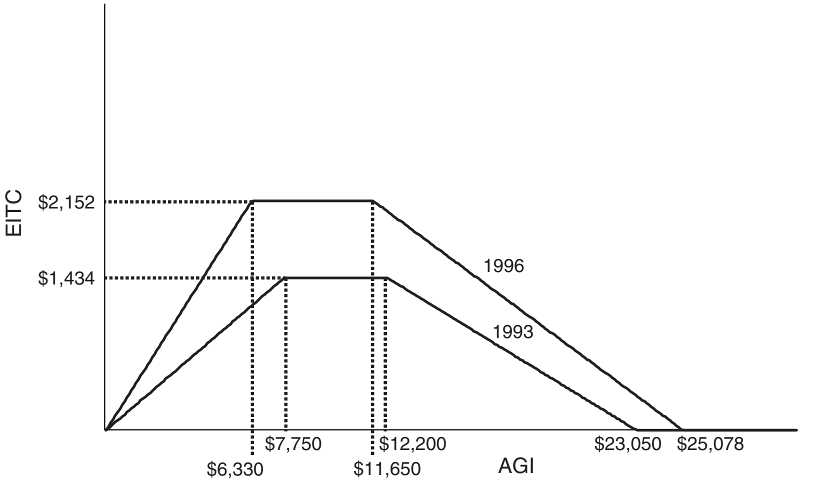

Figure 2. EITC Payments for Families with One Child, 1993 and 1996

The impact of this expansion on families with two or more children is illustrated in Figure 1, where the horizontal axis represents adjusted gross income and the vertical axis is the size of the credit. As a result of the OBRA93 expansion, the subsidy during the phase-in range for these families increased from 19.5 percent to 40 percent, and the maximum benefit increased from $1,511 to $3,556. The effect of the expansion on families with only one child is detailed in Figure 2. In this case, OBRA93 increased the size of the subsidy rate in the phase in range from 18.5 percent to 34 percent, increased the maximum benefit from $1,434 to $2,152, and decreased the phase-out rate from 21 to just under 14 percent, which extended the maximum AGI that will receive the credit from $23,000 to roughly $25,000.

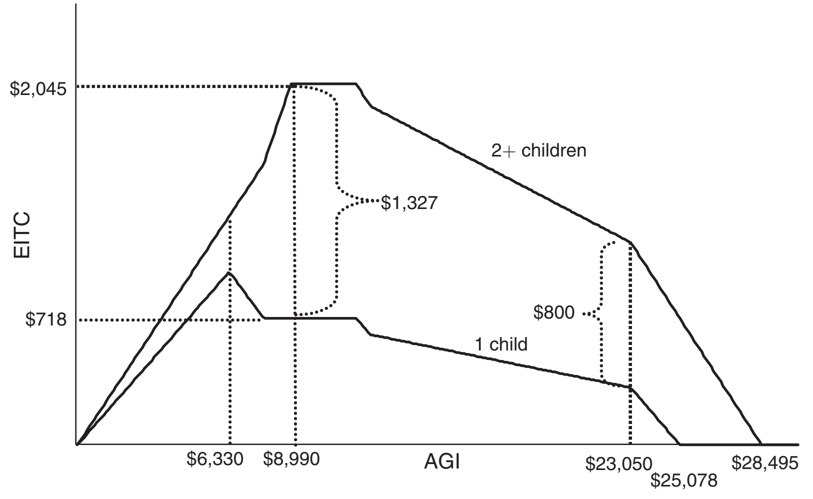

Figure 3. Difference in EITC Benefit in Families with Different Numbers of Children, 1996 and 1993

Of particular interest to this analysis are the differences in the size of the credit between families with one versus two or more children that were generated by the expansion. In Figure 3, we note the difference in the EITC between 1993 and 1996 at various levels of AGI for one and two plus children families. Following OBRA93, families with two or more children had an 18 percent greater subsidy rate and were eligible for 65 percent more in maximum benefits. As a result, between $8,900 and $23,050 in AGI, the OBRA93 expansions increased the maximum benefit by between $800 and $1,327. The difference in the maximum benefit for individuals earning $8,900 is nearly 15 percent of family income.

II. Existing Literature on the Earned Income Tax Credit

There is a large literature that examines the effects of the EITC and its expansions on a wide variety of economic outcomes and this literature is reviewed in Hotz and Scholz (2003). The most studied outcome is labor supply. In many of these papers, authors utilize difference-in-differences models and exploit changes in the structure of the program over time to identify program impacts. To isolate the EITC effects from secular changes, the authors typically use data from a comparison sample that is composed of people unlikely impacted by the reform. For example, Eissa and Liebman (1996) and Meyer and Rosenbaum (2001) examine the impact of the EITC on the labor supply of single mothers by comparing the time series changes in labor supply for women with and without children. Eissa and Hoynes (2004) used a similar methodology to examine the effect of the EITC on the labor supply of married mothers. This work suggests that the EITC raises the labor supply of single mothers but reduces the labor supply of married mothers. The results tend to be larger for women with lower years of education. Meyer and Rosenbaum (2001) estimated the EITC expansions increase in the probability of working for single women over the 1984–1996 period by 10.7 percentage points.

The evidence on whether the EITC alters hours worked is less clear with Liebman (1998) and Eissa and Leibman (1996) finding little evidence that EITC expansions altered this measure of labor supply while Dickert, Houser, and Scholz (1995), Keane and Moffitt (1998), and Meyer and Rosenbaum (2001), finding modest impacts of EITC expansions on hours of work.

The most salient article for our purposes is Hotz, Mullin, and Scholz (2006) who used administrative data from California to estimate the labor supply effects of the 1993 EITC expansions of mothers on welfare. The authors compared the changes in labor supply of women with two or more children to those of women with one child—two groups that have arguably more similar preexpansion trends in labor force participation than the typical comparisons which are women with and without children. These authors found large, positive effects of the EITC expansions on employment. Similarly, Adireksombat (2010) used data from the Current Population Survey (CPS) and implemented a similar identification strategy comparing the labor supply of women with two or more children compared to those with only one child. This analysis found large and statistically significant increases in labor supply for unmarried women with two children following the expansion of the EITC compared to similar women with only one child.

Since the amount of credit is based on family income and size, it is possible that EITC expansions impacted other family outcomes, but in general, there is little empirical evidence that the EITC has altered marriage or family formation rates (Dickert-Conlin and Houser 2002; Eissa and Hoynes 1998; Ellwood 2000) or fertility (Baughman and Dickert-Conlin 2003, 2009).

III. Identifying the Income/Health Gradient in the BRFSS Samples

In many of the papers utilizing quasi-experimental variation in income or education to assess the causal impact of socioeconomic status on health, the primary outcome of interest has been mortality. Since most beneficiaries of the EITC are relatively young, mortality rates are low and there is little hope of finding an impact of income on mortality even for large changes in income.[5] Identifying a relationship between income and health for a younger population requires thinking more broadly about the set of health outcomes. Existing research examining correlations in health disparities by socioeconomic status provides some guide as to where to look for such outcomes. Most of this literature to date has demonstrated that some of the likely mechanisms (e.g., poor health habits, environmental conditions, health insurance) explain only a small fraction of the SES/health gradient (Lantz

[5]For example, using data from the National Health Interview Survey Multiple Cause of Death data for the 1997–1999 period for women aged 21–50, we find a one-year mortality rate for this group of 0.184 percent. In a regression where the dependent variable is a dummy that equals one if a person died within one year of the survey and the covariates include controls for age, race/ethnicity status, and marital status plus the natural log of family income, the coefficient (standard error) on this last variable is −0.00064 (0.00024). Consider an experiment that would increase income by 20 percent for a randomly selected group of N people with an equally large control group. If the OLS estimate above were a "causal" impact of income on mortality, the reduced-form regression of one-year mortality on treatment assignment would generate a difference in mortality between the two groups of only 0.000128 and a simple power calculation indicates that one would need a sample of 836,000 in the treatment group (and a total sample of 1.672 million observations) to detect a statistically significant ($\alpha = 0.05$) reduced-form difference in mortality between the two groups.

et al. 1998; Cutler and Lleras-Muney 2008). A more promising line of research has focused on the potential physiological linkages between SES and health. This line of literature notes that stress has been demonstrated to produce dysfunction in the body's regulatory systems such as fight-or-flight, metabolic, immune, and the hypothalamic-pituitary-adrenal systems (Sterling and Eyer 1988; McEwan and Stellar 1993), and this stress may accelerate cell aging (Epel et al. 2004; Cherkas et al. 2006). Research has also demonstrated that those in lower socioeconomic groups have higher levels of biochemicals associated with stress such as cortisol, C-reactive protein, fibrinogen, low density lipoproteins, and blood pressure (Steptoe et al. 2001; Steptoe, Brydon, and Kunz-Ebrecht 2005; Seeman et al. 2008). This work is suggestive that stress-induced physiological responses may partly explain the health/SES gradient. As a result, we focus on outcomes that are precursors for later negative health events such as self-reported health, mental health status, as well as biomarkers that measure stress and other physiological characteristics.

Initially, we utilize data from the BRFSS, which is an annual, state-based telephone survey designed to measure the health and health habits of the US population. The survey is administered by individual states and data is then aggregated into a single annual file by the Centers for Disease Control (CDC). It is a very large annual survey with the survey size increasing from 102,263 in 1994 to 212,510 in 2001, and 414,509 observations in 2004. BRFSS is an excellent survey for our purposes because it has detailed demographic data, including the number of children in the household, plus a host of health outcomes and health habits, such as self-reported health status and the number of bad physical and mental health days in the past month.

The econometric model we utilize is similar to that employed by Hotz, Mullin, and Scholz (2006) and Adireksombat (2010) in their analysis of the 1993 EITC expansions on female labor supply. Specifically, as we note in Figure 3, the 1993 expansions increased in absolute and relative terms the size of the benefit for low income families with two or more children compared to families with one child. Therefore, if income is protective of health, we should find an increase in the health of families with two or more children over time relative to the same time series change for families with one child. To the extent that we are concerned about other policy changes or events impacting maternal health, families with more than one child should experience similar conditions over this time period to families with only one child.[6] This assumption is discussed in more detail below.

A key question within this research framework is how to restrict the sample to include people *likely* to be eligible for the EITC? Although the EITC is an income-based benefit, the literature summarized above indicates that there are important labor supply consequences of the program so an income-based criterion would select the sample based on an outcome that would potentially contaminate

[6] As we outline in Section V, we also estimate supplementary models comparing women with two or more children to childless women. While there is a bigger change in income between these two groups, it is important to note that, unlike our main results, this specification cannot control for contemporaneous events impacting women with children differentially from those without children.

Table 1—Earned Income Tax Receipt by Education and Number of Children, Mothers Age 21–40, Tax Years 1993–1995 and 1998–2001

	≤ High school education		College graduate	
	1 Child	2+ kids	1 Child	2+ kids
Panel A. Percent receiving the EITC				
Tax years 1993–1995	25.66	18.32	7.80	3.89
Tax years 1998–2001	31.97	27.71	7.94	4.86
Panel B. Size of EITC payment				
Tax years 1993–1995	$314.16	$271.02	$86.75	$46.31
Tax years 1998–2001	$487.49	$642.86	$99.94	$90.30
Panel C. Size of EITC payment among recipients				
Tax years 1993–1995	$1,224.43	$1,479.29	$1,091.15	$1,190.50
Tax years 1998–2001	$1,524.43	$2,320.21	$1,281.41	$1,857.04

Source: Current Population Survey 1994–1996 and 1999–2002

results due to a sample selection bias. A strategy used in the past is to select likely recipients by level of education and this is the method employed here.

