Explaining Income-Related Disparities in Pap Smear Utilization:

A regression-based decomposition analysis of differences in Pap smear utilization following implementation of the Affordable Care Act

A PLAN B PAPER

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Breanna J. Arndt

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Advisor: Pinar Karaca-Mandic

Committee:

Elizabeth Davis, Chair

Pinar Karaca-Mandic, Reader

Katy Kozhimannil, Reader

November 2017

Acknowledgments

I would like to thank my advisor, Pinar Karaca-Mandic, and committee members, Elizabeth Davis and Katy Kozhimannil, as well as the faculty and staff of the Department of Applied Economics of the University of Minnesota, for their support and guidance throughout the process of writing this paper and my time as a Master's student. I would also like to thank the Department of Applied Economics for awarding me funding in the form of Graduate Fellowships—including the Willis L. and Dorothy L. Peterson Graduate Fellowship—and Teaching Assistantships over the past two years.

Dedication

This paper is dedicated to my mother—who pursued her dream of becoming a nurse while caring for my five older siblings—and my father—who ensured thousands of students received the highest quality education possible as a teacher, coach, principal, and superintendent of schools. You have both inspired me to invest in my passions for health policy and teaching. Without your unwavering support, this paper and my academic achievements would not have been possible.

Table of Contents

List of Tables

List of Figures

I. Introduction

Despite a wealth of evidence showing the benefits of appropriate use of preventive health services, less than half of women in a 2010 survey had met their recommended use of preventive screenings and services (McMorrow, Kenney, and Goin, 2014; IOM, 2011). A key factor associated with reduced use of some preventive services is a lack of health insurance (Rezayatmand, Pavlova, and Groot, 2012). Therefore, improving access to health care and emphasizing prevention were timely focuses of the Affordable Care Act (ACA) of 2010. These goals were addressed by increasing health insurance coverage and coverage generosity, particularly through the possibility of states expanding their Medicaid program eligibility to adults up to 138% of the Federal Poverty Level (FPL); 32 states including the District of Columbia adopted Medicaid expansions as of January 2017 (Kaiser Family Foundation, 2017). Additionally, the ACA's Federal subsidies were available for those with incomes up to 400% of the FPL to purchase coverage in the health insurance exchanges (McMorrow, Kenney, and Goin, 2014).

Along with these changes in insurance coverage, the ACA also emphasized prevention and coverage generosity by requiring most non-grandfathered private plans to provide coverage of key recommended preventive health services rated "A" or "B" by the US Preventive Services Task Force (USPSTF, 2014) at no cost sharing. In addition to an annual well-woman visit which may include a pelvic examination, beginning August of 2012, one such preventive service includes receiving the Pap smear test, a cost-effective preventive service for identifying precancerous or abnormal cervical cells before they advance to a cancerous stage for women aged 21-65 years (HHS.gov, 2011; USPSTF, 2014; NIH, 2010; Chen, et al., 2011). While the zero cost sharing obligation, along with increased insurance coverage, is expected to encourage increased utilization of these preventive services, it is important to note the March 2012 USPSTF guideline change recommending the Pap test only every three years. Studying utilization of the Pap smear in the years following the ACA's implementation is important because these changes may lead to lower use of Pap testing for some women who previously used the service more frequently or increased use by others who had previously not had access to the preventive service. The role of insurance coverage in years prior to the ACA was found to be an important driver in explaining income-related gaps in preventive service utilization patterns, including the Pap smear (McMorrow, Kenney, and Goin, 2014).

In the years following ACA implementation, women's rate of uninsurance fell from 19% before the ACA to 12%; however, this differed by demographic groups (Jones and Sonfield, 2016). Low-income women from Medicaid expansion states experienced significant reduction in reported lack of insurance coverage (Jones and Sonfield, 2016). However, low-income women from states that did not adopt the Medicaid expansion or undocumented women may be left without assistance. In fact, the Kaiser Women's Health Survey found that 4 out of 10 low-income women reported being uninsured even in 2013, post ACA-implementation (Kaiser 2013).

Moreover, cost sharing is associated with lower rates of utilization of the Pap smear, an effect which is particularly significant for low-income women (Rezayatmand, Pavlova, and Groot, 2012). Moderate copayments in particular have also been shown to deter women from receiving the Pap test (IOM, 2011). There is evidence that only six out of ten women are aware that pap tests are to be covered without cost sharing, suggesting there may be women under-utilizing this service because of incomplete information regarding the cost they may have to pay for it (Salganicoff, Alina, et al., 2014). Furthermore, it has been shown that there are differences in utilization of this important test across demographic groups. Lower-income women were found to have utilized the Pap test at a rate statistically significantly lower than that of higher-income women in the years preceding the ACA (McMorrow, Kenney, and Goin, 2014). Thus, while there is evidence of improved rates of health insurance coverage and access to preventive services at no cost sharing for some, it is clear that these changes have not impacted women uniformly across demographic and socioeconomic groups.

In light of the ACA's elimination of cost sharing for the Pap smear and other clinical preventive women's health services, as well as the guidelines for Pap smear utilization changing in the years following the ACA implementation, this paper expanded upon the work done by McMorrow, Kenney, and Goin (2014) by providing more recent evidence regarding the gap in Pap testing among lower- and higher-income women. Specifically, this analysis used Blinder-Oaxaca regression-based decomposition to identify the roles and significance of insurance coverage and other factors in explaining income-related disparities in Pap smear utilization in the United States from 2011-2015.

II. Conceptual Framework

Characteristics included in this analysis are expected to impact a woman's demand for the preventive service of the Pap smear. Covariates were considered based upon those included in related literature (McMorrow, Kenney, and Goin, 2014; Shen and Long, 2006; Hargraves and Hadley, 2003; Han, et al. 2015). Women seek to maximize their utility subject to their budget constraints through deciding to use their time and resources to consume either health services or other goods and services. This paper, like the study by McMorrow and colleagues (2014) incorporated a human capital conceptual framework in which demand for medical care is derived from the demand for health (or the demand for the stock of health), which is expected to be impacted by health status, age, and education (McMorrow, Kenney, and Goin, 2014). In such an intertemporal model of consumption of health services within a human capital framework, a woman's stock of health can improve her productivity and the dynamics of demand for the service can depend on her current health status (stock of health) or change over the woman's life cycle (Jack, 1999). Therefore, the framework also recognizes the likely impact of an individual's rate of time preference and notes the distinction between an investment and consumption perspective when individuals demand preventive services. For example, older individuals might expect a shorter remaining lifespan and invest less in preventive services relative to treatment services. Healthier individuals may be more future-oriented than those in poor health status and invest more in preventive services, but individuals in poor health may consume more health services in general from a consumption perspective (McMorrow, Kenney, and Goin, 2014).

Informed by this framework, the covariates included in this study were enabling, predisposing, and need factors: income status (<400% FPL or ≥400%FPL), employment status, educational attainment (highest level achieved), health insurance coverage status, marital status, race and ethnicity, U.S. Census region of residence, U.S. citizenship status, self-reported health status, respondents' age at the time of the survey, and the survey year.

Health insurance coverage was expected to positively impact use of preventive services. Past research provides evidence that those gaining access to health insurance coverage are expected to increase their utilization of medical care (Finkelstein et al., 2012, Kolstad and Kowalski, 2012). Coverage may lead to price effects, in which demand for health services increases due to the reduction in the cost of care. Previous studies have also shown a positive impact of additional income on health care utilization (Barbaresco, et al., 2015; Acemoglu, et al, 2005).

Higher levels of educational attainment have been found to be positively associated with preventive service use, and were expected to be positively associated with Pap smear utilization in this analysis (McMorrow, Kenney, and Goin, 2014). Citizenship was expected to be positively associated with Pap smear receipt due to potential for greater access to care. The expected association between the survey year and Pap testing was ambiguous, due to the concurrent cost sharing elimination and recommended utilization guidelines changing.

III. Data

To analyze gaps in Pap smear receipt among lower- and higher-income adult women in the United States, this analysis utilized 2011-2015 National Health Interview Survey (NHIS) data. This data is collected from a cross-sectional household survey conducted in-person continuously throughout each year by the National Center for Health Statistics (NCHS). The data typically covers about 100,000 individuals from about 40,000 households with an annual response rate of about 80 percent of eligible households The survey is nationally representative of the United States non-institutionalized civilian population; it does not include Americans who are living in foreign countries or in long-term care facilities, on active military duty, or incarcerated. The data includes details on demographic characteristics, employment status, health care access and use, and insurance coverage status of respondents, for example (nhis.gov, n.d.). Additionally, for observations missing family income and personal earnings data, NHIS provides imputed values generated through multiple imputation methods. Missing race and ethnicity data are also imputed using “hotdeck” imputation. The NHIS data utilized in this study was obtained from a harmonized version—IPUMS Health Surveys—provided by the Minnesota Population Center (Blewett, et al., 2016).

The outcome (dependent) variable used in this study as a measure of women's preventive care involves whether women respondents had a Pap smear within the past year. This binary variable was derived from the survey question included in the sample adult questionnaire, which asked sample adult women respondents aged 18 years or older whether or not they had a Pap smear within the preceding twelve months. The interviewer reminded respondents about what the test consists of, in the event they were unsure. Responses of "refused to answer" and "don't know" were recoded as missing. A detailed inclusion diagram for the sample size of the analytic sample is shown in Figure 1. Since this question is only asked of sample adult women, the universe for this variable is limited to a subset of women 18 and older in the sample, rather than all women respondents over 18 years. Because current guidelines recommend the women aged 21-65 years receive the Pap smear, the analytic sample was restricted to sample adult women aged 21-65 years (USPSTF, 2014). The analytic sample was further restricted to the 60,800 observations without missing or unknown values for the outcome variable or covariates used in the analysis through Stata survey subpopulation estimation using *svy* commands.

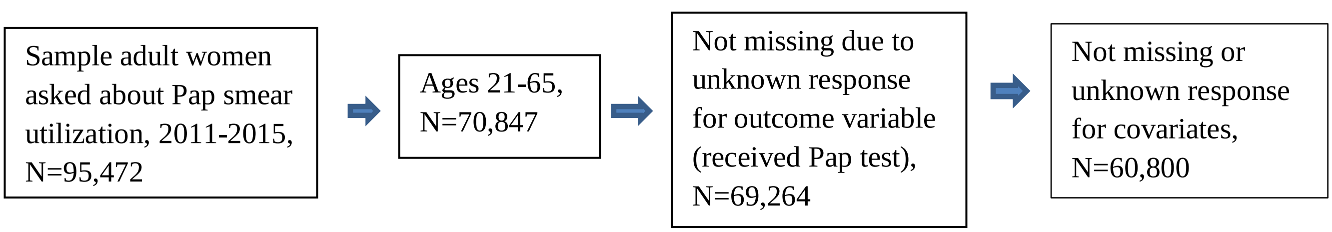


Figure 1. Analytic sample inclusion diagram.