The other consideration concerns the age range of the mothers in the sample. According to the enabling legislation, qualifying children for the EITC must be under age 19 or under 24 for full time students. It was not until 1993 that BRFSS first asked respondents to identify the number of children in the household less than 18 years of age, meaning this is the first pre-treatment year in the analysis sample.[7] As we increase the maximum age of the mothers in the sample, we increase the likelihood of including families that have potentially qualifying children older than 18 and hence misplacing mothers in one versus two or more children families. At the same time, reducing the maximum age eliminates women potentially "treated" by the EITC and increases the chance of a Type II error. To balance these two interests, we restrict the sample to women 21 to 40 years of age with reported children in the household.[8]

Table 1 reports data from the Annual Demographic file from the 1994–1996 and 1999–2002 March Current Population Survey (CPS) that reports the estimated percentage of women aged 21–40 who received the EITC, categorized by their education status and number of children, and pre- and post-1996 time periods.[9,10] The estimated amount of the credit received by each CPS respondent is generated by the United States Census Bureau tax model and the calculation assumes that all those eligible for the credit actually applied.[11] The results in Table 1 demonstrate that

[7] Prior to that year, the survey asks respondents the number of children in grades K–8 and the age of the youngest child, eliminating any pre-1993 surveys from use.

[8] This sample reduces the chance of having families with qualifying older children in the sample. In the 2000 census One-Percent Public Use Micro Sample, the fraction of mothers aged 21–40 with a high school degree or below in families with older qualifying children (e.g., children aged 19–24 and in school) was only 1.8 percent.

[9] The March CPS data was downloaded from www.ipums.org, King et al. (2010).

[10] Because the March CPS asks about income earned in the previous year, data from the 1994–1996 CPS presents data for the 1993–1995 tax years.

[11] This is an assumption that previous research has established is clearly wrong. Data from the 1996 tax year suggest that the between 12.8 and 17.8 of those eligible for the program never applied (Tax Policy Center 2002). At the same time, the Internal Revenue Service (2002) estimates that approximately 30 percent of the benefits paid out by the EITC in 2000 went to individuals who were not eligible for the benefit.

the probability of receiving any EITC benefit is decreasing in education, holding the number of children constant. Furthermore, the group that received the largest increase in benefits between the tax years 1993–1995 and the years 1998–2001 were women with two or more children and a high school diploma or less in education.

Table 1 also contains information about the estimated amount of the benefit received, again assuming all eligible women applied. The numbers in these tables are in nominal terms. As would be expected given the structure of the OBRA93 expansion, women with two or more children received much larger increases in their estimated EITC payment. For example, in the last two rows of Table 1, women with two or more children and a high school diploma or less experienced an increase in their estimated EITC payment of roughly \$840 (57 percent). On the other hand, women of a similar educational background but with only one child had an increase of only approximately \$300 (25 percent). The numbers in Table 1 indicate that among those with children, the most likely recipients of the EITC are the low-educated women and hence, these comprise the population eligible for the program in our econometric models. While the average dollar amount of benefits among all women may appear small, it is important to note that this average amount represents a mixture of people receiving large increases in annual income and those receiving relatively small changes in benefit levels. If there is a nonlinearity in the health benefits of income, this mixture of large and small beneficiaries could lead to larger health benefits than the average income change may imply. This is discussed in more detail below.

In order to estimate the simple difference-in-differences model outlined above, at a minimum, we need information on mother's age, education, and the number of children in the household. Because the first checks under the new EITC schedule for families with two or more children are distributed in 1996 (for tax year 1995),[12] we look at data from 1993 through 2001, giving us three years pre-EITC expansion and six years post.[13]

Sample means from the BRFSS dataset for the pre-EITC expansion period are reported in Table 2. In the first two columns, we report estimates for women, age 21–40 with a high school education or less with one and two plus kids respectively. In the next column, we report the p-value on the test of the null hypothesis that the means are the same across the two columns allowing for observations within

[12] The EITC expansions of 1993 were phased in over tax years 1994 and 1995. Because most recipients take the credit in the form of a tax rebate, the full effect of the expansions should be witnessed in calendar year 1996 which is how we have defined the treatment dummy above. However, there could be a concern that some of the women exposed to first portion of the EITC expansion in calendar year 1995 are incorrectly classified as being in the pre-treatment period. Even if this were the case, we should expect that this would bias the estimate against finding an effect on health outcomes. This can be seen in the data. If we change the treatment period to start in calendar year 1995 when rebates from tax year 1994 were distributed, our results are slightly stronger. The coefficients (standard errors) on the OLS regression-adjusted difference-in-differences estimates for the "at work" and "excellent/very good" linear probability estimates are 0.0189 (0.0075) and 0.0202 (0.009), which are similar to estimates in final column of Table 5. Likewise, the estimate on the treatment effect for the regression-adjusted negative binomial model for the number of bad mental health days is −0.0846 (0.040).

[13] The EITC was passed in 1993 and became effective in tax year 1995. The Advance EITC (AEITC) allows taxpayers to collect their EITC throughout the year in the form of lower tax withholdings in their paycheck (http://www.irs.gov/individuals/article/0,,id=96515,00.html). The US Government Accountability Office (2007), however, estimates that only 3 percent of eligible taxpayers in the 2002–2004 period collected the AEITC. Therefore, calendar year 1996 is the first year of increased benefits from the "1993 expansion."

Table 2—Sample Characteristics, Mothers Aged 21–40, 1993–1996 BRFSS

	≤ High school education			College graduates		
Variable	1 child	2+ kids	p-value	1 child	2+ kids	p-value
Average age	31.0	32.0	0.000	32.2	34.6	0.000
% currently employed	0.679	0.580	0.000	0.796	0.714	0.000
Race						
% white, non-Hispanic	0.761	0.713	0.000	0.789	0.846	0.000
% black, non-Hispanic	0.129	0.141	0.139	0.104	0.069	0.000
% Hispanic	0.076	0.101	0.006	0.044	0.036	0.070
% other race	0.033	0.044	0.036	0.063	0.049	0.002
Marital status						
% married	0.548	0.655	0.000	0.709	0.846	0.000
% separated/divorced/widowed	0.238	0.210	0.000	0.162	0.120	0.000
% never married	0.183	0.111	0.000	0.111	0.025	0.000
Family income						
% < \$20K	0.405	0.386	0.057	0.110	0.082	0.000
% ≥ \$20K, < \$50K	0.421	0.442	0.063	0.441	0.410	0.002
% ≥ \$50K	0.078	0.086	0.060	0.374	0.432	0.000
% income missing	0.096	0.088	0.016	0.075	0.076	0.961
Health outcome						
% excellent/very good health	0.582	0.577	0.446	0.805	0.809	0.580
% with any bad mental health days[a]	0.432	0.447	0.039	0.418	0.424	0.603
% with any bad physical days[a]	0.351	0.343	0.218	0.357	0.356	0.969
Number of bad mental health days[a]	4.27	4.52	0.030	2.93	2.89	0.753
Number of bad physical days[a]	2.81	2.65	0.072	1.89	1.95	0.656
Observations	7,315	15,737		3,881	6,740	

Notes: The p-value is for the test of the null hypothesis that the means across the samples are the same. The test is performed allowing for an arbitrary correlation for observations within a state.

[a]In the past 30 days.

a state to be correlated. In the final three columns of the table, we repeat the same basic structure but for mothers with a college degree.[14]

In the sample with high EITC eligibility, there are noticeable differences in the observed characteristics of the mothers with one versus two plus children. Women with two plus children tend to be slightly older, have higher fraction minority, are more likely to be married, and have lower incomes. Not surprisingly, women with more children are less attached to the labor force as well. Most of these differences are statistically significant. In the bottom of the table, we report sample means for the measures of health status and health habits. The first outcome is a dummy that equals 1 if a person self-reports they are in excellent or very good health.[15] The second and third variables are, respectively, the number of bad mental and physical health days reported in the past 30 days. Mothers with two or more children are less likely to report excellent or very good health and report more bad mental health days, but they report fewer bad physical days. Interestingly, unlike the demographic variables, there are much smaller differences in the health characteristics between

[14]We utilize this final group in a difference-in-difference-in-differences model and for completeness, report basic sample means for this group as well.

[15]The original question in the survey is the standard one where respondents report whether their current health is excellent, very good, good, fair, or poor.

women with only one child and those with two children. For women with a high school degree or less there are statistically significant but small differences in the number of bad mental health days in the past month (95 percent confidence level).

IV. Econometric Models

Our econometric model exploits the fact that after tax year 1995, low income mothers with two or more children received a substantial rise in income relative to similar women with only one child due to the EITC expansions. As we outline below, the model is a straightforward difference-in-differences (DD) specification. At its most basic level, the DD specification can be expressed as a comparison of pre-treatment and post-treatment means for the affected groups versus the comparison sample. We can enhance the explanatory power of this simple DD model by adding a set of covariates to control for individual characteristics, state level fixed effects, and time effects in the following equation:

$$(1) \qquad y_i = \alpha + Two_i \phi + \sum_{t=1993}^{2000} T(t)\pi_t + X_i \gamma + \sum_{m=1}^{50} S(m)\lambda_m + (Two_i Expand_i)\delta_{dd} + \varepsilon_i,$$

where y_i is the outcome of interest for person i; Two_i is an indicator variable for having two or more children; $T(t)$ equals 1 if an observation is from year t; $S(m)$ equals 1 if that observation is from state m; X_i is a set of explanatory variables; and $EXPAND_i$ is an indicator variable equal to 1 for the years after 1995. The final term ε_i is an idiosyncratic error, and the reduced-form impact of additional income generated by the EITC is captured by δ_{dd}. In our results, we call the estimates from the comparison of means as the "simple" DD estimates and the results from equation (1) as the "regression-adjusted" DD estimates. In these models we allow for an arbitrary correlation in errors for observations within a state over time.

As in any DD model, the key identifying assumption is that the trends in the comparison sample provide an estimate of the time path of outcomes that would have occurred in the treatment group had there been no intervention. We can never directly test this hypothesis but we can provide some evidence that the trends for these two groups were similar in the pre-treatment period. Specifically, we take model (1), restrict the sample to include data from the pre-treatment period only and allow the year effects to vary across mothers with one and two children. We can then test the null hypothesis that the year effects are the same across the two groups. Using BRFSS data, in the currently employed and excellent/very good self-reported health equations, the p-values on the test of the null hypothesis that the trends are the same across the two groups are 0.65 and 0.32 respectively.[16]

[16] One limitation of this test using the BRFSS data is that the pre-treatment period is relatively short. While there is little that we can do to overcome this limitation, we can supplement this evidence with a similar test using data from the National Health Interview Survey (NHIS). These data have the advantage of a longer pre-treatment time period, but are not appropriate for the full analysis because the NHIS was dramatically redesigned in 1996—the same year as the full implementation of the EITC expansion. We pool NHIS data for women with a high school

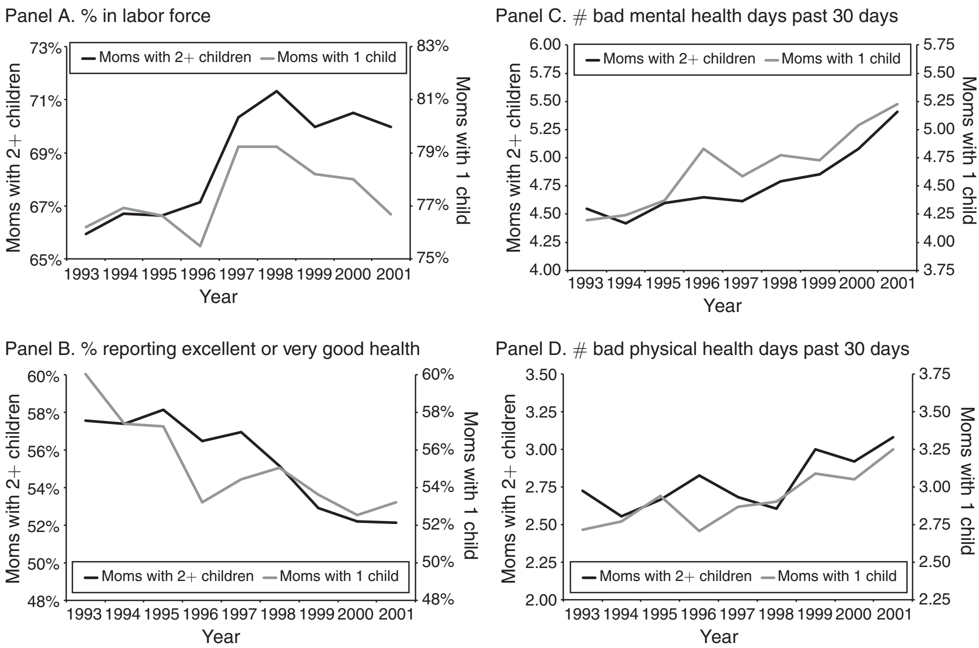

FIGURE 4. DIFFERENCE-IN-DIFFERENCES TIME SERIES GRAPHS BRFSS, 1993–2001

If there are unmeasured factors that differentially impacted low educated mothers with two kids compared to mothers with one child then the estimate δ_{dd} will be biased. We can potentially reduce this bias by increasing the dimensions of the problem and exploit data on a group of mothers with similar fertility experiences but not subject to the EITC shocks in a DDD framework. Specifically, noting the results in Table 1 that few college-educated mothers are EITC recipients, differential trends in health outcomes for college-educated mothers with two plus children versus one child can be used to control for parity-specific trends in the lower educated and higher EITC eligible populations. The DDD estimate is the parameter under the assumptions that the health status of mothers with a college degree has a similar pre-treatment trend as those for women with a high school degree or less and that this group will react similarly to postexpansion shocks, the DDD estimate will provide an unbiased estimate of the effect of the EITC on health outcomes. A tradeoff is that these models use less variation in the data to identify parameter estimates and as a result, standard errors tend to rise considerably.

V. Maternal Health Results from the BRFSS Samples

The basic DD results can be highlighted in a simple set of graphs that compare the treatment and control groups over time. Figure 4, panels A–D contain time series

degree or less and with at least one child from the 1988–1993 NHIS. Using this NHIS sample, the p-value on the same test implemented above for the BRFSS for a dependent variable of reporting excellent or very good health is 0.21.

graphs for four outcomes in the BRFSS data. In each graph the black line represents mothers with two or more children while the grey line presents mothers with one child. The left vertical axis always reports outcomes for the treatment group and the right axis is the scale for the comparison group. In all cases, both vertical axes have the same distance moving bottom to top. The dashed vertical line in indicates the last year of the pre-treatment period.

To provide a benchmarking comparison with previous research, in panel A we report the fraction of mothers who report they are at work. In the two pre-treatment years, outcomes appear to be trending similarly for both groups. However, after the OBRAA expansions, labor supply for both treatment and comparison groups increases with the change much smaller among the comparison sample, indicating the EITC expansions increased work among women with two or more children. This result is consistent with the consensus of the literature on the impacts of the EITC on labor supply.

In panel B, we report the fraction of mothers reporting excellent or very good health. In this case, immediately following the EITC expansions, we see a large decline in this outcome for mothers in the comparison sample with little change in the treatment group. This suggests some potential impact of the EITC expansions on this outcome but the impact vanishes towards the end of our sample. In panels C and D, we report the number of self-reported poor mental health days and poor physical health days, respectively. Panel C provides the cleanest graphical evidence of the impact of the EITC on maternal health in that there is a similar pre-treatment trend in outcomes for the treatment and comparison samples and after the EITC expansion mothers with two or more children have a lower number of poor mental health days compared to mothers with only one child. This difference persists over time.

The robustness of these graphical results can be tested using the simple and regression-adjusted DD methods described about. Table 3 contains these coefficients for different measures of self-reported health allowing for within-state correlation in errors. In this table, the numbers in parentheses are standard errors while the numbers in brackets are p-values on the null hypothesis that the parameter is zero.[17] The first row of results are for a dependent variable that equals one if an individual reports being at work. Similar to the previous literature on the EITC and the graphical results in Figure 4, we find a statistically significant 2.04-percentage-point increase in labor supply for mothers with two children compared to similar mothers with only one child. Using the same identification strategy with a sample of CPS data, Adireksombat (2010) found a of 5.2-percentage-point increase in labor supply.

The second row of results are for a dependent variable that equals one if an individual reports being in either excellent or very good health. The regression-adjusted

[17] In statistical tests for single parameters, we reject the null hypothesis if the p-value falls below a critical value of α. Because we are producing multiple tests from the same dataset, we report the p-value so that readers can use the Bonferroni correction for statistical tests of multiple parameters. Specifically, Bonferroni suggests that to reduce Type I error, with j tests from the sample, the confidence level for a statistical test should be expanded to be $1 - \alpha/j$ and nulls are rejected if the p-values are lower than α/j. In this case, with three health outcomes, critical p-values would be 0.0167 for a 95 percent confidence level and 0.033 for a 90 percent confidence level. With this correction, our estimates for the impact of EITC on bad mental health days are still statistically significant at the 90 percent confidence level. We should note that while this simple correction reduces the chance of a Type I error, it does greatly increase the probability of a Type II error.

Table 3—Difference-in-Differences OLS and Negative Binomial Estimates
Mothers Aged 21–40, 1993–2001 BRFSS

Outcome	Preexpansion mean of outcome for treatment group	Estimation method	Difference-in-differences estimates (82,907 observations): Simple	Regression adjusted
At work	0.580	OLS	0.0170	0.0204
			(0.0073)	(0.0074)
			[0.023]	[0.008]
Excellent/very good health?	0.577	OLS	0.0095	0.0135
			(0.0079)	(0.0075)
			[0.233]	[0.078]
Number of bad mental health days in past month	4.52	Negative binomial	−0.0474	−0.0754
			(0.0306)	(0.0328)
			[0.121]	[0.021]
Number of bad physical health days in past month	2.65	Negative binomial	0.0140	0.0105
			(0.0390)	(0.0390)
			[0.719]	[0.788]

Notes: Standard errors are reported in parentheses and *p*-values on the test of the null that the coefficient is zero are reported in square brackets. All standard errors allow for arbitrary correlations between observations within the same state. Other covariates in the difference-in-differences model include: Complete set of dummies for age, race, marital status, and number of children for the respondent, plus a complete set of dummies for the month of survey, year of survey, and state of residence.

coefficient suggests that the EITC increased the probability of women with a high school degree or less and with two or more children reporting these high levels of health by 1.35 percentage points (p-value 0.10). In the third and fourth rows of the table, we report estimates for a model where the outcomes of interest are the counts of poor mental and physical health days in the past 30 days, respectively. Since these data are nonzero integer counts, we estimate a negative binomial model which allows for over-dispersion in the dependent variable.[18] The third row contains the estimates for a negative binomial model with the number of bad mental days as the dependent variable. The regression-adjusted coefficient shows that following the expansion of the EITC, women with two or more children and a high school degree or less experienced a statistically significant (p-value <0.05) 7.5 percent reduction in the number of bad mental health days compared to similarly educated women with only one child. The final row contains a similar set of estimates for the presence of bad physical days. These results, however, are generally small, positive, and imprecisely estimated.

For the first three outcomes in Table 3, the p-values decline as we move from the simple to the regression-adjusted DD estimates. This is due primarily to an increase in size of the treatment effect coefficient rather than a change in the precision of the estimates. This indicates that the covariates matter in our analysis, which should not be surprising given that the summary statistics in Table 2 indicate that there are

[18] In the negative binomial model, the variance to mean ratio is defined as $1 + \delta$ where δ is the over dispersion parameter. If $\delta = 0$, the model collapses to a Poisson where the variance equals the mean. In the models in Table 3, we can easily reject the null that $\delta = 0$ which is no surprise given that the responses vary from 0 to 30.

systematic differences in the observed characteristics of mothers with two or more children and those with only one child.

Although the estimates in Table 3 are in most cases of marginal statistical significance, they are large responses to the expansion. From Table 1, we see that the EITC expansions increased the average payment to families of two or more kids by about $200 relative to change for families with one child. Among recipients, this increase was about $540. Focusing on the bad mental health days results from Table 3, the "treatment on the treated" effect obtained by dividing −0.075 by 200 which is −3.75E-4. This number suggests that increasing EITC payments by $500 would reduce the number of bad mental health days by 19 percent, which is a steep income/health gradient. Four points are worth mentioning. First, there is a standard error on this estimate with a *p*-value of 0.03 noting that we can reject the null that income has no impact on health. Focusing solely on the null hypothesis, the evidence is suggestive of a causal impact of evidence although it is important to note that the range of possible values is quite wide.

Second, although the treatment on the treated effect is large, the changes in lifestyle brought about by EITC payments are also potentially very large. Receiving a large EITC benefit may loosen a liquidity constraint regarding important expenditures that would not occur with a much smaller transfer. For example, previous analyses of the expenditure patterns of EITC recipients found increased consumption in categories such as housing, transportation, and durable goods. In a survey of EITC recipients, Smeeding, Phillips, and O'Connor (2000) found that over half of the respondents spent their refunds on goods to improve social mobility such as moving and transportation. In a more systematic analysis using data from the consumer expenditure survey, Patel (2012) found that the majority of increased durable goods expenditures (furniture, appliances, vehicles, and large home electronics) from the EITC occur during the first quarter—providing suggestive evidence that large EITC refund payments may slacken liquidity constraints for some families. Similar to Smeeding et al. (2000), this analysis found that EITC payments also increased housing expenditures. These housing expenditures increased in each quarter by roughly the same size and are therefore suggestive of these families consuming higher quality housing throughout the year. This could be in the form of a larger home or a neighborhood with better amenities. In other settings, families moving to a better quality neighborhood through the Moving-to-Opportunity program experienced increased mental health indicators and decreased biomarker evidence of diabetes (Ludwig et al. 2012). If these liquidity constraints for durable goods and housing are important, a combination of large and small beneficiaries that results in a relatively small average benefit might result in a greater change in health than the case where all beneficiaries receive the same relatively small benefit.

Third, while there may be remaining concerns that the average income benefits from the EITC expansion are not big enough to generate the large observed health changes, this particular policy change has been shown in other research to generate substantial health benefits. For example, using a similar identification strategy to this paper Hoynes, Miller, and Simon (2012) found that a $1,000 increase in the EITC reduced low birth weight rates by 6.7 to 10.8 percent. While our estimates are much larger than this number, our standard errors encompass these effects.

TABLE 4—ROBUSTNESS TESTS, WOMEN AGED 21–40, 1993–2001 BRFSS

Outcome	Method	DD results 82,907 observations	DD results state × year 82,907 observations	Two children versus no children 77,679 observations	DD results married 49,217 observations	DD results single 33,690 observations	DDD results 127,209 observations
At work	OLS	0.0204	0.0206	0.0635	0.0183	0.0457	0.0130
		(0.0074)	(0.0076)	(0.0097)	(0.0109)	(0.0110)	(0.0123)
		[0.008]	[0.009]	[<0.0001]	[0.099]	[<0.0001]	[0.298]
Excellent/ very good health?	OLS	0.0135	0.0119	0.0148	0.0211	0.0082	−0.0045
		(0.0075)	(0.0075)	(0.0100)	(0.0099)	(0.0127)	(0.0114)
		[0.078]	[0.118]	[0.146]	[0.021]	[0.521]	[0.695]
Number of bad mental health days in past 30 days	Negative binomial	−0.0754	−0.0615	−0.0195	−0.1027	−0.0514	−0.0622
		(0.0328)	(0.0326)	(0.0423)	(0.0519)	(0.0519)	(0.0729)
		[0.021]	[0.059]	[0.644]	[0.048]	[0.322]	[0.394]
Number of bad physical health days in past 30 days	Negative binomial	0.0105	0.0249	−0.05	0.0432	−0.0377	0.1226
		(0.0390)	(0.0391)	(0.0521)	(0.0508)	(0.0675)	(0.0911)
		[0.788]	[0.525]	[0.388]	[0.394]	[0.577]	[0.178]

Notes: Standard errors are reported in parentheses and p-values on the test of the null that the coefficient is zero are reported in square brackets. All standard errors allow for arbitrary correlations between observations within the same state. Other covariates in the difference-in-differences model include: complete set of dummies for age, race, marital status, and number of children for the respondent, plus a complete set of dummies for the month of survey, year of survey, and state of residence. Other covariates in the difference-in-difference-in-differences model include: Complete set of dummies for age, race, marital status, education, and number of children for the respondent; a complete set of dummies for the month of survey, year of survey, state of residence, plus interactions between the education and the year effects, the number of children and the year effect, the education and number of children effects.