Although the March 2012 US Preventive Services Task Force guidelines recommend women in this age range receive this test every three years, the recall period for this variable is the past 12 months prior to the survey (USPSTF, 2014). This complicates assessing whether women are complying with this guideline (or under or over using the service). However, this variable was still useful in examining differences in utilization across income groups for the purposes of this analysis.

IV. Empirical Analysis Background

Blinder-Oaxaca Regression Decomposition

Blinder-Oaxaca regression-based decomposition methods are useful in studying differences in an outcome by groups. The overall difference can be parsed into two parts: the explained and the unexplained portions. The explained portion is attributable to differences in the means of observed characteristics in the two populations (insurance coverage or race, for example). The explained portion can be further broken down into the parts explained by each of the observed population characteristics. The unexplained portion is not attributable to differences in observed characteristics, but rather due to either the returns to the population characteristics (i.e. estimated coefficients—how effective insurance coverage is at encouraging preventive service use, for example) or caused by unobserved characteristics (i.e. discrimination, culture, attitudes, etc.). It is important to note that unobserved characteristics might also include potentially measurable characteristics which are not available in a given analytic data set (i.e. distance to the nearest health facility or the languages spoken there, etc.).

Blinder-Oaxaca decomposition is based on linear probability model regressions stratified by group and relies on the key property of linear regression that the mean of the dependent variable is equal to the sum of the mean values of the independent variables multiplied by their estimated coefficients (O'Donnell, et al., 2008):

$$\bar{Y} = \bar{X}\hat{\beta} \qquad (1)$$

Considering the example of differences in Pap smear use by higher- and lower-income populations of women explored in this paper, the means of Pap smear utilization for higher- and lower-income groups are:

$$\bar{Y}_h = \bar{X}_h \hat{\beta}_h \text{ and } \bar{Y}_l = \bar{X}_l \hat{\beta}_l, \quad (2)$$

respectively. The difference in mean outcomes--receipt of the Pap smear--among these groups is represented by:

$$\bar{Y}_h - \bar{Y}_l = \bar{X}_h \hat{\beta}_h - \bar{X}_l \hat{\beta}_l \quad (3)$$

Adding and subtracting $\bar{X}_h \hat{\beta}_l$ and rearranging equation (3) gives the following:

$$\bar{Y}_h - \bar{Y}_l = \hat{\beta}_l (\bar{X}_h - \bar{X}_l) + \bar{X}_h (\hat{\beta}_h - \hat{\beta}_l) \quad (4)$$

The portion of the difference in Pap smear use between lower- and higher-income women that can be explained by differences in their mean observable characteristics is represented by the term,

$$\hat{\beta}_l (\bar{X}_h - \bar{X}_l), \quad (5)$$

which reflects the expected change in Pap smear use for the lower-income population if they had the characteristics of the higher-income population. If lower-income women have lower average education and insurance coverage, for example, and it is assumed that these characteristics are associated with greater barriers to receiving care, they would likely have more problems with access to utilizing the pap smear (Hargraves and Hadley, 2003).

The portion of the difference that cannot be explained by differences in observed characteristics is represented by the term,

$$\bar{X}_h (\hat{\beta}_h - \hat{\beta}_l), \quad (6)$$

which reflects either differences attributable to varying effects of the observable characteristics or differences caused by unobservable characteristics. Term (6) can also be thought of as the differences in returns to characteristics evaluated at the means of the higher-income population's characteristics. Essentially, this simulates a model with all individuals having the characteristics of the average higher-income woman and assesses if there would be a difference in the returns to those same characteristics for lower-income women (Hargraves and Hadley, 2003). In other words, this portion of the overall difference represents both differences in how the observed factors affect a woman's probability of using the Pap test and other unobservable factors not captured by the model (Shen and Long, 2006). With this specification, often used in previous literature, differences in mean population characteristics are weighted by coefficients of the lower-income group (McMorrow, Kenney, and Goin, 2014; Shen and Long, 2006; Hargraves and Hadley, 2003). Differences in coefficients are weighted by the mean population characteristics of the higher-income group.

However, it is equally valid to decompose equation (3) by instead adding and subtracting $\bar{X}_l\hat{\beta}_h$ and rearranging to give the following:

$$\bar{Y}_h - \bar{Y}_l = \hat{\beta}_h(\bar{X}_h - \bar{X}_l) + \bar{X}_l(\hat{\beta}_h - \hat{\beta}_l) \quad (7)$$

where differences in mean population characteristics are weighted by coefficients of the higher-income group, and differences in coefficients are weighted by the mean population characteristics of the lower-income group. The presumption with this specification is that higher-income women receive the Pap smear according to their characteristics (need, predisposition, and enabling factors), whereas lower-income women might be discriminated against or face unduly barriers to access, for example. This might be thought of as discrimination against lower-income women, assuming there are no relevant unobserved predictors, which is a strong assumption.

Conversely, the presumption with specification (4) is that lower-income women use the Pap smear due to their characteristics giving them access to or preference/need for the service, and higher-income women have unduly generous access to the service (or knowledge of it, etc.). This might be thought of as discrimination in favor of higher-income women, again making the strong assumption that there are no relevant unobserved predictors.

O'Donnell, et al. (2008) showed that the specifications in equations (4) and (7) are both special cases of the more general, threefold decomposition. This form includes an interaction term between observed characteristics, endowments (E), and coefficients (C) to account for the fact that income-related gaps in observed characteristics and their coefficients occur simultaneously (Jann, 2008). This specification is represented by

$$\overline{Y_h} - \overline{Y_l} = \hat{\beta}_l(\bar{X}_h - \bar{X}_l) + \bar{X}_l\left(\hat{\beta}_h - \hat{\beta}_l\right) + (\bar{X}_h - \bar{X}_l)\left(\hat{\beta}_h - \hat{\beta}_l\right), \qquad (8)$$

which can also be represented as

$$\overline{Y_h} - \overline{Y_l} = E + C + C * E \qquad (9)$$

Essentially, the form shown in (4) assumes the interaction is placed in the unexplained portion of the income-related disparity in Pap smear use,

$$\overline{Y_h} - \overline{Y_l} = \hat{\beta}_l(\bar{X}_h - \bar{X}_l) + \bar{X}_h\left(\hat{\beta}_h - \hat{\beta}_l\right) = E + (C * E + C), \qquad (10)$$

whereas the form shown in (7) assumes the interaction is placed in the explained portion,

$$\overline{Y_h} - \overline{Y_l} = \hat{\beta}_h(\bar{X}_h - \bar{X}_l) + \bar{X}_l(\hat{\beta}_h - \hat{\beta}_l) = (E + C * E) + C. \qquad (11)$$

The assumptions made by specifications (4) and (7) previously described may lead to undervaluation of one group and overvaluation of the other (Jann, 2008). Additional specifications discussed in Jann's (2008) paper, address this issue. For example, one such specification proposed by Neumark (1988),

$$\overline{Y_h} - \overline{Y_l} = \Delta\hat{\beta}_{Pooled} + [\bar{X}_l(\hat{\beta}_h - \hat{\beta}_{Pooled}) + \bar{X}_h(\hat{\beta}_{Pooled} - \hat{\beta}_l)], \quad (12)$$

uses coefficients estimated from a regression including the full sample of data pooling observations from both groups, rather than using just one set of coefficients estimated from one of the stratified regressions.

<u>Related Decomposition Literature</u>

Originally applied to study wage gaps by race and gender, Blinder-Oaxaca regression based decomposition has been used in several analyses exploring disparities in health care access, utilization, and outcomes (Blinder, 1973; Oaxaca, 1973). These studies employed Blinder-Oaxaca linear regression decomposition techniques to examine differences across groups in binary outcomes [Shen and Long, 2006; Hargraves and Hadley, 2003; McMorrow, Kenney, and Goin, 2014]. Past studies have employed Blinder-Oaxaca linear regression decomposition techniques to examine differences across groups in binary outcomes [Shen and Long, 2006; Hargraves and Hadley, 2003; McMorrow, Kenney, and Goin, 2014]. This method estimates separate linear probability models for each group, which may yield predictive probabilities out of the 0 to 1 range for binary dependent variables. However, they are useful to estimate consistent coefficient estimates. Another key property of linear regression useful for decomposition is that the mean of the dependent variable is equal to the sum of the mean values of the independent variables multiplied by their estimated coefficients.

Shen and Long (2006) explored changes in employer sponsored health insurance offers and take up of offers from 1999 to 2002. They estimated linear probability models of the probability of having an offer using two stage least squares regression on the 2002 sample separately for low- and middle-income workers. Hargraves and Hadley (2003) examined differences in access to care and health service utilization among racial/ethnic groups using Blinder-Oaxaca regression decomposition. They estimated linear models using Ordinary Least Squares. O'Donnell et al. (2008) decomposed differences in child malnutrition among the poor and non-poor in Vietnam using Blinder-Oaxaca linear decomposition, and McMorrow and colleagues (2014) also implemented Oaxaca-Blinder decomposition methods using linear probability models for lower- and higher-income groups to explore gaps in preventive service utilization.

Bustamante, et al. (2009) studied the disparities in healthcare access and utilization among Latino adults of Mexican and non-Mexican ancestry in the United States. Bustamante, et al. (2012) decomposed differences in healthcare utilization and access among Mexican immigrants by documentation status. These analyses used non-linear decomposition methods described in Fairlie (2014), who expanded upon the application of Blinder-Oaxaca linear decomposition techniques to incorporate coefficients from non-linear models, such as logit or probit models, and to produce the appropriate standard errors.

V. Methods

Statistical Analysis:

Pearson chi-square test (linearized standard errors of weighted column proportions were produced) comparing rates of Pap testing within the past 12 months among lower- and higher-income women aged 21-65. Similarly, Pearson chi-square tests were conducted to compare rates of each covariate among lower- and higher-income women. Stata version 12 statistical software was utilized to employ Taylor-series linearization techniques to calculate appropriate standard errors and confidence intervals, used in significance tests.

Decomposition:

To decompose the difference in rate of Pap testing by income group into the portion attributable to group differences in observed characteristics and the portion unexplained by observed characteristics, the Oaxaca-Blinder estimation and standard error calculation methods outlined by Jann (2008) were used. This method was performed using the user-written *oaxaca* command (Jann, 2008). This method estimated survey linear probability models stratified by income group (<400% FPL and ≥400% FPL) and accounted for the complex survey design of the NHIS databy employing the delta method to calculate linearized standard errors (Jann, 2008). For consistency with methods employed in related literature decomposing preventive service use disparities by income, the decomposition in this analysis assumed the previously described specification (4).

Robustness Checks:

As described above, it has been noted in previous literature that decomposition results may vary based on the underlying assumptions associated with the specification used, potentially leading to undervaluation of one group and overvaluation of the other (Jann, 2008). Thus, alternative decomposition results were estimated to compare results from this paper's main model specification (4) to the results when specifications (7)—using the higher-income regression coefficients–and (12)—using pooled higher- and lower-income regression coefficients–are estimated. Lastly, decomposition results were estimated using stratified survey logistic regression coefficients and, like specification (4), weighting differences in mean population characteristics by coefficients of the lower-income group and differences in coefficients by the mean population characteristics of the higher-income group. This method was performed using the user-written *nldecompose* command in Stata 12 using methods outlined by Bauer and Sinning (2008) and producing bootstrapped standard errors also accounting for the survey design of the data source (Bauer, Hahn, and Sinning, 2008).

Key Variable Definitions

The sample is split into higher-income and lower-income groups based on respondents' reported total family income compared to the U.S. Census Bureau's poverty thresholds (based on income, family size, and the number of children under age 18) for the year the respondent was asked about. As was done in the previous study by McMorrow, Kenney, and Goin (2014), which examined differences in preventive services across income groups prior to the Affordable Care Act, this analysis defined higher-income status as having reported family income at or above 400% of the Federal Poverty Level (FPL) and lower-income status as having reported family income less than 400% of the FPL. This definition is based on the Affordable Care Act's (ACA) optional Medicaid expansions being available for most adults with incomes less than 138% FPL in 32 states, and the ACA's federal subsidies being available for those with incomes up to 400% of the FPL to purchase coverage in the health insurance exchanges (Kaiser Family Foundation, 2017; McMorrow, Kenney, and Goin, 2014).

A woman was considered insured if she had any health insurance coverage at the time of the survey. Respondent race/ethnicity was recoded using an IHIS source variable which was categorized using the post- 1997 OMB classifications of race/ethnicity, and responses of "race category not releasable" were recoded as missing. Women were considered employed if they reported working for pay in the last year or if they did not answer the question about working for pay in the past year and were imputed the value "employed" by IPUMS staff. Table A6 provides detailed information regarding the survey questions used to obtain the original source variables and how original variables were recoded for this analysis.

VI. Results

Statistical Analysis

The 2011-2015 estimates show the rate of Pap smear receipt varied significantly by income (Table 1). While 65.4% of women at or above the 400% FPL used the service, only 52.1% of women below the 400% FPL utilized the Pap smear. This 13.4 percentage point difference in utilization was statistically significant at the 1% level.

In addition, the higher- and lower-income populations of women in the sample varied significantly for many observed characteristics (Table 2). Higher-income women reported a statistically significantly higher rate of having any form of health insurance coverage compared to lower-income women at the 1% level (96.4% and 76.1% insured, respectively). Older women (aged 40-65) were more likely to report income levels at or above 400%FPL than younger women (21-39 years), significant at the 1% level. The higher-income group of women were less likely to be Black, non-Hispanic or Hispanic and more likely to be US citizens than lower-income women. The higher-income group was more likely to attain higher levels of education than the lower-income group of women; lower-income women reported a 37.1 percentage point lower rate of having attained a Bachelor's degree or higher. The higher-income group of women was more likely to report being married and to report being employed, both significant at the 1% level. Lastly, higher-income women were less likely to report poor health status (12.3 percentage points lower, $p<0.01$).

Table 1. Receipt of the Pap Smear[1] for Adult Women (Aged 21–65 Years) by Income[2]: 2011–2015 National Health Interview Survey, United States

	All Incomes		<400% of the FPL		≥400% of the FPL		Percentage Point Difference
	Percent (unweighted n)	SE	Percent (unweighted n)	SE	Percent (unweighted n)	SE	
Utilized Pap Smear	57.0 (33,995)	0.32	52.1 (21,954)	0.38	65.4 (12,041)	0.45	13.4***

Source: Data pooled from 2011-2015 National Health Interview Survey (NHIS) data sample adult files (n= 60,800 unweighted; Higher-income N= 18,651; Lower-income N: 42,149) obtained from the Integrated Health Interview Series (IHIS).
Note: FPL = federal poverty level; Pap = Papanicolaou.
[1]within the past 12 months of the survey date.
[2]%FPL was determined by using the reported total family income compared to the U.S. Census Bureau's poverty thresholds (based on income, family size, and the number of children under age 18) for the year the respondent was asked about .
Note: Stata Version 12 svy commands were used to account for the probability of selection, stratification and clustering due to the NHIS complex survey design. Percentage is weighted column percent. P values were based on Pearson chi-square test (linearized standard errors of weighted column proportions were produced) comparing lower- and higher-income women aged 21-65. ***P <0.01 and **P<0.05 for difference between < 400% FPL and≥400% FPL.

Table 2. Characteristics of Adult Women (Aged 21–65 Years) by Income[1]: 2011–2015 National Health Interview Survey, United States

	<400% FPL		≥400% FPL		Percentage Point Difference	
Characteristic	Percent	SE	Percent	SE		
Health Insurance Status						
Insured	76.10	0.3	96.40	0.17	20.30	***
Age						
21-25	15.10	0.28	6.30	0.28	-8.80	***
26-29	10.30	0.21	7.40	0.25	-2.90	***
30-39	24.30	0.27	20.40	0.4	-3.90	***
40-49	21.40	0.26	24.20	0.4	2.80	***
50-65	29.00	0.32	41.70	0.45	12.70	***
Race/Ethnicity						
White, non-Hispanic	55.60	0.48	77.50	0.41	21.90	***
Black, non-Hispanic	16.20	0.37	7.00	0.26	-9.20	***
Other race, non-Hispanic	7.30	0.23	8.30	0.26	1.00	***
Hispanic	21.00	0.39	7.20	0.24	-13.80	***
US Citizenship Status						
US citizen	87.70	0.31	96.00	0.19	8.30	***
Educational Attainment						
Less than high school diploma or GED	17.20	0.3	1.80	0.12	-15.40	***
High School Diploma or GED	27.40	0.3	13.60	0.34	-13.80	***
Some college or complete Associate's degree	36.00	0.34	28.10	0.43	-7.90	***

Bachelor's, Master's, professional, or doctoral degree	19.40	0.31	56.50	0.49	37.10	***
Employment Status						
Employed	64.80	0.37	84.30	0.37	19.50	***
Marital Status						
Married	45.70	0.42	71.00	0.47	25.30	***
Self-Reported Health Status						
Poor health	17.00	0.25	4.70	0.22	-12.30	***
Region						
Northeast	14.70	0.36	20.60	0.54	5.90	***
Northcentral/Midwest	22.90	0.5	21.40	0.55	-1.50	**
South	39.80	0.52	34.40	0.72	-5.40	***
West	22.70	0.48	23.60	0.53	0.90	
Survey Year						
2011	19.60	0.28	19.80	0.39	0.20	
2012	19.90	0.24	19.60	0.38	-0.30	
2013	19.90	0.25	19.70	0.36	-0.20	
2014	20.50	0.29	19.80	0.37	-0.70	
2015	20.10	0.28	21.10	0.4	1.00	**

Source: Data pooled from 2011-2015 National Health Interview Survey (NHIS) data sample adult files (n= 60,800 unweighted) obtained from the Integrated Health Interview Series (IHIS).
Note: FPL = federal poverty level; Pap = Papanicolaou.
[1]%FPL was determined by using the reported total family income compared to the U.S. Census Bureau's poverty thresholds (based on income, family size, and the number of children under age 18) for the year the respondent was asked about.
Note: Stata Version 12 svy commands were used to account for the probability of selection, stratification and clustering due to the NHIS complex survey design. Percentage is weighted column percent. P values were based on Pearson uncorrected chi-square test (linearized standard errors of weighted column proportions were produced) comparing lower- and higher-income women aged 21-65.
***P <0.01 and **P<0.05 for difference between < 400% FPL and ≥400% FPL.

Multivariate Linear Probability Models

The linear probability models estimated for decomposition analysis show that the characteristics included in the model were estimated to have associations with Pap smear use consistent with the previously discussed expectations (Tables A4 and A5). Having any form of health insurance and higher educational attainment was associated with a higher probability of using the preventive service. Being married, a US citizen, or employed were also associated with higher probabilities of utilizing the Pap test. Black, non-Hispanic women were more likely to use the Pap smear, consistent with past evidence. Surprisingly inconsistent with previous findings, being Hispanic was estimated to be associated with higher use of the service. Women of ages 40-65 were less likely than women 21-25 years of age to use the Pap smear, whereas women aged 26-39 were more likely than women in the youngest age group to use this preventive service. Poor health status was associated with lower probabilities of using the Pap smear.

Decomposition Analysis

Oaxaca-Blinder linear regression decomposition results are given in Table 3. There was a statistically significant observed gap in Pap smear use of 13.4 percentage points between lower- and higher-income populations of women in this sample ($p<0.01$). The portion of this difference attributable to differences in mean characteristics for these populations comprised 5.1 percentage points of the 13.4 total percentage point difference, or 38.4% of the observed income-related gap in Pap smear use. The remaining portion, 8.2 percentage points, was not attributable to population characteristics, but attributable to either coefficients (i.e. effectiveness of those characteristics) or unobservable characteristics.

Of particular interest, the difference in insurance coverage by income statistically significantly explained 4.6 percentage points of the 13.4 percentage point overall difference, comprising 34.3% of the total difference ($p<0.01$). Figure 2 depicts the actual receipt of the Pap smear for higher- and lower-income women aged 21-65 from 2011-2015. Figure 2 also shows the expected increase in utilization if the lower-income women had the insurance coverage levels of the higher-income population. If this were the case, the percentage of lower-women reporting having received a Pap smear within the past 12 months of the survey would have been 4.6 percentage points higher, an increase from 52.1% to 56.7%. However, it is clear from Figure 2 that even if lower-income women were to have had the insurance coverage characteristics of the higher-income population in this sample, their rate of Pap testing would still remain 8.7 percentage points below that of the higher-income women.

Table 3. Blinder-Oaxaca Linear[1] Decomposition of Differences in Pap Smear Utilization Among Women (Aged 21-65 Years) by Income[2]: 2011-2015 National Health Interview Survey, United States

		Linearized SE
Total Difference between Women ≥400%FPL and <400% FPL	0.134***	0.006
Difference not attributable to observable characteristics[3]	0.082***	0.007
Difference attributable to observable characteristics[4]	0.051***	0.004
Differences attributable to		
Health Insurance	0.046***	0.002
Age	-0.023***	0.001
Other Demographics[5]	-0.007***	0.003
Education	0.021***	0.003
Poor health status	0.004***	0.001
Employed	0.009***	0.001
Region and year	0.001*	0.001

Source: Data pooled from 2011-2015 National Health Interview Survey (NHIS) data sample adult files (n= 60,800 unweighted) obtained from the Integrated Health Interview Series (IHIS). Note: FPL = federal poverty level; Pap = Papanicolaou. N=60,800 unweighted observations (Higher-income N: 18,651; Lower-income N: 42,149). Note: Stata Version 12 svy commands were used to account for the probability of selection, stratification and clustering due to the NHIS complex survey design. P values were based on linearized standard errors produced using the Oaxaca command delta method. ***P <0.01, **P<0.05, and *P<0.10 .