Finally, in an analysis of a similar program in Canada, Milligan and Stabile (2011) also found large impacts of income transfers on the health of children. The authors find that a \$500 increase in benefits was associated with an approximately 11 percent reduction in a child's indirect aggression score. While this study primarily focuses on children's health, it does contain some estimates for maternal health. For example, the authors' instrumental variables estimates show that a \$500 increase in benefits was associated with a 12 percent decrease in a mother's depression score for women with a high school degree or less. Again, the standard errors of our maternal mental health results do not exclude an impact of this magnitude.

Table 4 contains the estimated coefficients for a number of alternative specifications. The first column reprints the regression-adjusted estimates from Table 3. The second column of results attempts to account for the potentially confounding effects of changes in other state based policies. For example, given our sample characteristics (low educated mothers), a large fraction in the sample are single mothers with low income and therefore, many will be eligible for welfare assistance. The 1990s witnessed tremendous changes in welfare policies such as the Personal Responsibility and Work Opportunity Reconciliation Act (PWRORA)[19] which placed time limits on welfare, instituted family caps on benefits, mandated work requirements, increased earnings limits, and provided more generous asset limits for eligibility (Meyer and Rosenbaum 2001; Blank 2002; Bitler, Gelbach, and Hoynes 2005). Welfare reform was accomplished piecemeal with many states adopting some of these policies prior to 1996 through waivers. Likewise, the PWRORA reforms

[19] Public Law 104-193.

were instituted in roughly half the states in 1997 and the other half in 1998. Using the same dataset as we use below, Bitler, Gelbach, and Hoynes 2005 found that welfare reform reduced insurance coverage, reduced preventive care but had no impact on self-reported health status or the number of poor physical or mental health days.

Changes in federal policies are a threat to our identification strategy if they *differentially* affect low-income families with two or more children compared to families with only one child. The variation in the implementation time of welfare reform across states could potentially contaminate our estimates. We guard against this by using low-educated moms with one child as a comparison sample. Welfare reform should in general impact low-income mothers with one and two children to similar degrees. Since events such as welfare reform vary across states and over time, we can capture their common impact by including state-specific year effects in the model. These results are contained in the second column of Table 4 and for most outcomes there is little change in the estimates. However, the inclusion of state-specific year effects increases the p-values on the treatment effect in the self-reported health status to 0.118. The negative binomial estimate for the total number of bad mental health days now has a p-value of 0.059.

The third column contains results for an alternative identification strategy that compares changes in health outcomes for women with two or more children compared to those with no children. This identification strategy has a particular set of benefits and costs. Certainly, the magnitude of the EITC expansion was largest between these two groups. However, the criterion for being an effective comparison group for a difference-in-differences analysis is the ability control for the time path of outcomes in the absence of an intervention. While women with two or more children in this age range are similar to women with only one child, this is not necessarily the case for women without any children. This is particularly a problem as it relates to the impact of other contemporaneous policy changes (such as the welfare reform policies discussed above) which impact women with two or more children differentially to those without any children. It is unclear if the costs of this strategy outweigh the benefits of the large income change. Therefore, these results are presented in the third column as a supplementary analysis to the main results. Using this alternate identification strategy the estimate for reporting excellent or very good health is very similar in magnitude to the main estimate in Table 3 but has a p-value of 0.14. In all cases, the standard errors on the treatment effect in models using childless women are larger than the corresponding values using mothers with one child as a control. However, the negative binomial estimate for the number of mental health days is much smaller in magnitude and not statistically significant at conventional levels. The estimated impact of the expansion on the number of bad physical days is large in magnitude by imprecisely estimated. Taken together these estimates provide suggestive support for our main results.

The fourth and fifth columns of results are for samples split by marital status. The negative binomial result for the total number of bad mental health days reported by married women is large and statistically significant at a p-value of 0.05. The estimated effect for single women is statistically insignificant but is negative and large in magnitude. Similarly, while the estimate on reported excellent or very good health is small in magnitude and statistically insignificant for single women, the result among married women is large (a 2.1-percentage-point increase) and has a small p-value. Across both columns, we cannot reject the null hypothesis that the single and married results are different in magnitude. The more precise estimates for the married sample may be caused by the much larger sample sizes for these women.[20]

The final column of Table 4 contains the estimates for the DDD identification strategy in equation (2). These results provide no statistically significant estimates—though this is not surprising. The basic results in the first column of Table 4 are of marginal statistical significance. Because the DDD models absorb additional dimension of the data, the model is using much smaller variation in the covariate of interest. Comparing the first (DD estimates) and last column (DDD estimates) of results in Table 4, the standard errors double in size. Holding the DD coefficient estimates constant, none of the parameters would be statistically significant at a p-value of 0.05 with the standard error estimates from the DDD models.

The estimates in Table 4 provide an important set of results for policy considerations. In contrast to nearly all of the previous sources of variation used to identify the effect of income on health, the EITC represents a feasible (if not the most feasible) means of distributing money to low-income Americans. Therefore, the totality of the effect of the program on health is important, even if some portion is a result of changes in labor-supply induced by the program.

Some concern may remain that changes in other government policies that occurred concurrent with the expansion are actually driving the estimates. For example, it is possible that the above results are driven by the start and rapid expansion of the State Child Health Insurance Program (SCHIP) in the late 1990s or the Medicaid expansion that began in the late 1980s. These programs would only contaminate our results if there was a *differential* change in insurance status for two-plus child families among low income women compared to single child families with similar incomes. Evans and Garthwaite (2010) used data from the March CPS and found that these was no differential effect on health insurance status from these government programs between families with one child and those with two or more children. This suggests that there is little credibility to the argument that the observed health improvements are a result of changes in health insurance status.[21]

[20] Previous work on the labor supply impacts of the EITC suggests that married women are less responsive than single women (Eissa and Hoynes 2004). However, it is important to consider that the theory predicts there should be positive labor supply effects for married women if their family income is in the phase-in range. Data from the 1994 March CPS indicates that roughly 12 percent of single married mothers with a high school degree or less are in families that are in the phase in range of the EITC for that year. In addition, 15 percent of married women have husbands with no labor earnings in the previous year and one third of married women have higher hourly wages than their husbands. It is also important to note that the incentive effects for labor supply should not be thought of as synonymous with the health impacts. Consider the case of a married women with two or more children whose family income is in the plateau range of the EITC. While she may not increase her labor supply as a result of the policy change, she still receives a positive income shock from the EITC expansion compared to a similar woman with only one child that could amount to 15 percent of her family income.

[21] Another federal policy that may be of concern is the Child Tax Credit. This credit, which went into effect in tax year 1998 (calendar year 1999), provided \$400 per child under the age of 17. Over time the credit grew and it currently provides \$1,000 per qualifying child. In the first three years of its existence the credit was not refundable. As of tax year 2001, families with children were eligible for the Additional Tax Credit which provided additional money to families with children that did not receive their full Child Tax Credit as a result of its nonrefundable nature. There are several reasons why the Child Tax Credit is not of great concern for this analysis. The first is that the Child Tax Credit is only in place for three years of the sample used for this analysis. Second, for the entire

VI. Maternal Health Results from the NHANES Samples

While the above results provide some evidence of the effect of higher transfer payments on health, the results are of marginal significance, all of the outcomes are self-reported, and all are subjective measures of health. Self-reported health is a relatively easy measure to collect. It is also an excellent predictor of objective measure of health such as mortality. In a review of 27 community studies, Idler and Benyami (1997) found that global self-reported health was an independent predictor of mortality, even when indicators of morbidity were included in the analysis. In a meta-analysis of 163 studies, DeSalvo et al. (2006) found similar results even after controlling for a variety of demographic factors and comorbidities. Similarly, Maddox and Douglass (1973) found that self-reported health status was a better predictor of future physician ratings than the reverse. This led the authors to claim that self-reported health data "clearly measure something more—and something less—than objective medical ratings."

However, the use of self-reported health as an outcome does have some drawbacks. The subjective nature of self-reported health survey questions lead to a lack of comparability across individuals which introduces classical measurement error into the model (Bound 1991). Because we use self-reported health as an outcome, this type of measurement error should primarily reduce precision which is costly in this case given the marginal statistical significance of our results from the BRFSS samples. In an attempt to overcome this measurement error, researchers have proposed using self-reported data regarding objective medical conditions as opposed to health status. These data, however, are also subject to measurement error. Baker, Stabile, and Deri (2004) analyzed a unique dataset that contained self-reports of disease presence and indicators of disease from insurance claims data. They found that these self-reported measures produced both false positive and negative indications of disease. Self-reported measures of health can also be subject to systematic measurement error. Using a self-reported measure of hypertension, Johnston, Propper, and Shields (2009) found no evidence of an income health gradient. When the authors used medically documented blood pressure readings for the same individuals, they found a large income-health gradient with respect to blood pressure.

A perhaps more important shortcoming of self-reported outcomes is that they are limited in their ability to provide information regarding the mechanism driving the observed increase in health. In this way, the BRFSS results should be seen as an important first step towards understanding the effect of the EITC on health. The second component of this analysis is to consider more detailed indicators of health.

BRFSS sample this tax credit is not refundable, limiting its availability to many low-income families. If you limit the sample to exclude the years where the nonrefundable credit was in existence, the DD estimates reporting excellent or very good health is 0.019 (0.0087), and the negative binomial estimate for the number of bad mental health days in the past month is −0.0806 (0.034). If anything, these estimates are stronger than the main results, showing that the Child Tax Credit is not a source of bias in these results.

Increasingly, researchers have turned their attention to biomarkers of physical and mental stress as indicators of health. Karlamangla, Gruenewald, and Seeman (2012) notes this movement has occurred for several reasons. First, individuals can experience significant reductions in health even without the presence of identifiable chronic conditions. Often, these decreases in health can be identified through the use of biomarkers even when specific diseases are not detectable. In addition, biomarkers have been found to be useful in predicting a wide variety of health outcomes. Finally, due to the fact that biomarkers precede the onset of major diseases they are believed to be more susceptible to external factors such as psychological stressors. In summary, biomarkers appear to be the ideal setting for comprehensively estimating the health effects of the EITC.

To obtain a better understanding of the mechanism underlying the identified changes in health we found in the previous section, we conduct a similar analysis using biomarker data. This data, obtained from several panels of the National Health and Nutrition Examination Survey (NHANES), directly confronts the two concerns about self-reported indicators discussed above. The biomarker data in the NHANES are measured by medical professionals—addressing any lingering concerns about relying on self-reported health outcomes. Biomarker data could also provide evidence about the causal pathways generating the previously documented relationship between socioeconomic status and health. This strategy does, however, come at a cost: the NHANES has much smaller sample sizes than other health datasets.

The NHANES is a national survey designed to measure the health and well-being of the American population. Dating back to the 1960s, the survey component of the NHANES contains data on demographic, socioeconomic, and health related issues. The unique aspect of NHANES is the examination component that is conducted in mobile examination centers staffed by medical professionals. The examination component provides detailed medical information including data from blood and urine tests, and medical exams.

The first three NHANES surveys were approximately 8–10 years apart. After NHANES III, which interviewed people from 1988–1994, surveys were fielded on two-year intervals but with smaller samples. Since NHANES III occurred during the pre-1993 expansion period, we pair this data with the first three samples from the new timing framework—the NHANES 1999–2000, NHANES 2001–2002, and NHANES 2003–2004. These four samples provide roughly equal samples sizes in the pre- and post-EITC expansion periods.

The econometric model outlined in Section V requires that we identify the number of EITC-eligible children in families. This is accomplished in different ways depending on the particular NHANES sample. In NHANES III, the sample respondent for the household was asked to identify the number of people in the family. We estimate the number of children as family size minus two for married heads of households and family size minus one for families with single mothers. We will overstate the number of qualified children if some of the children in the family are being claimed as a qualifying child by a noncustodial parent in another household or if some of the children are above the EITC-qualifying age. There is little we can do about the former situation, but by restricting the top end age range of mothers, we can eliminate counting "boomerang children" who do not qualify as an EITC-qualifying child because of their age. As with the BRFSS, we will restrict the sample to women aged 21–40.[22]

The final three NHANES surveys do not ask about family size, but rather, household size. In this instance we first eliminate all households where the woman reports zero live births in her lifetime since few women who never gave birth live in families with children from their spouse.[23] In these surveys, we estimate the number of children as two minus household size for married women and one minus household size for single mothers.[24, 25]

The NHANES has a wealth of information from physical, blood, and urine tests. Table 5 contains the definitions and sample means of the biomarkers we utilize from the NHANES datasets. We select as biomarkers those use by Seeman et al. (2008) and following these authors, we classify individuals based on whether they have risky levels of these biomarkers (e.g., high blood pressure, low levels of albumin), and we group the risky biomarkers into four groups: those that measure inflammation, cardiovascular conditions, metabolic disorders, and aggregate risks across all three groups.

The first two biomarkers are acute-phase proteins where concentration levels are altered in response to inflammation. For example, atherosclerosis (considered the main cause of coronary artery disease) is an inflammation process where fatty material collects on the walls of arteries. Acute-phase proteins are thought to be independent predictors of heart disease (Hansson 2005). The two acute-phase proteins we consider are C-reactive protein (CRP) and albumin.

CRP is produced by the liver and is only present in the blood when there is inflammation. It is measured as milligrams per deciliter of blood (mg/Dl). Because CRP is only produced during inflammation, medical researchers have investigated whether it is an independent predictor of coronary heart disease (Ridker 2003; Koenig et al. 1999). Owen et al. (2003) found elevated levels of CRP among lower employment classes in the Whitehall II survey while Alley et al. (2006) found higher levels of CRP in lower income groups. Respondents are defined to have risky CRP levels when concentrations are ≥ 0.3 mg/Dl.

[22] In the 2000 census One-Percent Public Use Micro Sample, the fraction of mothers aged 21–40 with a high school degree or lower in families with nonqualifying children (e.g., children aged 19–24 and not in school, or any child over the age of 24) was only 3 percent.

[23] Using data from the Fertility Supplement to the June 2000 CPS, only 6 percent of women aged 21–40 who have never had a live birth report they have their "own children" under the age of 18, a variable that measures not only biological children but step and foster children as well.

[24] Among families with children, the fraction of households with two or more children is very similar across the four surveys. In our sample, we find 77 percent have two or more children in the NHANES III survey and about 72 percent in the final three NHANES surveys.

[25] Although there may be concerns that this procedure has the potential to miscount thenumber of children in the household, the procedure appears to be very accurate. Using data from the 2000 1 percent PUMS (Ruggles et al. 2010), we generate a sample of mothers aged 21–40 with a high school degree or less. We compare the number of own children in the household (which may include stepchildren) which is generated from the detailed relationship codes in the census, with the number of children we calculate using the method employed for those reporting household size in the NHANES sample. Our estimate of whether the mother has one or two or more children in the family matches the number of children in 90 percent of the cases. Because we restrict the sample to only women who have had a birth, when our proposed method doesn't match we tend to overstate the number of children in the house. Therefore, in these cases we have too many single children families in our two plus child treated group, which should bias our estimates towards zero.

Table 5—Biomarkers for Mothers Aged 21–40 with a High School Education or Less, NHANES III, 1999/2000, 2001/2002, 2003/2004

Biomarker	Measured in:	Observations	Sample mean	Risky levels of biomarker	Percent with risky levels
Measures of inflammation					
C-reactive protein[a] (CRP)	mg/Dl	2,950	0.573	$\geq$ 0.3 mg/Dl	0.437
Albumin	g/Dl[b]	2,935	4.07	< 3.8 g/Dl	0.262
Number of risky inflammation conditions		2,934	0.699		
Any risky inflammation conditions		2,934	0.526		0.526
Measures of cardiovascular conditions					
Diastolic blood pressure	mmHg[c]	2.947	69.3	$\geq$ 140 mmHg	0.046
Systolic blood pressure	mmHg	2,952	112.2	$\geq$ 90 mmHg	0.035
Resting pulse	Beats/minute	3,090	74.97	$\geq$ 90 BPM	0.108
Number of risky cardiovascular conditions		2,947	0.184		
Any risky cardiovascular conditions		2,947	0.155		0.155
Measures of metabolic conditions					
Total cholesterol	mg/Dl[d]	2,949	189.95	$\geq$ 240 mg/Dl	0.102
High-density lipoproteins	mg/Dl	2,942	53.62	< 40 mg/Dl	0.156
Glycated hemoglobin	percent	2,992	5.2	$\geq$ 6.4 percent	0.026
Number of risky metabolic conditions		2,933	0.283		
Any risky metabolic conditions		2,933	0.259		0.259
Aggregate risks					
Number of risky conditions		2,683	1.156		
One or more risky conditions		2,683	0.657		0.657
Two or more risky conditions		2,683	0.333		0.333
Three or more risky conditions		2,683	0.127		0.127

[a]To make the datasets comparable over time, we censored the lower values of C-reactive protein at 0.21 in the final three NHANES samples.
[b]g/Dl = Grams per deciliter
[c]mmHg = Millimeters of mercury
[d]mg/Dl = Milligrams per deciliter

Albumin is a blood protein made by the liver and is measured as grams per deciliter (g/Dl). Albumin levels decline during inflammation (Gillium, Ingram, and Makuc 1994). Lower levels of albumin may indicate liver disease, and is predictive of coronary heart disease, cardiac events (Danesh et al. 1998), and stroke (Gabay and Kushner 1999). Risky albumin levels are defined as concentrations below 3.8 g/Dl (grams per deciliter). Seeman et al. (2008) found little correlation with low albumin levels and education but found risky albumin levels decline with income.[26]

Looking at the sample means in Table 5 for these inflammation biomarkers, roughly 44 percent of the mothers in our sample have elevated CRP levels while about a quarter have risky albumin levels. About 53 percent of women in the sample have at least one risky inflammation condition and the average number of risky inflammation conditions is about 0.7.

[26]High levels of albumin may also signal malnutrition, so a lower fraction of risky albumin could be due to either reduced inflammation or improved nutrition. We believe malnutrition is not a problem in our sample given the high obesity rates in this population. In the pre-EITC expansion period, roughly 30 percent of the women in our sample are obese and 70 percent are overweight. In contrast, there are only 6 percent of women in our sample during this time period that report a body mass index of 20 or under.

The second group of biomarkers measure cardiovascular conditions and we include three: diastolic blood pressure, systolic blood pressure, and resting pulse. Blood pressure is measured in millimeters of mercury (mmHg) while resting pulse is measured in beats per minute. High blood pressure is predictive of heart disease, heart failure, stroke, and kidney failure.[27] A detailed review by Colhoun, Hemingway, and Poulter (1998) noted that 30 years of research has found a consistent connection between low socioeconomic status and elevated blood pressure across several developed countries.

Elevated resting pulse rates are predictive of future coronary heart disease and other cardiovascular events (Gillum, Makuc, and Feldman 1991). Seeman et al. (2008) found a strong negative relationship between education, income, and elevated pulse rates. Respondents are defined to have risky blood pressure if the systolic levels are 90 and above or the diastolic levels are 140 and above. Likewise, a resting pulse rate of 90 beats or more per minute is considered risky. The sample means in Table 5 indicate that only about 4 percent of mothers have elevated blood pressure but roughly 11 percent have an elevated pulse rate. Approximately one in six mothers in our sample has at least one risky cardiovascular condition.

The third group of biomarkers indicates metabolic disorders and the conditions for this category include total cholesterol, the concentration of high-density lipoproteins (HDL), and the concentration of glycated hemoglobins. Total cholesterol and HDL are measured in mg/Dl. Total cholesterol levels of 240 mg/Dl and above and HDL levels below 40 are considered high risk and one in ten mothers have elevated cholesterol while one in 7 have elevated HDLs.

The third biomarker in this group is the level of glycated hemoglobin (HbA1c), which is a substance in red blood cells that is created when glucose attaches to hemoglobin (the protein in red blood cells that carries oxygen). This biomarker is measured as percent of the red blood cells that are composed of HbA1c and it is thought to be a better long-term measure of blood glucose than the point-in-time glucose tests done on a daily basis by diabetic patients. Elevated levels of HbA1c are associated with eye damage, kidney disease, heart disease, nerve damage, and stroke.[28] HbA1c levels have been found to be inversely associated with SES. Kelly, Stedman, and Leonardi-Bee (2000) used data from the NHANES 1999–2000 and found that HbA1c levels among nondiabetics were correlated with a variety of measures of SES. Concentrations of HbA1c of 6.4 percent or above are thought to be risky but only 2.6 percent of women have elevated levels of this biomarker. Although there are low levels of risky biomarkers for each of the elements in this group, the fraction of women in the sample with any risky cardiovascular biomarker is about 25 percent.

In the final group of biomarkers, we generate aggregate measures of risk by summing the number of risky conditions across all eight biomarkers. Aggregating the data in this manner is suggested by medical research which has shown that this count has more predictive capacity than the individual variables themselves. This is sometimes referred to as measure of "allostatic load" (McEwen and Stellar 1993;

[27] http://www.nhlbi.nih.gov/health/dci/Diseases/Hbp/HBP_WhatIs.html.
[28] http://www.nlm.nih.gov/medlineplus/ency/article/003640.htm.

McEwen 1998). Researchers have found that the stressors that accompany lower socioeconomic status are related to higher allostatic loads (Evans 2003). Research has demonstrated that those with low SES tend to have higher allostatic loads (Geronimus et al. 2006; Seeman et al. 2008) while Singer and Ryff (1999) found those with a history of low socioeconomic status had higher allostatic load levels in midlife. Crimmins, Kim, and Seeman (2009) found higher allostatic loads predicted a greater risk of mortality over a 6- to 12-year follow-up period while Karlamangla, Gruenewald, and Seeman (2012) found that all-cause mortality was monotonically increasing in an allostatic load measure containing nine biomarkers.

Medical studies show that unweighted count scores across a variety of biomarkers do a better job of predicting future outcomes such as mortality than any individual measure (Seeman, Singer, and Rowe 1997; Berenson et al. 1998). Therefore, we sum all eight risky biomarker measures into a composite score. In our sample, the average respondent has 1.2 risky conditions with this number ranging from 0 to 7. Two-thirds of women have at least one risky condition, a third have two or more, and an eighth have at least three conditions. Results for this composite measure of biomarkers provides the most complete picture of the health effects of the EITC expansion. The use of a composite measure of risk allows us to more effectively aggregate information into a single metric and hence increases the power of our test.

One benefit of using biomarkers is that they are reactive to current situations and predictive of future medical conditions. Given the age of the women in our sample, actual disease incidence rates will be low but as the results in Table 5 indicate, risky biomarker rates are high. Arguing for the use of biomarkers in health outcomes research, Karlamangla, Gruenewald, and Seeman (2012, 40) note that "… because changes in biomarkers precede the onset of clinically recognized diseases, they act as early indicators of pathology, and so are more sensitive to influences such as psychosocial stressors, health behaviors, and interventions." This is particularly true for the markers we use in this analysis compared to some others such as BMI or hip to waist ratio. Previous research has shown the markers in Table 5 are altered in the short term by stressful situations. Therefore, although the consequences of heightened blood pressure or LDLs may be years off the future, elevated biomarkers can be detected quickly after a health shock. For example, acute stress has been manipulated in experimental settings and in a short period after the experiments, subjects tend to have elevated measures of inflammation such as CRP (Steptoe, Hamer, and Chida 2007). Caregivers have been shown to have elevated measures of inflammation the day after providing care (Gouin et al. 2012). Average cholesterol levels are higher among medical students taking academic examinations (Grundy and Griffin 1959), among male accountants around urgent tax deadlines (Friedman et al. 1958), during periods of mental stress (Muldoon et al. 1995), and during periods of increased unemployment risk (Mattiasson et al. 1990). Studies indicate elevated blood pressure among medical students taking their final licensing exams (Zeller et al. 2004), among workers in high-stress jobs (Steptoe, Cropley, and Jokes 1999), and in families with increased financial strain (Steptoe, Brydon, and Kunz-Ebrecht 2005). Additionally, studies have connected reduced stress levels to decreases in blood pressure (Schneider et al. 2005). Research has also shown that changes in job stress can alter HbA1c levels in the blood (Netterstrøm et al. 1991; Kawakami et al. 2000).