[1]Decomposition based on stratified Linear Probability Models. Differences in estimated coefficients are weighted by the mean observed characteristics of the higher income group, and differences in mean observed characteristics are weighted by estimated coefficients of the lower income group (**D=0**, twofold decomposition where the interaction between the gap in endowments and the gap in coefficients is placed in the **unexplained** portion of the difference).

[2] %FPL was determined by using the reported total family income compared to the U.S. Census Bureau's poverty thresholds (based on income, family size, and the number of children under age 18) for the year the respondent was asked about.

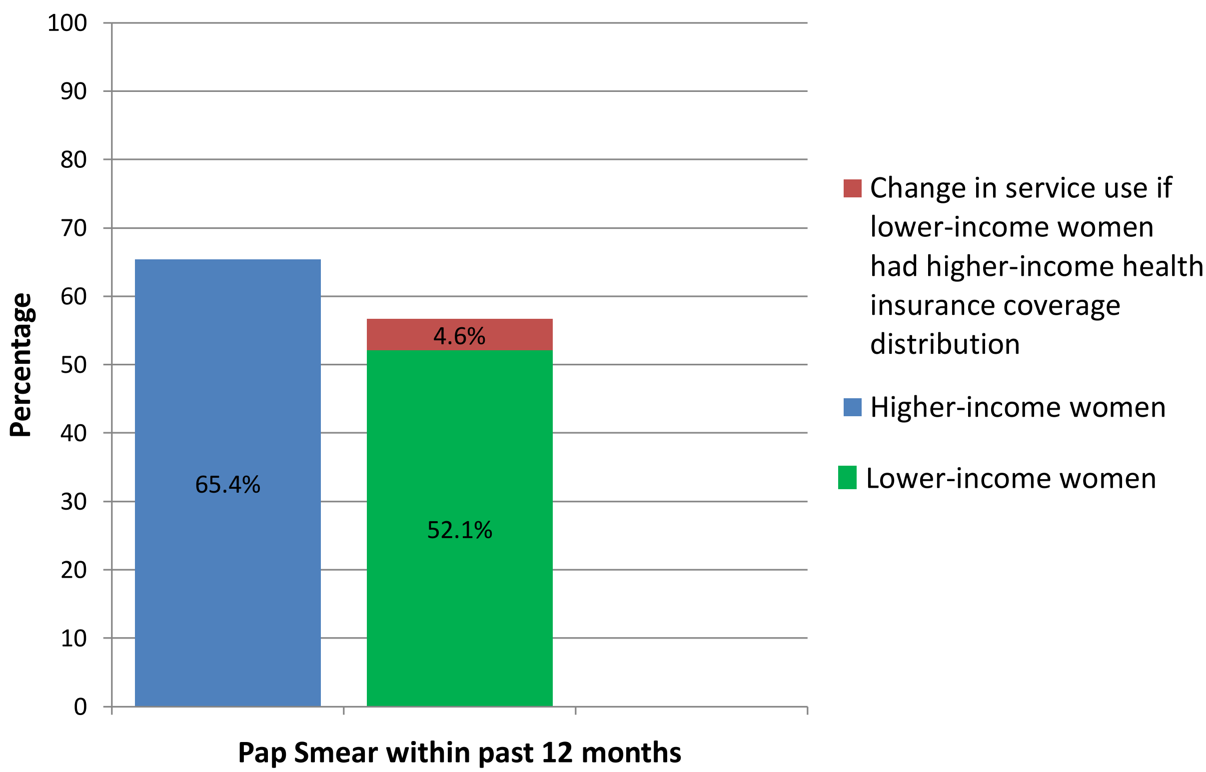
[3] (unexplained portion, due to coefficients)

[4] (explained portion, due to means)

[5]Other demographics include marital status, race/ethnicity, and US citizenship

Furthermore, differences in educational attainment played a significant role in explaining 15.7% of the difference in Pap smear utilization by income (p<0.01). Population differences in reporting poor health status explained 3.0% and differences in employment status explained 6.7% of the income-related gap in Pap testing. Table 3 shows that if lower-income women had the population age distribution of the higher-income group, their rate of Pap testing would actually have been 2.3 percentage points lower. This is not surprising given that older women were less likely to use the Pap smear (Table A4 and Table A5), and the higher income population was more likely to be older (Table 2).

Figure 2. Pap Smear Utilization by Adult Women (Aged 21-65 years): 2011-2015 National Health Interviews Survey, United States



Note: Pap = Papanicolaou. Lower (higher) income adults have incomes below (at or above) 400% of the FPL.

<u>Robustness checks</u>

Alternative decomposition results were estimated to compare with this paper's main model specification (4) given by Table 3. Tables A1 and A2 show the results when specifications (7) and (12) are estimated, rather than the main specification (4) used in this analysis. As shown in Table A1, there are sizable differences in the portions explained and unexplained by observable characteristics when form (7) is assumed, where coefficients used in the decomposition were generated from the higher income group's linear probability model, compared to form (4), where the lower income group's coefficients are used. Of the 13.4 percentage point gap, 10 percentage points were explained by observables in specification (7), rather than just 5.1 percentage points explained in specification (4). However, it is shown that even in specification (7), 5.4 percentage points of the disparity in service use is attributable to population differences in insurance coverage.

Table A2 gives estimated results from the Neumark specification (12) where coefficients used in the decomposition were generated from a pooled regression incorporating both the lower- and higher-income population. These decomposition results are more similar to those of the main analysis specification (4) shown in Table 3, than were the results of the (7) in Table A1. In this pooled version, 6.4 percentage points of the 13.4 percentage point gap in Pap use is explained by observable characteristics, compared to 5.1 percentage points explained in the original version (Table 3). The portion of the overall gap attributable to population differences in insurance coverage is the same as that found in the original specification: 4.6 percentage points ($p<0.01$).

Furthermore, additional alternative results estimated using non-linear Blinder-Oaxaca decomposition are shown in Table A3, which compares these estimates (which used logistic regressions stratified by income) to the previously discussed estimates (which used stratified linear probability model coefficients), using specification (4). It is evident that in this case, results are nearly identical whether the coefficients used in decomposition are based on logistic regression estimations as compared to linear probability model estimations used in this paper's main decomposition analysis. Linear decomposition estimated 5.13 percentage points of the total 13.4 percentage point income-related gap in Pap smear receipt were explained by observed characteristic differences, whereas the explained portion was estimated to comprise 5.14

percentage points of the gap using logistic decomposition. Linear decomposition results estimated with *nldecompose*, as compared to linear decomposition estimated using *oaxaca* commands, are also shown in Table A3; these estimations yielded identical estimates with similar standard errors.

VII. Discussion

This paper expanded upon the pre-ACA implementation analysis performed by McMorrow, Kenney, and Goin (2014) by providing more recent, post-ACA evidence regarding the gap in Pap testing among lower- and higher-income women and the role of insurance coverage and other factors. This post-ACA analysis found that 38.4% of the 13.4 percentage point income-related gap in yearly Pap smear utilization was attributable to population differences in observable characteristics; specifically, insurance coverage explained 34.3% of the gap. For comparison, the pre-ACA analysis—which examined income-related disparities in having had a Pap smear within the previous 3 years, rather than the current analysis' outcome of Pap smear utilization within the previous 12 months—had found that 46% of the 7.9 percentage point gap in Pap smear utilization (receipt every 3 years) was explained, and insurance coverage similarly played a role in explaining 35% of that pre-ACA gap. Interestingly, educational attainment was found to explain a smaller portion of the gap in Pap testing (15.7% during 2011-2015) compared to the previous period (2005-2010), when it explained 30% of the gap (McMorrow, Kenney, and Goin, 2014).

Furthermore, individual characteristics stratified by income in the current, post-ACA analysis are similar to those in the previous, pre-ACA study. As in the 2005-2010 MEPS sample used pre-ACA, lower-income women in the post-ACA, 2011-2015 NHIS sample reported higher rates of uninsurance (23.9%) compared to the higher-income group (3.6%); McMorrow and colleagues (2014) had found the rates of uninsurance for lower- and higher-income women to be 21.8% and 4.3%, respectively. Further, the stratified population characteristics for age, race/ethnicity, educational attainment, employment, and marital status were distributed similarly in both analyses.

In both the pre- and post-ACA analyses, a sizable portion of income-related disparities in the receipt of the Pap smear remained unexplained by insurance coverage. Based on this paper's findings, even if lower-income women had the insurance coverage characteristics of the higher-income population, their rate of receiving the Pap smear would have been 4.6 percentage points higher. However, this would mean their rate of Pap testing would still remain 8.7 percentage points below that of the higher-income women.

It is possible that women had access to Pap testing independent of being insured, for example through cervical and breast cancer prevention programs, which are not accounted for in this paper. Furthermore, it is not clear from these findings whether moral hazard is an issue; it is possible the higher-income group of women reporting higher rates of insurance coverage and higher use of the Pap smear could be overusing the preventive service, contrary to the recommended guidelines for appropriate use. Similarly, it is not clear whether the lower-income group of women reporting lower rates of Pap testing are actually underusing the service. However, while this paper's findings do not sufficiently identify underuse or overuse, the results are helpful in identifying that lower- and higher-income women are not utilizing the Pap test at the same rate and that factors beyond insurance coverage emphasized by the ACA may help explain the rest of this difference in preventive service utilization.

Overall, the portion of the gap remaining unexplained even after accounting for differences in individual characteristics could be due to the actual effect of those characteristics on the decision to get a Pap test varying by income. For example, a lower-income woman having a Bachelor's degree or higher educational attainment might not have the same impact on her access to or demand for care as it does for a higher-income woman of the same educational level. The remaining unexplained portion of the disparity could also be due to unobservable population characteristics not included in this model, such as systematic distrust of the healthcare system or preferences for risk. Furthermore, it may be that beyond uninsurance or cost sharing barriers, lower-income women face more logistical barriers to receiving the Pap smear. In a 2013 survey, lower-income women were more likely to report transportation problems, being unable to take time off of work, or having trouble finding childcare when trying to utilize health services (Salganicoff, et al., 2014).

Future research might decompose disparities in Pap smear utilization by racial/ethnic groups, as previous studies have used older data to decompose racial gaps in health care utilization, but have not explored Pap smear receipt (Hargraves and Hadley, 2003). There has been evidence of lower utilization among women identifying as Asian American, Native Hawaiian and Other Pacific Islander, American Indian and Alaska Native, or Hispanic had Pap smear utilization rates below the national average, and below the rates of women identifying as African-American or white (James, 2009). This difference in utilization of the Pap test may partially explain the higher levels of cervical cancer found in Hispanic and Asian women, and would be an important gap to explore further (NIH, 2010). Future analyses might also decompose these racial gaps in Pap testing separately by income groups in a methodology similar to Shen and Long (2006).

VIII. Limitations

There are several limitations with this paper. First, regression decomposition analysis does not estimate a causal impact of explanatory factors on Pap test receipt due to the possibility of omitted variable bias. Specifically, there may be unobservable factors affecting both the independent variable of insurance coverage and the dependent outcome of Pap testing. For example, if insurance coverage and Pap smear use are positively correlated and unobservable attitudes of distrust toward the health care system might be negatively correlated with both insurance coverage and Pap smear utilization, the resulting omitted variable bias would be expected to overestimate the effects of insurance coverage in explaining gaps in utilization.

Furthermore, there may be measurable factors which are not captured in a data set, an issue of particular concern for this paper because information regarding whether a woman lived in an urban or rural setting was not available in the data set and was consequently left out of the model. Urban/rural status is likely correlated with both the outcome of Pap smear utilization and other covariates included in the model. In fact, past research has shown rural residents to have lower income levels and less educational attainment—both associated with less likelihood of utilizing preventive services—than their urban counterparts (Larson and Correa-de-Araujo, 2006). Furthermore, rural residents have been found to be more likely to face issues of uninsurance, less access to providers, longer distances to health facilities, and less access to costly medical technology (MacKinney, et al., 2014). Since rural residents tend to have lower income levels than their urban counterparts, splitting the sample into higher- and lower-income groups to examine the income-related gap in Pap testing without accounting for urban/rural status could potentially be capturing part of an urban/rural gap in testing. Additionally, much like the above example, omitted variable bias is expected to lead to an under- or overestimation of the effects of the included covariates known to be correlated with urban/rural status (i.e. educational attainment, insurance coverage, etc.).

Furthermore, linear probability models were used in the decomposition analysis for the binary dependent variable of receiving a Pap test to facilitate comparisons between this paper and the McMorrow, Kenney, and Goin (2014) study. Although linear decomposition has also been utilized in several related studies to explain disparities in binary outcomes, it has been shown to be problematic given that predicted probabilities from these estimations can fall out of the 0-1 range (Hargraves and Hadley, 2003). However, this paper does address the robustness of using linear decomposition by also estimating decomposition analysis using stratified logistic regression models.

As the data used in this analysis was survey data, recall error for women self-reporting their screening history over the past 12 months may be present. In addition, while insurance coverage was expected to influence women's decision about utilizing the Pap test, the timing of the survey questions regarding Pap smear testing and health insurance coverage are not the same. Women were asked whether they had received a Pap smear within the past 12 months and whether they had insurance coverage at the time of the survey. It is possible that an insured woman at the time of the survey might have had gaps in coverage over the previous 12 months in which she made her decision. Future analyses should incorporate information available in the NHIS data set into the model regarding whether currently insured women had some time in the past 12 months in which they did not have any insurance coverage.

Additionally, 11.8% of observations were excluded from the analytic sample of women aged 21-65 years asked about Pap test utilization due to missing values for the ratio of family income to the Federal poverty level. The final analytic sample for this study was obtained after listwise deletion of observations with missing values for included covariates, the outcome of Pap smear utilization, and income level (used to define the two subgroups of higher- and lower-income women). Beyond affecting statistical power of tests of statistical inference, if the values are not missing completely at random (MCAR) or not missing at random (MAR), the high percentage of observations excluded due to missing income information may bias the study's results (Kang, 2013). The characteristics of women who refused to respond or reported unknown response to a survey question giving rise to a variable used to describe Pap smear utilization might vary systematically from those of women who did provide this information. It is possible that lower-income women were less likely to report their income than their higher-income counterparts, for example. Listwise deletion of those observations would remove that systematically different subset of women from the analysis, biasing the findings. Further analyses might incorporate multiple imputation techniques to address this potential bias by imputing missing income values so that currently excluded observations could be included in the sample. Tables A6 and A7 in the Appendix give more information about the number of missing observations for the women's income level and Pap smear utilization, and the percent of the original sample these represent. Moreover, just one of the five imputed employment status data from NHIS was used in the present analysis. Future analyses should employ multiple imputation techniques using all five of these provided variables.

This paper uses NHIS data, rather than the MEPS data used by McMorrow and colleagues (2014). The NHIS contains information on whether sample adult women had received a Pap test within the previous 12 months, whereas the MEPS asks women whether they used the service within the past 1, 3, or 5 years (if at all). Thus, the Pap test outcome variable in the previous study was a measure of use within the past 3 years, rather than the past 12 months as is the case in this paper, which complicates comparisons across studies and directly assessing whether women are up to date with their recommended preventive services. Nevertheless, this variable was still useful in examining differences in utilization across income groups for the purposes of this paper. The NHIS data also does not contain the variables used in the previous work to measure risk behavior (i.e. seatbelt use habits) as a covariate. Future analyses could improve upon this paper's comparison to previous studies by utilizing more recent data from the Medical Expenditure Panel Survey (MEPS) used by McMorrow and colleagues (2014) and including more detailed insurance coverage information, as well. However, the current analysis had the advantage of producing standard errors allowing statistical inference for the estimates of explained components of the income-related disparities in Pap testing, which were not estimated in the McMorrow, Kenney, and Goin (2014) analysis.

Lastly, only one measure of women's preventive service use, receipt of the Pap smear was studied in this analysis. Further exploration of income-related gaps in other measures of preventive service utilization or access (such as having a usual source of care when sick or delay in receiving necessary care or prescription drugs) would improve upon this paper.

IX. Conclusion

This analysis used Oaxaca-Blinder regression-based decomposition to identify the roles and significance of insurance coverage and other factors in explaining the 13.4 percentage point income-related gap in Pap smear utilization in the United States from 2011-2015 among lower- and higher-income adult women. The results are helpful in comparing different methods of estimating the Oaxaca-Blinder decomposition for binary outcomes. The findings are especially useful to compare factors explaining more recent income-related gaps in Pap testing in the years in which the Affordable Care Act (ACA) was in effect to estimates previously studied from the years prior to the law's implementation.

This analysis estimated that lower- and higher-income population differences in insurance coverage explain 34.3% of the income-related gap in utilization. Given the ACA's elimination of cost sharing for the Pap smear and other clinical preventive women's health services and the guidelines for Pap smear utilization changing in the years following the ACA implementation, this more recent evidence is useful for policy makers interested in understanding this disparity in preventive service utilization.

X. References

Acemoglu, D., Finkelstein, A., & Notowidigdo, M. J. (2013). Income and health spending: Evidence from oil price shocks. *Review of Economics and Statistics*, *95*(4), 1079-1095.

Barbaresco S, Courtemanche CJ, Qi Y. (2015). Impacts of the Affordable Care Act dependent coverage provision on health-related outcomes of young adults. *Journal of Health Economics*, Volume 40, Pages 54-68, ISSN 0167-6296.http://dx.doi.org/10.1016/j.jhealeco.2014.12.004.

Bauer, T. K., & Sinning, M. (2008). An extension of the Blinder–Oaxaca decomposition to nonlinear models. *AStA Advances in Statistical Analysis*, *92*(2), 197-206.

Blewett, L. A., Rivera Drew, J. A., Griffin, R. M., King, L., & Williams, K. C.W.. (2016). IPUMS Health Surveys: National Health Interview Survey, Version 6.2 [dataset]. Minneapolis, MN: University of Minnesota. http://doi.org/10.18128/D070.V6.2

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, 436-455.

Bustamante, A. V., Fang, H., Rizzo, J. A., & Ortega, A. N. (2009). Understanding observed and unobserved health care access and utilization disparities among US Latino adults. *Medical Care Research and Review*, *66*(5), 561-577.

Bustamante, A. V., Fang, H., Garza, J., Carter-Pokras, O., Wallace, S. P., Rizzo, J. A., & Ortega, A. N. (2012). Variations in healthcare access and utilization among Mexican immigrants: the role of documentation status. *Journal of Immigrant and Minority Health*, *14*(1), 146-155.

Chen, Meng-Kan, et al. (2011). "Cost-effectiveness analysis for Pap smear screening and human papillomavirus DNA testing and vaccination." *Journal of evaluation in clinical practice* 17.6 (2011): 1050-1058.

Fairlie, R. (2014). An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models. Journal of Economic and Social Measurement. UC Santa Cruz: Department of Economics, UCSC. Retrieved from: http://escholarship.org/uc/item/7j86n5q3.

Han, X., Yabroff, K. R., Guy, G. P., Zheng, Z., & Jemal, A. (2015). Has recommended preventive service use increased after elimination of cost-sharing as part of the Affordable Care Act in the United States?. *Preventive medicine*, *78*, 85-91.

Hargraves, J. L., & Hadley, J. (2003). The contribution of insurance coverage and community resources to reducing racial/ethnic disparities in access to care. *Health services research*, *38*(3), 809-829.

IOM (Institute of Medicine). (2011). Clinical Preventive Services for Women: Closing the Gaps. Washington, DC: The National Academies Press.

Jack, W. (1999). *Principles of health economics for developing countries*. Washington, DC: World Bank Publications.

James, C. (2009). Putting Women's Health Care Disparities on the Map: Examining Racial and Ethnic Disparities at the State Level. *Kaiser Family Foundation*. June 2009.

Jann, B. (2008). The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal*, *8*(4), 453-479.

Jones, R. K., & Sonfield, A. (2016). Health insurance coverage among women of reproductive age before and after implementation of the affordable care act. *Contraception*, *93*(5), 386-391.

Kaiser Family Foundation. (2017). Current status of state medicaid expansion decisions. Retrieved September 10, 2017, from http://www.kff.org/health-reform/slide/current-status-of-the-medicaid-expansion-decision/

Kang, H. (2013). The prevention and handling of the missing data. Korean Journal of Anesthesiology, 64(5), 402-406.

Larson, S., & Correa-de-Araujo, R. (2006). Preventive health examinations: a comparison along the rural-urban continuum. *Women's Health Issues*, 16, pp. 80-88, doi:10.1016/j.whi.2006.03.001

MacKinney, A., Coburn, A., Lundblad, J., McBride, T., Mueller, K., & Watson, S. (2014). Access to Rural Health Care–A Literature Review and New Synthesis. Policy Report. Rupri: Rural Policy Research Institute.Martinez ME, Zammitti EP, Cohen RA. (2017). Health insurance coverage: Early release of estimates from the National Health Interview Survey, January–September 2016. National Center for Health Statistics. Retrieved from: http://www.cdc.gov/nchs/nhis/releases.htm.