Table 6—Regression-Adjusted DD and DDD Estimates for Effect of EITC Expansion on Allostatic Load for Women Aged 21–40, NHANES III, 1999/2000, 2001/2002, 2003/2004

Outcome	Preexpansion mean for treatment group	DD	DDD	Two children versus no children
One or more risky conditions	0.642	−0.091	−0.093	−0.165
		(0.050)	(0.072)	(0.076)
		[0.066]	[0.199]	[0.031]
Two or more risky conditions	0.308	−0.094	−0.054	−0.038
		(0.053)	(0.071)	(0.085)
		[0.074]	[0.448]	[0.658]
Three or more risky conditions	0.113	−0.061	−0.038	−0.0384
		(0.039)	(0.053)	(0.071)
		[0.118]	[0.507]	[0.587]
Poisson model: total number of risky conditions	1.108	−0.235	−0.207	−0.174
		(0.095)	(0.150)	(0.151)
		[0.014]	[0.169]	[0.249]

Notes: Standard errors are reported in parentheses and p-values on the test of the null that the coefficient is zero are reported in square brackets. All standard errors allow for arbitrary form of heteroskedasticity. Other covariates in the DD model include: complete set of dummies for age, race, marital status, and the year of survey. Other covariates in the DDD model include: complete set of dummies for age, race, marital status, education, plus interactions between the education and the year effects, the number of children and the year effect, the education and number of children effects.

In Table 6, we report estimates for difference-in-differences models of the effect of the EITC expansion on maternal health as measured by allostatic load—the measures of risky biomarkers thought by the medical community to be most predictive of negative health outcomes. The sample includes women aged 21–40 with a high school degree or lower. The covariates include dummies for the survey year plus the mother's age, race, marital status, and number of children. The treatment effect is captured by a simple interaction: respondents with two or more children in the final three NHANES surveys. In all models, we estimate standard errors that allow for an arbitrary form of heteroskedasticity across observations.

In the first three rows of the table, we report estimates from linear probability models where we estimate the impact of EITC expansion on having one or more, two or more, or three or more negative conditions. For the first two models, we estimate that the EITC expansion increased the probability of having one or more or two or more conditions by 9 percentage points, and both of these results have p-values less than 0.10. Moving to three or more conditions, the marginal effect declines to 6.1 percentage points (p-value of 0.118) but the impact as a percent of the baseline sample mean is very large (60 percent). In the fourth row of the table, we utilize the total number of counts as the dependent variable and estimate a simple Poisson model that explicitly accounts for the count nature of the data.[29] In this case,

[29] A potential concern with the Poisson model in this case is that the PDF is defined over all counts from 0 to infinity but by construction, our counts vary only from 0 to 8. We can easily adjust for this fact in any econometric estimation. If $f(y_i | x_i, \beta)$ is the PDF of the Poisson for person i and $F(8 | x_i, \beta)$ is the CDF evaluated from 0 to 8, the actual value of the likelihood for individual i is then $f(y_i | x_i, \beta)/F(8 | x_i, \beta)$. Programming a maximum likelihood version of this censored Poisson, the estimated coefficient on the EITC expansion variable and the standard error are unchanged out to three decimal places. Programming a maximum likelihood version of this censored Poisson model, we estimate a value of the EITC expansion coefficient (standard error) that equals −0.236 (0.098).

the coefficient on the EITC expansion suggests that counts of risky conditions are 23 percent lower for mothers who received the larger EITC payments. This estimate is statistically significant at a p-value of 0.015.[30] Similar to above, we supplement our main estimates with a difference-in-differences specification comparing mothers with two or more children to women with no children. These estimates are contained in the third column of Table 6 and are subject to the same caveats discussed above concerning the appropriateness of the identification strategy in the face of events such as welfare reform that likely had differential impact on the treatment and comparison group. The estimated impact of the expansion on reporting at least one risky biomarker is larger in magnitude using this identification strategy and is statistically significant at a p-value of 0.05. In contrast, the estimated impact on reporting two more conditions is smaller in magnitude than the main estimate and statistically insignificant at conventional levels. The Poisson estimate on the total number of conditions is large in magnitude but imprecisely estimated. These supplementary estimates generally support the main biomarker results that show improved a reduction in the presence of a risky biomarker following the EITC expansion.

The results in Table 6 suggest a large increase in the quality of the biomarkers for mothers impacted by the EITC expansions. In Table 7, we attempt to disaggregate the data and detect the source of this advantage by estimating results for particular metabolic, cardiovascular, and inflammation disorders in that order for the same sample as in Table 6. It is important to note that any one of these individual biomarkers is less indicative of health than the aggregate measures above, and given the small NHANES sample size these biomarker-level results are less precisely estimated. This is particularly true for the two or more children compared to childless women estimates in the third column which are generally statistically insignificant

We group results by metabolic, cardiovascular, and inflammation disorders. For each group, we estimate linear probability estimates for whether the respondent has a risky level of the individual biomarker, a linear probability for whether the person has any risky biomarker in the group, and a count-data model for total subgroup counts.

Among metabolic disorders, we find a persistent decline in risky biomarkers (cholesterol, HDL, and glycated hemoglobin) but in all cases, the standard errors are larger than the parameter estimates. The estimated effect for having any metabolic disorder is large but is statistically insignificant. Similarly, the estimated EITC treatment effect from a Poisson model with the outcome the number of metabolic disorders is large but statistically insignificant.

The second block of results in Table 7 contains estimates for the presence of cardiovascular disorders. The results suggest a 3.2-percentage-point decrease in the

[30] The Poisson model is restrictive in that the expected value of outcomes is equal to the variance. In many cases, data is subject to over-dispersion where the variance grows faster than the mean and when over-dispersion is present, imposing the Poisson distribution on the data will tend to bias standard error estimates down (Hausman, Hall, and Griliches 1984). In our sample, over-dispersion is not an issue since the maximum count value is 7. Estimating the model with a negative binomial model allows for a variance to mean ratio of $1 + \delta$ but if $\delta = 0$ the model collapses to the Poisson. In our case, when the model is estimated as a negative binomial, we estimate δ to be 0.054 with a standard error of 0.028 indicating some but very little over-dispersion. It is therefore no surprise that we estimate a value of the EITC expansion treatment to have a coefficient (standard error) of -0.234 (0.096) in the negative binomial model.

Table 7—Regression-Adjusted DD and DDD Estimates for Individual Biomarkers, Women Aged 21–40, NHANES III, 1999/2000, 2001/2002, 2003/2004

Outcome	Preexpansion mean for treatment group	DD	DDD	Two children versus no children
Panel A. Metabolic biomarkers				
Risky glycated hemoglobin	0.026	−0.004	−0.012	−0.05
		(0.013)	(0.019)	(0.034)
		[0.773]	[0.529]	[0.129]
Risky total cholesterol	0.102	−0.022	0.043	0.087
		(0.034)	(0.046)	(0.048)
		[0.520]	[0.351]	[0.071]
Risky HDL	0.156	−0.027	−0.044	−0.048
		(0.036)	(0.047)	(0.061)
		[0.453]	[0.348]	[0.428]
Any risky metabolic condition	0.251	−0.042	−0.007	0.035
		(0.045)	(0.078)	(0.073)
		[0.348]	[0.776]	[0.628]
Poisson model: number of risky metabolic conditions	0.277	−0.185	0.030	−0.025
		(0.177)	(0.276)	(0.29)
		[0.294]	[0.914]	[0.930]
Panel B. Cardiovascular biomarkers				
Risky diastolic blood pressure	0.045	−0.032	−0.030	0.010
		(0.017)	(0.026)	(0.036)
		[0.059]	[0.251]	[0.770]
Risky systolic blood pressure	0.035	0.004	−0.0005	0.009
		(0.014)	(0.023)	(0.027)
		[0.800]	[0.984]	[0.748]
Risky pulse	0.108	−0.016	−0.043	−0.002
		(0.037)	(0.049)	(0.062)
		[0.669]	[0.381]	[0.971]
Any risky cardiovascular condition	0.131	−0.034	−0.048	0.01
		(0.041)	(0.055)	(0.073)
		[0.407]	[0.392]	[0.891]
Poisson model: number of risky cardiovascular conditions	0.164	−0.317	−0.423	0.175
		(0.233)	(0.343)	(0.357)
		[0.174]	[0.217]	[0.624]
Panel C. Inflammation biomarkers				
Risky albumin	0.262	−0.088	−0.087	−0.126
		(0.045)	(0.063)	(0.0742)
		[0.052]	[0.167]	[0.089]
Risky c-reactive protein	0.437	−0.083	−0.012	−0.047
		(0.050)	(0.070)	(0.08)
		[0.098]	[0.778]	[0.555]
Any risky inflammatory condition	0.493	−0.096	−0.060	−0.094
		(0.050)	(0.071)	(0.078)
		[0.057]	[0.402]	[0.232]
Poisson model: number of risky inflammatory conditions	0.493	−0.217	−0.136	−0.225
		(0.099)	(0.159)	(0.157)
		[0.029]	[0.394]	[0.153]

Notes: Standard errors are reported in parentheses and p-values on the test of the null that the coefficient is zero are reported in square brackets. All standard errors allow for arbitrary form of heteroskedasticity. Other covariates in the DD model include: complete set of dummies for age, race, marital status, and the year of survey. Other covariates in the DDD model include: complete set of dummies for age, race, marital status, education, plus interactions between the education and the year effects, the number of children and the year effect, the education and number of children effects.

probability of reporting high diastolic blood pressure. This estimate is statistically significant at a p-value of 0.10. While no other results in this section are statistically significant, the coefficient for the Poisson model is large, negative, but with a large p-value.

The final block of results is for the presence of inflammation biomarkers. All of the results in this section are statistically significant at least with a p-value of 0.10. The estimates suggest that the expansion of the EITC decrease the probability of reporting risky levels of Albumin by 8.8 percentage points. The estimated effect on CRP is a decrease of 8.3 percentage points. The probability of reporting any risky inflammation biomarker falls by 9.6 percentage points and the preexpansion mean for this variable is about 50 percent. The Poisson model estimate for the number of inflammatory biomarkers shows that the EITC expansion is associated with a 21.7 percent decrease in the number of these biomarkers (p-value < 0.05). These results for inflammatory biomarkers are the most precisely estimated of the three subgroupings which is not surprising given the high incidence rate for these outcomes relative to the other biomarkers. As we note above, the medical literature has also found that inflammatory biomarkers are independently predictive of outcomes such as heart attacks, strokes, and mortality, so there are a vast array of physical insults that can be captured by these outcomes (Tracy et al. 1997; Tice et al. 2003; Ridker et al. 2002; Schmidt et al. 2002; Ridker 2003).

While there are a large number of estimates in Table 7, the analysis primarily considers the effect of the EITC on eight biomarkers using different specifications. Given that none of the linear probability estimates of the impact of the EITC on an individual risky biomarker are statistically significant at the 95 percent confidence level, any correction for multiple comparisons such as the Bonferroni adjustment will reduce the confidence in these results. There is however a striking persistence in the biomarker results for our main estimates. In seven of eight cases, risky levels declined for mothers in the treatment group. If obtaining a negative estimate in any regression is a Bernoulli draw with a probability of 0.5, the probability we would obtain seven or more negative coefficients is only 3.5 percent. In four of eight cases we obtained estimates that were statistically significant at the 10 percent level. The probability of obtaining four estimates of the same sign with this precision from eight trials is 0.04 percent. The persistence in these individual level results, even with a general lack of precision, helps explain why the results are so much more precise when we aggregate the biomarkers into allostatic load.

VII. Conclusion

One of the more promising avenues that can potentially explain the pathway linking SES and health involves stress. A large medical literature has demonstrated that those in poor economic conditions exhibit more stress and this manifests itself in physiological transformations in the body. Those with more stress tend to have higher pulse, higher blood pressure, higher cholesterol, and more inflammation—physiological conditions that are predictive of future disease incidence and mortality. The literature to date has primarily generated a number of robust correlations but this work has failed to provide convincing evidence that exogenously changing underlying economic conditions would alter markers of stress. In this paper, we exploit the OBRA93 expansions of the EITC that gave dramatically more money to families with two or more children compared to other families with one child to examine whether this change in income translates into better health. Utilizing self-reported data from the large sample of respondents to the BRFSS, we find that the expansion of the EITC decreased the number of reported bad mental health days for mothers with a high school degree or lower and two or more children compared to a similar woman with only one child. Suggestive evidence was also found that the increase in payments increased the probability of reporting excellent or very good health status. We also find strong evidence that the expansion of the EITC lowered the counts of the total number of risky biomarkers for women with two or more children and a high school degree or less compared to similar women with only one child. These effects were strongest for measures of inflammation and suggestive evidence was found for a decrease in women with risky levels of diastolic blood pressure.

This work also creates a new dimension to the understanding of the EITC and other income maintenance programs. While a vast literature has developed about this large program, its potential effect on health has gone relatively unnoticed. Given that the size of these programs results from an implicit and explicit discussion of costs and benefits, demonstrating a clear (and previously not discussed) set of benefits from the nation's largest anti-poverty program can lead to a more fruitful and concrete discussion about appropriate program size. This could lead to more optimal allocation of government resources. The results also highlight that from a statistical standpoint, there is tremendous amount that can be gained by aggregating many different biomarkers into omnibus measures of health. The literature on allostatic load has stressed the enhanced predictive power of aggregating multiple measures into one outcome rather than using any one measure in isolation. In much the same way, although there was a consistent pattern in results across most of the eight biomarkers used in this analysis, few were statistically significant. We did, however, obtain much more precise estimates of a reduction in aggregate poor health from the combined measures of risk than from any individual marker in particular.

REFERENCES

Adda, Jérôme, James Banks, and Hans-Martin von Gaudecker. 2009. "The Impact of Income Shocks on Health: Evidence from Cohort Data." *Journal of the European Economic Association* 7 (6): 1361–99.

Adireksombat, Kampon. 2010. "The Effects of the 1993 Earned Income Tax Credit Expansion on the Labor Supply of Unmarried Women." *Public Finance Review* 38 (1): 11–40.

Adler, Nancy E., and Katherine Newman. 2002. "Socioeconomic Disparities in Health: Pathways and Policies." *Health Affairs* 21 (2): 60–76.

Alley, Dawn E., Teresa E. Seeman, Jung Ki Kim, Arun Karlamangla, Peifeng Hu, and Eileen M. Crimmins. 2006. "Socioeconomic status and C-reactive protein levels in the US population: NHANES IV." *Brain, Behavior, and Immunity* 20 (5): 498–504.

Baker, Michael, Mark Stabile, and Catherine Deri. 2004. "What Do Self-Reported, Objective, Measures of Health Measure?" *Journal of Human Resources* 39 (4): 1067–93.

Banerjee, Abhijit, Esther Duflo, Gilles Postel-Vinay, and Timothy Watts. 2007. "Long-Run Health Impacts of Income Shocks: Wine and Phylloxera in 19th Century France." National Bureau of Economic Research (NBER) Working Paper 12895.

Barrow, Lisa, and Leslie McGranahan. 2000. "The Effects of the Earned Income Credit on the Seasonality of Household Expenditures." *National Tax Journal* 53 (4 Part 2): 1211–43.

Baughman, Reagan, and Stacy Dickert-Conlin. 2003. "Did Expanding the EITC Promote Motherhood?" *American Economic Review* 93 (2): 247–50.

Baughman, Reagan, and Stacy Dickert-Conlin. 2009. "The earned income tax credit and fertility." *Population Economics* 22 (3): 537–63.

Berenson, Gerald S., Sathanur R. Srinivasan, Weihang Bao, William P. Newman, Richard E. Tracy, and Wendy A. Wattigney. 1998. "Association Between Multiple Cardiovascular Risk Factors and Atherosclerosis in Children and Young Adults." *New England Journal of Medicine* 338 (23): 1650–56.

Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes. 2005. "Welfare Reform and Health." *Journal of Human Resources* 40 (2): 309–34.

Blank, Rebecca M. 2002. "Evaluating Welfare Reform in the United States." *Journal of Economic Literature* 40 (4): 1105–66.

Bound, John. 1991. "Self-Reported versus Objective Measures of Health in Retirement Models." *Journal of Human Resources* 26 (1): 106–38.

Card, David, Carlos Dobkin, and Nicole Maestas. 2009. "Does Medicare Save Lives?" *Quarterly Journal of Economics* 124 (2): 597–636.

Case, Anne. 2004. "Does Money Protect Health Status? Evidence from South African Pensions." In *Perspectives on the Economics of Aging*, edited by David A. Wise, 287–311. Chicago: University of Chicago Press.

Cherkas, Lynn, Abraham Aviv, Anna Valdes, Janice Hunkin, Jeffrey Gardner, Gabriela Surdulescu, Masayuki Kimura, and Tim Spector. 2006. "The effect of social status on biological aging as measured by white-blood-cell telomere length." *Aging Cell* 5 (5): 361–65.

Colhoun, Helen, Harry Hemingway, and Neil Poulter. 1998. "Socio-economic status and blood pressure: an overview analysis." *Journal of Human Hypertension* 12 (2): 91–110.

Crimmins, Eileen M., Jung K. Kim, and Teresa E. Seeman. 2009. "Poverty and Biological Risk: The Earlier 'Aging' of the Poor." *Journals of Gerontology*: *Biological Sciences and Medical Sciences* 64A (2): 286–92.

Currie, Janet, and Jonathan Gruber. 1996a. "Health Insurance Eligibility, Utilization of Medical Care, and Child Health." *Quarterly Journal of Economics* 111 (2): 431–66.

Currie, Janet, and Jonathan Gruber. 1996b. "Saving Babies: The Efficacy and Cost of Recent Changes in Medicaid Eligibility of Pregnant Women." *Journal of Political Economy* 104 (6): 1263–96.

Cutler, David M., and Adriana Lleras-Muney. 2008. "Education and Health: Evaluating Theories and Evidence." In *Making Americans Healthier: Social and Economic Policies as Health Policy*, edited by Robert F. Schoeni, James S. House, George Kaplan, and Harold Pollack, 29–60. New York: Russell Sage Foundation.

Danesh, John, Rory Collins, Paul Appleby, and Richard Peto. 1998. "Association of Fibrinogen, C-reactive Protein, Albumin, or Leukocyte Count With Coronary Heart Disease: Meta-analyses of Prospective Studies." *Journal of the American Medical Association* 279 (18): 1477–82.

Deaton, Angus. 2003. "Health, Inequality, and Economic Development." *Journal of Economic Literature* 41 (1): 113–58.

DeSalvo, Karen B., Nicole Bloser, Kristi Reynolds, Jiang He, and Paul Muntner. 2006. "Mortality Prediction with a Single General Self-Rated Health Question: A Meta-Analysis." *Journal of General Internal Medicine* 21 (3): 267–75.

Dickert, Stacy, Scott Houser, and John Karl Scholz. 1995. "The Earned Income Tax Credit and Transfer Programs: A Study of Labor Market and Program Participation." *Tax Policy and the Economy* 9 (1): 1–50.

Dickert-Conlin, Stacy, and Scott Houser. 2002. "EITC and Marriage." *National Tax Journal* 55 (1): 25–40.

Eissa, Nada, and Hilary Williamson Hoynes. 2004. "Taxes and the labor market participation of married couples: the earned income tax credit." *Journal of Public Economics* 88 (9–10): 1931–58.

Eissa, Nada, and Jeffrey B. Leibman. 1996. "Labor Supply Response to the Earned Income Tax Credit." *Quarterly Journal of Economics* 111 (2): 605–37.

Ellwood, David. 2000. "The Impact of the Earned Income Tax Credit and Social Policy Reforms on Work, Marriage, and Living Arrangements." *National Tax Journal* 53 (4 Part 2): 1063–1105.

Epel, Elissa S., Elizabeth H. Blackburn, Jue Lin, Firdaus S. Dhabhar, Nancy E. Adler, Jason D. Morrow, and Richard M. Cawthon. 2004. "Accelerated Telomere Shortening in Response to Life Stress." *Proceedings of the National Academy of Sciences* 101 (49): 17312–315.

Evans, Gary W. 2003. "A multimethodological analysis of cumulative risk and allostatic load among rural children." *Developmental Psychology* 39 (5): 924–33.

Evans, William N., and Craig L. Garthwaite. 2010. "Giving Mom a Break: The Impact of Higher EITC Payments on Maternal Health." National Bureau of Economic Research (NBER) Working Paper 16296.

Evans, William N., and Craig L. Garthwaite. 2014. "Giving Mom a Break: The Impact of Higher EITC Payments on Maternal Health: Dataset." *American Economic Journal: Economic Policy*. http://dx.doi.org/10.1257/pol.6.2.258.

Friedman, Meyer, Ray H. Rosenman, Vernice Carroll, and Russell J. Tat. 1958. "Changes in the Serum Cholesterol and Blood Clotting Time in Men Subjected to Cyclic Variation of Occupational Stress." *Circulation* 17 (5): 852–61.

Frijters, Paul, John P. Haisken-DeNew, and Michael A. Shields. 2005. "The causal effect of income on health: Evidence from Germany reunification." *Journal of Health Economics* 24 (5): 997–1017.

Gabay, Cem, and Irving Kushner. 1999. "Mechanics of Disease: Acute-Phase Proteins and Other Systemic Response to Inflammation." *New England Journal of Medicine* 340 (6): 448–54.

Geronimus, Arline T., Margaret Hicken, Dayna Keene, and John Bound. 2006. "'Weathering' and Age Patterns of Allostatic Load Scores Among Blacks and Whites in the United States." *American Journal of Public Health* 96 (5): 826–33.

Gillum, Richard F., Deborah D. Ingram, and Diane M. Makuc. 1994. "Relation between serum albumin concentration and stroke incidence and death: then NHANES I Epidemiological Follow-up Study." *American Journal of Epidemiology* 140 (10): 876–88.

Gillum, Richard F., Diane M. Makuc, and Jacob J. Feldman. 1991. "Pulse rate, coronary heart disease, and death: the NHANES I Epidemiological Follow-up Study." *American Heart Journal* 121 (1 Part 1): 172–77.

Goiun, Jean-Philippe, Ronald Glaser, William B. Malarkey, David Beversdorf, and Janice Kiecolt-Glaser. 2012. "Chronic Stress, Daily Stressors, and Circulating Inflammatory Markers." *Health Psychology* 31 (2): 264–68.

Grundy, Scott M., and A. Clark Griffin. 1959. "Relationship of Periodic Mental Stress to Serum Lipoprotein and Cholesterol Levels." *Journal of the American Medical Association* 171 (13): 1794–96.

Hansson, Göran K. 2005. "Inflammation, Atherosclerosis, and Coronary Artery Disease." *New England Journal of Medicine* 352 (16): 1685–95.

Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches. 1984. "Econometric Models for Count Data with an Application to Patents-R & D Relationship." *Econometrica* 52 (4): 909–38.

Hotz, V. Joseph, Charles H. Mullin, and John Karl Scholz. 2006. "Examining the Effect of the Earned Income Tax Credit on the Labor Market Participation of Families on Welfare." National Bureau of Economic Research (NBER) Working Paper 11968.

Hotz, V. Joseph, and John Karl Scholz. 2003. "The Earned Income Tax Credit." In *Means-Tested Transfer Programs in the United States*, edited by Robert A. Moffitt, 141–97. Chicago: University of Chicago Press.