McMorrow, S., Kenney, G. M., & Goin, D. (2014). Determinants of receipt of recommended preventive services: implications for the Affordable Care Act. *American journal of public health*, *104*(12), 2392-2399.

National Institutes of Health. (2010). *Cervical Cancer* [Fact Sheet]. Retrieved from https://report.nih.gov/nihfactsheets/Pdfs/CervicalCancer(NCI).pdf.

Neumark, D. (1988). Employers' discriminatory behavior and the estimation of wage discrimination. *Journal of Human Resources*, 23: 279–295.

O'Donnell, O. Van Doorslaer, E., Wagstaff, A., & Lindelow, M. (2008). Explaining differences between groups: Oaxaca decomposition. *Analysing health equity using household survey data. Inst Learn Resourc Ser*, 147-157.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, 693-709.

Rezayatmand, R., Pavlova, M., Groot, W. (2012). The impact of out-of-pocket payments on prevention and health-related lifestyle: a systematic literature review. *European Journal of Public Health, 21*(1), 74-79.

Salganicoff, Alina, et al. (2014). "Women and health care in the early years of the ACA: Key findings from the 2013 Kaiser Women's Health Survey." *Menlo Park (CA): The Henry J. Kaiser Family Foundation.* https://report.nih.gov/nihfactsheets/viewfactsheet.aspx?csid=76.

Shen, Y. C., & Long, S. K. (2006). What's Driving the Downward Trend in Employer-Sponsored Health Insurance?. *Health services research*, *41*(6), 2074-2096.

Sinning, M., Hahn, M., & Bauer, T. K. (2008). The Blinder-Oaxaca decomposition for nonlinear regression models. *The Stata Journal*, *8*(4), 480-492.

StataCorp. (2011). Stata Statistical Software: Release 12. College Station, TX: StataCorp LP.

U.S. Cancer Statistics Working Group. (2015). *United States Cancer Statistics: 1999–2012 Incidence and Mortality Web-based Report.* Atlanta (GA): Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute.

U.S. Department of Health & Human Services. (2011). *ACA Rules on Expanding Access to Preventive Services for Women*. Retrieved from HHS.gov.

U.S. Preventive Services Task Force. (2014). Final Recommendation Statement: Cervical Cancer: Screening. Retrieved from: http://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/cervical-cancer-screening.

XI. Appendix

Table A1. Alternative Blinder-Oaxaca Linear[1] Decomposition of Differences in Pap Smear Utilization Among Women (Aged 21-65 Years) by Income[2]: 2011-2015 National Health Interview Survey

		Linearized SE
Total Difference between Women ≥400%FPL and <400% FPL	0.134***	0.006
Difference not attributable to observable characteristics[3]	0.034***	0.010
Difference attributable to observable characteristics[4]	0.100***	0.009
Differences attributable to,		
Health Insurance	0.053***	0.005
Age	-0.013***	0.002
Other Demographics[5]	0.007***	0.004
Education	0.035***	0.006
Poor health status	0.011***	0.003
Employed	0.008***	0.003
Region and year	0.000	0.001

Source: Data pooled from 2011-2015 National Health Interview Survey (NHIS) data sample adult files (n= 60,800 unweighted) obtained from the Integrated Health Interview Series (IHIS).

Note: FPL = federal poverty level; Pap = Papanicolaou. N=60,800 unweighted observations (Higher-income N: 18,651; Lower-income N: 42,149).

[1]Decomposition based on stratified Linear Probability Models. Differences in estimated coefficients are weighted by the mean observed characteristics of the lower income group, and differences in mean observed characteristics are weighted by estimated coefficients of the higher income group (**D=I**, twofold decomposition where the interaction between the gap in endowments and the gap in coefficients is placed in the **explained** portion of the difference).

Note: Stata Version 12 svy commands were used to account for the probability of selection, stratification and clustering due to the NHIS complex survey design. P values were based on linearized standard errors produced using the Oaxaca command delta method. ***P <0.01, **P<0.05, and *P<0.10 .

[2] %FPL was determined by using the reported total family income compared to the U.S. Census Bureau's poverty thresholds (based on income, family size, and the number of children under age 18) for the year the respondent was asked about.

[3] (unexplained portion, due to coefficients)

[4] (explained portion, due to means)

[5]Other demographics include marital status, race/ethnicity, and US citizenship

Table A2. Alternative Blinder-Oaxaca Pooled Linear[1] Decomposition of Differences in Pap Smear Utilization Among Women (Aged 21-65 Years) by Income[2]: 2011-2015 National Health Interview Survey		
		Linearized SE
Total Difference between Women ≥400%FPL and <400% FPL	0.134***	0.006
Difference not attributable to observable characteristics,[3]	0.070***	0.006
Difference attributable to observable characteristics[4],	0.064***	0.004
Differences attributable to, %		
Health Insurance	0.046***	0.002
Age	-0.020***	0.001
Other Demographics	-0.004**	0.002
Education	0.026***	0.003
Poor health status	0.005***	0.001
Employed	0.009***	0.001
Region and year	0.001	0.001

Source: Data pooled from 2011-2015 National Health Interview Survey (NHIS) data sample adult files (n= 60,800 unweighted) obtained from the Integrated Health Interview Series (IHIS).

Note: FPL = federal poverty level; Pap = Papanicolaou. N=60,800 unweighted observations (Higher-income N: 18,651; Lower-income N: 42,149).

[1]Decomposition based on Linear Probability Models. Differences in the mean observed characteristics are weighted by coefficients obtained from a pooled data regression.

Note: Stata Version 12 svy commands were used to account for the probability of selection, stratification and clustering due to the NHIS complex survey design. P values were based on linearized standard errors produced using the Oaxaca command delta method. ***P <0.01, **P<0.05, and *P<0.10 .

[2] %FPL was determined by using the reported total family income compared to the U.S. Census Bureau's poverty thresholds (based on income, family size, and the number of children under age 18) for the year the respondent was asked about.

[3] (unexplained portion, due to coefficients)

[4] (explained portion, due to means)

[5]Other demographics include marital status, race/ethnicity, and US citizenship

Table A3. Comparison of Linear and Non-Linear (Logistic) Blinder-Oaxaca Decomposition Results for Differences in Pap Smear Utilization Among Women (Aged 21-65 Years) by Income[1]: 2011-2015 National Health Interview Survey, United States			
	LPM: oaxaca[2] (SE)	LPM: nldecompose[3] (SE)	Logit: nldecompose[4] (SE)
Total Difference between Women ≥400%FPL and <400% FPL	0.1336305*** (0.0055572)	0.1336305*** (0.0063791)	0.1336305*** (0.0047412)
Difference not attributable to observable characteristics,[5]	0.0823626*** (0.0067277)	0.0823626*** (0.0067146)	0.0822237*** (0.0061671)
Difference attributable to observable characteristics,[6]	0.0512679*** (0.0042264)	0.0512679*** (0.0038411)	0.0514068*** (0.0041614)

Source: Data pooled from 2011-2015 National Health Interview Survey (NHIS) data sample adult files (n= 60,800 unweighted) obtained from the Integrated Health Interview Series (IHIS). Note: FPL = federal poverty level; Pap = Papanicolaou; LPM = Linear Probability Model. N=60,800 unweighted observations (Higher-income N: 18,651; Lower-income N: 42,149).

[1]%FPL was determined by using the reported total family income compared to the U.S. Census Bureau's poverty thresholds (based on income, family size, and the number of children under age 18) for the year the respondent was asked about.

[2]Decomposition based on linear regression models estimated using the Stata oaxaca command.

[3]Decomposition based on linear regression models estimated using the Stata nldecompose command.

[4]Decomposition based on logistic regression models estimated using the Stata nldecompose command.

Note: For each column, differences in estimated coefficients are weighted by the mean observed characteristics of the higher income group, and differences in mean observed characteristics are weighted by estimated coefficients of the lower income group (D=0, twofold decomposition where the interaction between the gap in endowments and the gap in coefficients is placed in the unexplained portion of the difference).

Note: Stata Version 12 svy commands were used for all estimations to account for the probability of selection, stratification and clustering due to the NHIS complex survey design. P values were based on standard errors produced using the Oaxaca (linearized standard errors using the delta method) and nldecompose (bootstrapped standard errors) commands. ***P <0.01, **P<0.05, and *P<0.10 .

[5] (unexplained portion, due to coefficients)

[6] (explained portion, due to means)

Table A4. Factors Associated with Yearly Pap Smear Utilization Based on Linear Regression, U.S. Women ages 21-65 with Incomes Below 400% of the Federal Poverty Level, 2011-2015 National Health Interview Survey

	LPM Coef.	SE	95% CI LCI	95% CI UCI	P-value
Health Insurance Status					
Uninsured (ref.)					
Insured	0.227***	(0.007)	0.212	0.241	0.000
Age					
19-25 (ref.)					
26-29	0.048***	(0.011)	0.026	0.070	0.000
30-39	-0.004	(0.011)	-0.026	0.018	0.743
40-49	-0.066***	(0.011)	-0.088	-0.043	0.000
50-65	-0.159***	(0.011)	-0.180	-0.138	0.000
Race/Ethnicity					
White, non-Hispanic (ref.)					
Black, non-Hispanic (ref.)	0.122***	(0.009)	0.104	0.141	0.000
Hispanic	0.060***	(0.010)	0.039	0.081	0.000
Other race (AIAN, Asian, other/multiple race), non-Hispanic	-0.055***	(0.014)	-0.084	-0.027	0.000
Educational Attainment (highest level achieved)					
Less than high school diploma or GED (ref.)					
High school diploma or GED	0.006	(0.010)	-0.013	0.025	0.553
Some college or complete Associate's degree	0.046***	(0.010)	0.026	0.066	0.000
Bachelor's, Master's, professional, or doctoral degree	0.068***	(0.011)	0.047	0.089	0.000
Marital Status					
Not Married (ref.)					
Married	0.043***	(0.006)	0.030	0.056	0.000
Region					
Northeast (ref.)					
Northcentral/Midwest	-0.032***	(0.012)	-0.057	-0.008	0.010
West	-0.035***	(0.011)	-0.057	-0.013	0.002
South	-0.023**	(0.011)	-0.045	-0.001	0.038
Self-Reported Health Stats					
Good health (ref.)					
Poor health	-0.030***	(0.009)	-0.047	-0.013	0.001
U.S. Citizenship Status					
Non-citizen (ref.)					
U.S. Citizen	0.030***	(0.011)	0.008	0.051	0.007
Employment Status					
Unemployed (ref.)					
Employed	0.048***	(0.007)	0.034	0.062	0.000
Survey Year					
2011 (ref.)					
2012	-0.006	(0.010)	-0.025	0.013	0.549
2013	-0.016*	(0.010)	-0.035	0.003	0.100

2014	-0.064***	(0.009)	-0.083	-0.046	0.000
2015	-0.058***	(0.011)	-0.079	-0.037	0.000
Constant	0.327***	(0.021)	0.285	0.369	0.000

Source: Data pooled from 2011-2015 National Health Interview Survey (NHIS) data sample adult files obtained from the Integrated Health Interview Series (IHIS).
Note: FPL = federal poverty level; Pap = Papanicolaou Lower-income N= 42,149 unweighted observations.
Note: Estimates and standard errors account for probability of selection, stratification and clustering due to the NHIS complex survey design. P values were based on linearized standard errors produced from Taylor Series Linearization using Stata Version 12. ***P <0.01, **P<0.05, and *P<0.10 .
Note: %FPL was determined by using the reported total family income compared to the U.S. Census Bureau's poverty thresholds (based on income, family size, and the number of children under age 18) for the year the respondent was asked about.

Table A5. Factors Associated with Yearly Pap Smear Utilization Based on Linear Regression, U.S. Women ages 21-65 with Incomes At or Above 400% of the Federal Poverty Level, 2011-2015 National Health Interview Survey

			95% CI		P-value
	LPM Coef.	SE	LCI	UCI	
Health Insurance Status					
Uninsured (ref.)					
Insured	0.260***	(0.023)	0.214	0.306	0.000
Age					
19-25 (ref.)					
26-29	0.066**	(0.026)	0.015	0.117	0.011
30-39	0.053**	(0.025)	0.003	0.104	0.036
40-49	-0.020	(0.025)	-0.070	0.030	0.426
50-65	-0.067***	(0.025)	-0.116	-0.018	0.007
Race/Ethnicity					
White, non-Hispanic (ref.)					
Black, non-Hispanic	0.080***	(0.015)	0.050	0.111	0.000
Hispanic	0.028*	(0.016)	-0.003	0.060	0.076
Other race (AIAN, Asian, other/multiple race), non-Hispanic	-0.047***	(0.015)	-0.077	-0.017	0.002
Educational Attainment (highest level achieved)					
Less than high school diploma or GED (ref.)					

High school diploma or GED	0.036	(0.038)	-0.040	0.111	0.355
Some college or complete Associate's degree	0.051	(0.037)	-0.021	0.124	0.166
Bachelor's, Master's, professional, or doctoral degree	0.118***	(0.036)	0.048	0.188	0.001
Marital Status					
Not Married (ref.)					
Married	0.035***	(0.010)	0.015	0.054	0.000
Region					
Northeast (ref.)					
Northcentral/Midwest	-0.042***	(0.014)	-0.069	-0.015	0.003
West	-0.069***	(0.013)	-0.093	-0.044	0.000
South	-0.010	(0.011)	-0.031	0.011	0.342
Self-Reported Health Stats					
Good health (ref.)					
Poor health	-0.086***	(0.022)	-0.130	-0.043	0.000
U.S. Citizenship Status					
Non-citizen (ref.)					
U.S. Citizen	0.121***	(0.024)	0.073	0.168	0.000
Employment Status					
Unemployed (ref.)					
Employed	0.039***	(0.014)	0.012	0.066	0.004
Survey Year					
2011 (ref.)					
2012	-0.017	(0.012)	-0.041	0.007	0.169
2013	-0.047***	(0.012)	-0.071	-0.023	0.000
2014	-0.061***	(0.012)	-0.085	-0.037	0.000
2015	-0.100***	(0.013)	-0.125	-0.074	0.000

Constant	0.237***	(0.051)	0.136	0.338	0.000

Source: Data pooled from 2011-2015 National Health Interview Survey (NHIS) data sample adult files obtained from the Integrated Health Interview Series (IHIS).
Note: FPL = federal poverty level; Pap = Papanicolaou; Higher-income N= 18,651unweighted observations.
Note: Estimates and standard errors account for probability of selection, stratification and clustering due to the NHIS complex survey design. P values were based on linearized standard errors produced from Taylor Series Linearization using Stata Version 12. ***P <0.01, **P<0.05, and *P<0.10 .
Note: %FPL was determined by using the reported total family income compared to the U.S. Census Bureau's poverty thresholds (based on income, family size, and the number of children under age 18) for the year the respondent was asked about.

Table A6. Number and Percent[1] of Missing Observations for Pap Smear Utilization[2] and Income Level[3] Among Women (Aged 21-65 Years): 2011-2015 National Health Interview Survey, United States

Pap smear in the past 12 months		Income Level (At/Above or Below 400% FPL)	
Number of observations missing	Percent	Number of observations missing	Percent
1,583	2.23	8,330	11.76

Source: Data pooled from 2011-2015 National Health Interview Survey (NHIS) data sample adult files (n= 70,847 unweighted; includes missing observations which were not included in the study's final analytical sample) obtained from the Integrated Health Interview Series (IHIS). Note: FPL = federal poverty level; Pap = Papanicolaou;
[1] Percent missing observations out of all sample adult women aged 21-65 asked about Pap smear utilization, 2011-2015
[2] within the past 12 months of the survey date.
[3] %FPL was determined by using the reported total family income compared to the U.S. Census Bureau's poverty thresholds (based on income, family size, and the number of children under age 18) for the year the respondent was asked about.

Table A7. Two-way Tabulation of Pap Smear Utilization[1] and Income Level[2] Among Women (Aged 21-65 Years): 2011-2015 National Health Interview Survey, United States

	Income Level			
Pap smear utilization	At/Above 400% FPL	Below 400% FPL	Missing	Total
No	6,655	20,449	3,439	30,543
Yes	12,099	22,128	4,494	38,721
Missing	296	890	397	1,583
Total	19,050	43,467	8,330	70,847

Source: Data pooled from 2011-2015 National Health Interview Survey (NHIS) data sample adult files (n= 70,847 unweighted; includes missing observations which were not included in the study's final analytical sample) obtained from the Integrated Health Interview Series (IHIS). Note: FPL = federal poverty level; Pap = Papanicolaou;
[1] within the past 12 months of the survey date.
[2] %FPL was determined by using the reported total family income compared to the U.S. Census Bureau's poverty thresholds (based on income, family size, and the number of children under age 18) for the year the respondent was asked about.

Table A8. Variable Descriptions						
New variable name	**Concept New Variable Operationalizes**	**Definition and codes for new variable**	**Original IHIS variable name**	**Definition and codes for the original variable**	**Survey question(s) used to obtain the original variable**	**Universe**
pap	Women's preventative care utilization--whether or not women respondents have had a Pap smear within a year of the survey	=0 if answered 'no' (recode from 2); =1 if answered 'yes' (already coded 1); =. (missing) if value is 'Refused' (recode from 7) or 'unknown' (recode from 8 and 9)	**Paphad1yr**	0=NIU (not in universe); 2=Yes; 1=No; 7=Refused; 8=unknown-not ascertained; 9=unknown-Don't Know (weight: sampweight)	From sample adult questionnaire: "'Have you had a Pap smear or Pap test during the past 12 months?' *Read if necessary. 'A Pap smear or Pap test is a routine test for women in which the doctor examines the cervix, takes a cell sample from the cervix with a small stick or brush, and sends it to the lab.'"	Female sample adults age 18+
hisprace	Race/ethnicity	=1 if white, non-Hispanic (recode from racenew=1 and hispyn=1) =2 if Black/African American, non-Hispanic (recode from racenew=2 and hispyn=1) =3 if AIAN, non-Hispanic (recode from racenew=3 and hispyn=1) =4 if Asian, non-Hispanic (recode from racenew=4 and hispyn=1) =5 if multiple race, non-Hispanic (recode from racenew=5 and hispyn=1) =6 if Hispanic	**Hispyn** and **racenew**	**Hispyn:** 2=Hispanic, 1=not Hispanic, 7=unknown-refused, 8=unknown-not ascertained, 9=unknown-don't know **Racenew (post 1997 OMB standards):** 1=white, 2=Black/African American, 3=AIAN, 4=Asian, 5=Multiple Race, 6=Other, 61=not releasable, 97=Unknown-refused, 98=unknown-not ascertained, 99=unknown-don't know (weight: perweight for both)	**Hispyn:** "'Do you consider yourself to be Hispanic or Latino?' * Read if necessary. 'Puerto Rican, Cuban/Cuban American, Dominican (Republic), Mexican/Mexican American, Central or South American, Other Latin American, Other Hispanic/Latino. Where did your ancestors come from?'" **Racenew:** "'What race or races do you consider yourself to be? Please select 1 or more of these categories.' * Enter all that apply: White, Black/African American, Indian (American), Alaska Native, Native Hawaiian, Guamanian or Chamorro, Samoan, Other Pacific Islander, Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, Other Asian, Some other race, Refused, Don't know"	All persons

		(recode from hispyn=2) =. (missing) if value is 'Race group not releasable' (recode from racenew=61) or if unknown (recode from hispyn=7, 8, or 9 or if racenew=97, 98, 99				
ethrace	Race/ethnicity (comparable categories used in McMorrow, Kenney, and Goin, 2014)	=1 if white, non-Hispanic (recode from racenew=1 and hispyn=1) =2 if Black/African American, non-Hispanic (recode from racenew=2 and hispyn=1) =3 if Other, non-Hispanic (recode to combine AIAN ,Asian and multirace non-Hispanic into one category: racenew=3 or racenew=4 or racenew=5 and hispyn=1) =4 if Hispanic (recode from hispyn=2) =. (missing) if value is 'Race group not releasable' (recode from racenew=61)	**Hispyn** and **racenew**	**See above**	**See above**	See above

		or if unknown (recode from hispyn=7, 8, or 9 or if racenew=97, 98, 99				
meduc	Educational attainment (similar categories as used in McMorrow, Kenney, Goin, 2014)	=. (missing) if value is ‘refused’ (97) or ‘don’t know’ (99) or ‘child under 5 years old’ (96) =1 if less than high school diploma or GED (recoded from educ values 00-12) =2 if highest level completed is high school diploma or GED (recoded from educ values 13-14) =3 if highest level completed is some college or complete associate degree (recoded from educ values 15-17) =4 if highest level competed is bachelor’s, master’s, professional, or doctoral degree (recoded from educ values 18-21)	**educ**	00=NIU, 01=Never attended/kindergarten only, 02=1st grade, 03=2nd grade, 04=3rd grade, 05=4th grade, 06=5th grade, 07=6th grade, 08=7th grade, 09=8th grade, 10=9th grade, 11=10th grade, 12=11th grade, 13=12th grade-no diploma, 14=GED or equivalent, 15=High School Graduate, 16=Some college-no degree, 17=Associate degree: occupational, technical, or vocational program 18=Associate degree: academic program 19=Bachelor's degree (Example: BA, AB, BS, BBA) 20=Master's degree (Example: MA, MS, MEng, MEd, MBA) 21=Professional School degree (Example: MD, DDS, DVM, JD) 22=Doctoral degree (Example: PhD, EdD) 97=unknown-Refused, 98=unkown-not ascertained, 99=unknown-Don't know (weight: perweight)	“’What is the HIGHEST level of school you have completed or the highest degree you have received? Please tell me the number from the card. ’ * Enter highest level of school completed.” (the numbers refer to the codes in the column to the left)	Persons age 5+