Hoynes, Hilary W., Douglas L. Miller, and David Simon. 2012. "Income, the Earned Income Tax Credit, and Infant Health." National Bureau of Economic Research (NBER) Working Paper 18206.

Hoynes, Hilary W., Marianne E. Page, and Ann Huff Stevens. 2009. "Is a WIC Start a Better Start? Evaluation WIC's Impact on Infant Health Using Program Introduction." National Bureau of Economic Research (NBER) Working Paper 15589.

Idler, Ellen L., and Yael Benyamini. 1997. "Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies." *Journal of Health and Social Behavior* 38 (1): 21–37.

Internal Revenue Service. 2002. *Compliance Estimates for Earned Income Tax Credit Claimed on 1999 Returns*. U.S. Department of Treasury. Washington, DC, February.

Johnston, David W., Carol Propper, and Michael A. Shields. 2009. "Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient." *Journal of Health Economics* 28 (3): 540–52.

Karlamangla, Arun S., Tara L. Gruenewald, and Teresa E. Seeman. 2012. "Promise of Biomarkers in Assessing and Predicting Health." In *The Biological Consequences of Socioeconomic Inequalities*, edited by Barbara Wolfe, William Evans, and Teresa E. Seeman, 38–62. New York: Russel Sage.

Kawakami, N., K. Akachi, H. Shimizu, T. Haratani, F. Kobayashi, M. Ishizaki, T. Hayashi, et al. 2000. "Job strain, social support in the workplace, and haemoglobin A1c in Japanese men." *Occupational and Environmental Medicine* 57 (12): 805–09.

Keane, Michael, and Robert Moffitt. 1998. "A Structural Model of Multiple Welfare Program Participation and Labor Supply." *International Economic Review* 39 (3): 553–89.

Kelly, Shona J., Jonathan Stedman, and J. Leonardi-Bee. 2000. "Is hemoglobin A1c level associated with measures of socio-economic status in non-diabetics after controlling for known explanatory factors?" *Stress and Health* 21 (3): 185–92.

King, Miriam, Steven Ruggles, J. Trent Alexander, Sarah Flood, Katie Genadek, Matthew B. Schroeder, Brandon Trampe, and Rebecca Vick. 2010. *Integrated Public Use Microdata Series, Current Population Survey: Version 3.0*. [Machine-readable database]. Minneapolis: University of Minnesota.

Koenig, Wolfgang, Malte Sund, Margit Fröhlich, Hans-Günther Fischer, Hannelore Löwel, Angela Döring, Winston L. Hutchinson, and Mark B. Pepys. 1999. "C-Reactive Protein: A Sensitive Marker of Inflammation, Predicts Future Risk of Coronary Heart Disease in Initially Healthy Middle-Aged Men: Results from the MONICA Augsburg Cohort Study, 1984 to 1992." *Circulation* 99 (2): 237–42.

Lantz, Paula M., James S. House, James M. Lepkowski, David R. Williams, Richard P. Mero, and Jieming Chen. 1998. "Socioeconomic Factors, Health Behaviors, and Mortality." *Journal of the American Medical Association* 279 (21): 1703–08.

Liebman, Jeffrey B. 1998. "The Impact of the Earned Income Tax Credit on Incentives and Income Distribution." *Tax Policy and the Economy* 12 (1): 83–119.

Lindahl, Mikael. 2005. "Estimating the Effect of Income on Health and Mortality Using Lottery Prizes as an Exogenous Source of Variation in Income." *Journal of Human Resources* 40 (1): 144–68.

Ludwig, Jen, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, and Lisa Sanbonmatsu. 2012. "Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults." *Science* 337 (6101): 1505–10.

Mackenback, Johan, Caspar Looman, and Joost van der Meer. 1996. "Differences in the misreporting of chronic conditions, by level of education: the effect on inequalities in prevalence rates." *American Journal of Public Health* 86 (5): 706–11.

Maddox, George L., and Elizabeth B. Douglass. 1973. "Self-Assessment of Health: A Longitudinal Study of Elderly Subjects." *Journal of Health and Social Behavior* 14 (1): 87–93.

Mattiasson, Ingrid, Folke Lindgärde, Jan Åke Nilsson, and Töres Theorell. 1990. "Threat of Unemployment and Cardiovascular Risk Factors: Longitudinal Study of Quality of Sleep and Serum Cholesterol Concentrations in Men Threatened with Redundancy." *British Medical Journal* 301 (6750): 461–66.

McEwan, Bruce S. 1998. "Stress, Adaptation, and Disease: Allostasis and Alloctatic Load." *Annals of the New York Academy of Science* 840 (1): 33–44.

McEwan, Bruce S., and Eliot Stellar. 1993. "Stress and the Individual: Mechanisms Leading to Disease." *Archives of Internal Medicine* 153 (18): 2093–2101.

Meer, Jonathan, Douglas L. Miller, and Harvey S. Rosen. 2003. "Exploring the health–wealth nexus." *Journal of Health Economics* 22 (5): 713–30.

Meyer, Bruce D., and Dan T. Rosenbaum. 2001. "Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers." *Quarterly Journal of Economics* 116 (3): 1063–1114.

Milligan, Kevin, and Mark Stabile. 2011. "Do Child Tax Benefits Affect the Well-Being of Children? Evidence from Canadian Child Benefits Expansion." *American Economic Journal: Economic Policy* 3 (3): 175–205.

Muldoon, Matthew, Tracy Herbert, Stephen Patterson, Marina Kameneva, Robert Raible, and Stephen Manuck. 1995. "Effects of Acute Psychological Stress on Serum Lipid Levels, Hemoconcentration, and Blood Viscosity." *Archives of Internal Medicine* 155 (6): 615–20.

Netterstrøm, Bo, Tage Søndergard Kristensen, Mogens Trab Damsgaard, Ole Olsen, and Anette Sjøl. 1991. "Job strain and cardiovascular risk factors: a cross sectional study of employed Danish men and women." *British Journal of Industrial Medicine* 48 (10): 684–89.

Owen, Natalie, Terry Poulton, Frank C. Hay, Vidya Mohamed-Ali, and Andrew Steptoe. 2003. "Socioeconomic status, C-reactive protein, immune factors, and responses to acute mental stress." *Brain, Behavior, and Immunity* 17 (4): 286–95.

Patel, Ankur. 2012. "The Earned Income Tax Credit and Expenditures." Unpublished.

Rainforth, Maxwell V., Robert H. Schneider, Sanford I. Nidich, Carolyn Gaylord-King, John W. Salerno, and James W. Anderson. 2007. "Stress Reduction Programs in Patients with Elevated Blood Pressure: A Systematic Review and Meta-Analysis." *Current Hypertension Reports* 9 (6): 520–28.

Ridker, Paul M. 2003. "C-Reactive Protein: A Simple Test to Help Predict Risk of Heart Attack and Stroke." *Circulation* 108 (3): 81–85.

Ridker, Paul M., Nader Rifai, Lynda Rose, Julie E. Buring, and Nancy R. Cook. 2002. "Comparison of C-Reactive Protein and Low-Density Lipoprotein Cholesterol Levels in the Prediction of First Cardiovascular Events." *New England Journal of Medicine* 347 (20): 1557–65.

Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2010. *Integrated Public Use Microdata Series: Version 5.0.* [Machine-readable database]. Minneapolis: University of Minnesota.

Schmidt, Reinhold, Helena Schmidt, J. David Curb, Kamal Masaki, Lon R. White, and Lenore J. Launer. 2002. "Early inflammation and dementia: A 25-year follow-up of the Honolulu-Asia aging study." *Annals of Neurology* 52 (2): 168–74.

Schneider, Robert, Charles Alexander, Frank Staggers, Maxwell Rainforth, John Salerno, Arthur Hartz, Stephen Arndt, Vernon Barnes, and Sanford Nidich. 2005. "Long-Term Effects of Stress Reduction on Mortality in Persons $\geq$55 Years of Age With Systemic Hypertension." *American Journal of Cardiology* 95 (9): 1060–64.

Scholz, John Karl. 1994. "The Earned Income Tax Credit: Participation, Compliance, and Anti-Poverty Effectiveness." *National Tax Journal* 47 (1): 63–87.

Seeman, Teresa E., Sharon S. Merkin, Eileen Crimmins, Brandon Koretz, Susan Charette, and Arun Karlamangla. 2008. "Education, income and ethnic differences in cumulative biological risk profiles in a national sample of U.S. adults: NHANES III (1988–1994)." *Social Science and Medicine* 66 (1): 72–87.

Seeman, Teresa E., Burton H. Singer, John W. Rowe, Ralph I. Horowitz, and Bruce S. McEwen. 1997. "Price of Adaptation — Allostatic Load and Its Health Consequences." *Archives of Internal Medicine* 157 (19): 2259–68.

Singer, Burton, and Carol D. Ryff. 1999."Hierarchies of Life Histories and Associated Health Risks." *Annals of the New York Academy of Sciences* 896 (1): 96–115.

Smeeding, Timothy M., Katherin Ross Phillips, and Michael O'Connor. 2000. "The EITC: Expectation, Knowledge, Use, and Economic and Social Mobility." *National Tax Journal* 53 (4 Part 2): 1187–1210.

Smith, James P. 1999. "Healthy Bodies and Thick Wallets: The Dual Relation Between Health and Economic Status." *Journal of Economic Perspectives* 13 (2): 145–66.

Snyder, Stephen E., and William N. Evans. 2006. "The Impact of Income on Mortality: Evidence from the Social Security Notch." *Review of Economics and Statistics* 88 (3): 482–95.

Steptoe, Andrew, Lena Brydon, and Sabine Kunz-Ebrecht. 2005. "Changes in Financial Strain Over Three Years, Ambulatory Blood Pressure, and Cortisol Responses to Awakening." *Psychosomatic Medicine* 67 (2): 281–87.

Steptoe, Andrew, Mark Cropley, and Katherine Joekes. 1999. "Job strain, blood pressure, and response to uncontrollable stress." *Journal of Hypertension* 17 (2): 193–200.

Steptoe, Andrew, Mark Hamer, and Yoichi Chida. 2007. "The effects of acute psychological stress on circulating inflammatory factors in humans: A review and meta-analysis." *Brain, Behavior, and Immunity* 21 (7): 901–12.

Steptoe, Andrew, Gonneke Willemsen, Natalie Owen, Louise Flower, Vidya Mohamed-Ali. 2001. "Acute mental stress elicits delayed increase in circulatory inflammatory cytokine levels." *Clinical Science* 101 (2): 185–92.

Sterling, Peter, and Joseph Eyer. 1988. "Allostasis: A New Paradigm to Explain Arousal Pathology." In *Handbook of Life Stress, Cognition and Health*, edited by Shirley Fisher and James Reason, 629–49. Oxford, UK: John Wiley and Sons.

Tax Policy Center. 2002. "Participation in the Earned Income Tax Credit Program for Tax Year 1996." Unpublished.

Tice, Jeffrey A., Warren Browner, Russell P. Tracy, and Steven R. Cumming. 2003. "The relation of C-reactive protein levels to total and cardiovascular mortality in older US women." *American Journal of Medicine* 114 (3): 199–205.

Tracy, Russell, Rozenn Lemaitre, Bruce Psaty, Dianne Ives, Robert Evans, Mary Cushman, Elaine Meilahn, and Lewis Kuller. 1997. "Relationship of C-Reactive Protein to Risk of Cardiovascular Disease in the Elderly: Results from the Cardiovascular Health Study and the Rural Health Promotion Project." *Arteriosclerosis, Thrombosis, and Vascular Biology* 117 (6): 1121–27.

US Government Accountability Office. 2007. *Advanced Earned Income Tax Credit: Low Use and Small Dollars Impede IRS Efforts to Reduce High Noncompliance*. US Government Accountability Office (GAO) GAO-07-1110. Washington, DC, August.

Zeller, Andreas, Daniel Handschin, Niklaus Gyr, Benedict Martina, and Edouard Battegay. 2004. "Blood pressure and heart rate of students undergoing a medical licensing examination." *Blood Pressure* 13 (1): 20–24.