married	Marital status	=1 if legally married =. (missing) if value is 'unknown marital status' (recoded from marstat=99) =0 otherwise	**Marstat**	00=NIU, 10=married, 20=widowed, 30=divorced, 40=separated, 50=never married, 99=unknown marital status (weight: perweight)	"'Are you now married, widowed, divorced, separated, never married, or living with a partner? Is your spouse living in the household? Have you ever been married? What is your current legal marital status?'"	Persons age 14+
female	Gender-whether or not the respondent was female	=1 if answered 'female' (recode from sex=2) =0 if answered 'male' (recode from sex=1) =. If sex is missing	**Sex**	1=male 2=female (weight: perweight)	"'Are you male or female?' * If don't know or refused enter your best guess."	All persons
syear	Survey Year	=1 if survey year was 2011 (recode from year=2011) =2 if survey year was 2012 (recode from year=2012) =3 if survey year was 2013 (recode from year=2013) =4 if survey year was 2014 (recode from year=2014) =5 if survey year was 2015 (recode from year=2015)	**year**	Continuous, 2011=2011,…, 2015=2015, (no weight)	"YEAR is a four-digit variable reporting the calendar year (e.g., 2003) the survey was conducted and the data were collected. YEAR indicates the survey year reported on the household record."	All persons
povstat	Poverty status-Indicator for whether the respondent's family income (or individual income for those living alone or	=1 if family income (or individual income for those living alone or with unrelated persons) is below the poverty level	**pooryn**	1=At or above poverty threshold, 2=Below poverty threshold 9=unknown (or undefined) (weight: perweight)	"'When answering this next question, please remember to include your income PLUS the income of all family members living in this household. What is your best estimate of your total income/the total income of all family members from all sources, s, in the last calendar year (in 4 digit format)?' * Enter '999,995' if the reported income is greater than $999,995. 'Was your total [family]	All persons

	with unrelated persons) is below the poverty level or if it is at/above this level (based on family size)	(recoded from pooryn= 2) =. If missing or unknown (recode from . or 9) =0 otherwise (income at or above poverty level, recoded from pooryn=1)			income from all sources less than $50,000 or $50,000 or more? Was your total [family] income from all sources less than $35,000 or $35,000 or more? Was your total [family] income from all sources less than [fill2: fill based on poverty threshold] or [fill2: fill based on poverty threshold] or more? Was your total [fill1: family/(blank)] income from all sources less than [fill2: fill based on 200% poverty threshold] or [fill2: fill based on 200% poverty threshold] or more? Was your total [fill: family] income from all sources less than $100,000 or $100,000 or more? Was your total [fill: family] income from all sources less than $150,000 or $150,000 or more? Was your total [fill: family] income from all sources less than $75,000 or $75,000 or more?'"	
lowerinc	Lower-income Status Indicator for whether the respondent's family income (or individual income for those living alone or with unrelated persons) is less than 400% FPL or at/above 400% FPL	=1 if income less than 400% FPL (recode from ratio of family income to US Census Bureau poverty threshold for the year in question less than 4) =0 if income greater than or equal to 400% FPL (recode from ratio of family income to US Census Bureau poverty threshold for the year in question greater than or equal to 4) =.	**Poverty**	10=less than 1.0; 11 =under .5; 12 =.5-.74; 13= .75-.99; 14= less than 1.0 (no other detail); 20= 1.00-1.99; … 34=3.5-3.99; 35=4.0-4.49; … (weight: perweight)	"Was your total [fill1: family/] income from all sources less than [fill2: 250% of poverty threshold] or [fill2: 250% of poverty threshold] or more? Was your total [fill1: family/] income from all sources less than [fill2: 138% of poverty threshold] or [fill2: 138% of poverty threshold] or more? Was your total [fill1: family/] income from all sources less than [fill2: 100% poverty threshold] or [fill2: 100% poverty threshold] or more? Was your total [fill1: family/] income from all sources less than [fill2: 200% of poverty threshold] or [fill2: 200% of poverty threshold] or more? Was your total [fill1: family/] income from all sources less than [fill2: 400% of poverty threshold] or [fill2: 400% of poverty threshold] or more?	All persons
georeg	Region of residence (based on Census Bureau	=1 if 'northeast' (recode from region=01) =2 if 'north	**Region**	01=*Northeast* (Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New	"REGION reports the region of the U.S. where the housing unit containing survey participants was located. The geographic information included in REGION was added during processing, rather than	All households

	recognized U.S. regions)	central/midwest' (recode from region=02) =3 if 'south' (recode from region=03) =4 if 'west' (recode from region=04) =. if 'unknown' (recode from region=09)		Jersey, and Pennsylvania); 02=*North Central/Midwest* (Michigan, Ohio, Indiana, Illinois, Wisconsin, Minnesota, Iowa, Missouri, North Dakota, South Dakota, Kansas, and Nebraska); 03=*South* (Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida, Kentucky, Tennessee, Mississippi, and Alabama, Texas, Arkansas, Oklahoma, and Louisiana); 04=*West* (Washington, Alaska, Oregon, California, and Hawaii, Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, and Nevada); 09=*Unknown* (weight: hhweight)	ascertained via questioning."	
insured	Indicator of whether the respondent currently has any health insurance coverage or not	=1 if 'covered' (recode from hinotcove=2) =. if 'unknown' (recode from hinotcove=7, 8, or 9) =0 otherwise	**hinotcove**	1=not covered 2=covered 7=unknown-refused 8=unknown-not ascertained 9=unknown-don't know (weight: perweight)	Note: NCHS recoded variable based on response to many questions and back editing; IHIS users are encouraged to use the NCHS recoded insurance variables. "What kind of health insurance or health care coverage do you have? INCLUDE those that pay for only one type of service (nursing home care, accidents, or dental care). EXCLUDE private plans that only provide extra cash while hospitalized.' * Enter all that apply: Private health insurance, Medicare, Medi-Gap, Medicaid, SCHIP (CHIP/Children's Health Insurance Program), Military health care (TRICARE/VA/CHAMP-VA), Indian Health Service, State-sponsored health plan, Other government program, Single service plan (e.g., dental, vision, prescriptions), No coverage of any type, Refused, Don't know." Further questions to clarify and reaffirm the insurance type are asked. Original NHIS variables include HIMCAIDE, HISTATEE, HICHIPE, and HIOTHGOVE. Note:	All persons

					HIPUBCOV excludes military coverage (HIMILITE), Medicare (HIMCAREE), and Indian Health Service coverage (HIHSE).	
employed	Employment status	=1 if employed (no recode, already empstatimp1 is 1 if employed) =0 if not employed (recode from empstatimp1=2)	**Empstatimp1**	00=NIU; 1=employed, 2=not employed (weight: perweight)	NOTE: This is the first of 5 imputed employment status (dichotomous-employed or not employed) variables NHIS provides. "It was created as part of a set of variables that provide complete (i.e., without missing values) data on family income." Multiple imputation techniques are beyond the scope of this analysis, so I use the first of the 5 imputed variables instead of utilizing all 5 in multiple imputation techniques. "Beginning in 2001, persons were categorized as "employed" (IHIS code 1) under two circumstances: a) they reported working for pay in the last year; or b) they did not answer the question about working for pay in the past year and were imputed the value "employed." Conversely, for 2001 forward, persons were categorized as "not employed" (IHIS code 2) under two circumstances: a) they reported that they did not work for pay in the past year; or b) they did not answer that question but were imputed the value "not employed." "The next few questions are about employment status. Which of the following were you doing last week?" Possible answers: Working for pay at a job or business, With a job or business but not at work, Looking for work, Working but not for pay at a family-owned job or business, Not working at a job or business and not looking for work, Refused, Don't know.	Persons age 18+
poorhealth	Indicator of whether the respondent is in poor health (self-reported health status)	=1 if answered 'fair' (recode from health=4) or 'poor' (recode from health=5) =0 if answered	**health**	1=excellent; 2=very good; 3=good; 4=fair; 5=poor; 7=unknown-refused; 8=unknown-not ascertained; 9=unknown-don't know (weight: perweight)	"'Would you say your health in general is excellent, very good, good, fair, or poor?'"	All persons

		'excellent' (recode from health=1) or 'very good' (recode from health=2) or 'good' (recode from health=3) =. (missing) if answered 'unknown' or health is missing (recode from health=7, 8, 9, or .)				
uscitizen	Indicator for being a US citizen	=1 if yes-US citizen (recode from citizen= 2) =0 if no-not a US citizen (recode from citizen= 1) =. (missing) if citizen is missing or 'unknown' (recode from citizen=., 7, 8, or 9)	**citizen**	1=no, not US citizen; 2=Yes, US citizen; 7=unknown-refused; 8=unknown-not ascertained; 9=unknown-don't know (weight: perweight)	"Are you a CITIZEN of the United States?" And clarifying questions are asked as needed: "Previously, you refused to say if you were born in the United States. Would you like to change your answer to the question? Previously, you didn't know if you were born in the United States. Would you like to change your answer to the question?"	All persons
age	Age of respondent	No recode	**age**	Continuous, 00 (0)-85 (85+) This is topcoded at 85 plus (weight: perweight)	"'What is your age? And what is your date of birth? Please give month, day, and year for the date of birth.'"	All persons
agecat_papYA	Indicator for whether the respondent is age	=1 if respondent is between the ages of 21-25years	**age**	See above	See above	All persons

	21-25	=0 otherwise				
agecat_2629	Indicator for whether the respondent is age 26-29	=1 if respondent is between the ages of 26-29 years =0 otherwise	**age**	See above	See above	All persons
agecat_3039	Indicator for whether the respondent is age 30-39	=1 if respondent is between the ages of 30-39 years =0 otherwise	**age**	See above	See above	All persons
agecat_4049	Indicator for whether the respondent is age 40-49	=1 if respondent is between the ages of 40-49 years old =0 otherwise	**age**	See above	See above	All persons
Agecat_5065	Indicator for whether the respondent is age 50-65	=1 if the respondent is between the ages of 50-65 years	**age**	See above	See above	All persons