Evaluation Study Examining the Effect of the WIC Program Participation during Pregnancy on the Child Mathematics Achievement in 1997

Xinyuan Hu

School of Social Policy & Practice, University of Pennsylvania

MSSP 897: Applied Linear Modeling

Prof. Ezekiel Dixon-Román

Nov. 25, 2020

Evaluation Study Examining the Effect of the WIC Program Participation during Pregnancy on the Child Mathematics Achievement in 1997

Social policy makers have been devoted to safeguarding the health of women, infants, and children through different programs. The WIC program, serving as a nutrition program, provides low-income women, infants, and children up to age 5 who are at nutrition risk with nutritious foods, information on healthy eating, and referrals to health care (U.S. Department of Agriculture, 2019).

Using the data of the Child Development Supplement (CDS) to the Panel Study of Income Dynamics (PSID), this report examines the effect of the WIC program participation during pregnancy on child mathematics achievement (mathraw97), attempting to explore the connections between nutrition and child development from the prenatal period.

On the basis of assumption diagnostics and corrections, this report captures respectively the interaction effects between WIC program participation during pregnancy and (1) family income, (2) race, (3) the current age of the child. This allows the report not to be confined to the individual effects of the independent variables. Finally, the Conclusion section discusses the policy implications of the WIC program for child mathematics achievement and the suggestions on the methodological limitations of the study.

Methods

This section includes the description of the sample, operationalization of the study variables, and the rationale of the analyses.

Description of the Sample

The data

The Child Development Supplement (CDS) to the Panel Study of Income Dynamics (PSID) is conducted by Survey Research Center, the University of Michigan. CDS has gathered comprehensive and nationally representative information about children and their families. Its purpose is to study how economic and social differences affect child development (University of Michigan Institute for Social Research). In this report, we use the constructed dataset to perform the evaluation study.

The variables

This report examines the effect of the WIC program participation during pregnancy (independent variable: WICpreg) on the raw scores of child mathematics achievement (dependent variable: mathraw97). Race, AGE97, faminc97, bthwht and HOME97 are taken into consideration as control variables.

Variables are defined as follows:

Table 1

Definition of Variables

Variable	Definition
mathraw97	Woodcock-Johnson Revised Mathematics Achievement Test Raw Score.
WICpreg	Women, Infant and Children (WIC) Nutrition Program participant during pregnancy. (0 = No, 1 = Yes)
Race	Centered binary coding of race: -0.5 = Black, 0.5 = White.
AGE97	The child's age in 1997.
faminc97	Total family income in 1997 (in 2002 constant dollars).
bthwht	Low birth weight status of the child. (0 = non-low birth weight child, 1 = low birth weight child)
HOME97	A composite total score of the emotional and cognitive stimulation at home.

Note. Race is a center binary variable created according to variable CHRACE.

Operationalization of the Study Variables

The initial model used in the report can be written as:

$$mathraw97=\beta_0+\beta_1*AGE97+\beta_2*faminc97+\beta_3*bthwht+\beta_4*WICpreg+\beta_5*Race+\beta_6*HOME97+\varepsilon \quad (1)$$

After all the data transformations, we will estimate the following main effects model:

$$mathraw97=\beta_0+\beta_1*AGE97c+\beta_2*AGE97c^2+\beta_3*cinc+\beta_4*bthwht+\beta_5*WICpreg+\beta_6*chome+\beta_7*Race+\varepsilon \quad (2)$$

Next, we will add an interaction term each time to the main effects model to see if the effect of WICpreg on mathraw97 is moderated by cinc, race and the two age variables (AGE97c & AGE97c2). The general model of multiple regression with one interaction term is as follows:

$$Y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_1*x_2+\varepsilon \quad (3)$$

Rationale of the Analyses

The research methods used in this report is to analyze the three interactions with WICpreg separately on the basis of a rigorous main effects model. Firstly, the report diagnoses normality and general linear assumptions for the initial model then make violation corrections. At the same time, the report uses Cook's Distance and Variance Inflation Factor (VIF) to identify outliers and multicollinearity problems. Secondly, the report creates a main effects model based on which three interaction models are developed and estimated. Last, the analyses section will cover the interpretation about the estimated interaction effects through tables and graphs.

Results

There are two parts in this section, one is the descriptive statistics, the other is the regression analyses.

The Descriptive Statistics

As shown in Table 2 and Table 3, the sample size is 3563. 1715 observations (48.13%) are deleted in the initial model according to the listwise deletion. So, a total of 1848 observations are used to estimate the initial model.

Table 2

Observations and Missing Values of Variables

Variable	Obs	Missing Values
mathraw97	3563	1352
WICpreg	3563	241
Race	3563	466
AGE97	3563	1340
faminc97	3563	0
bthwht	3563	0
HOME97	3563	0

Descriptive statistical analysis of variables is shown in Table 3. The dependent variable mathraw97 has a wide range of values from 0.0 to 98.0 and a large standard deviation 22.34. This suggests that child mathematics achievement varies greatly among children. The mean of WICpreg is 0.4, which shows that more than half the families did not participate in the WIC program during pregnancy. Similarly, the mean of bthwht (Low birth weight status of the child) indicates that most children have a normal birth weight. The white children in the sample are slightly more than the black because the mean of Race is 0.02. Like mathraw97, the standard deviation (54909.07) and range (784610.6) of faminc97 are so large thus we can tell the observed values of family income are very spread-out.

Table 3

Descriptive Statistics

Variable	N	Mean	Median	Sd	Range	Min	Max
mathraw97	1848	35.98	37.0	22.34	98.0	0.0	98.0
WICpreg	1848	0.40	0.0	0.49	1.0	0.0	1.0
AGE97	1848	7.39	7.0	2.92	10.0	3.0	13.0
faminc97	1848	53452.15	43041.5	54909.07	784610.6	0.0	784610.6
bthwht	1848	0.37	0.0	0.48	1.0	0.0	1.0
HOME97	1848	20.23	20.6	3.10	19.1	7.9	27.0
Race	1848	0.02	0.5	0.50	1.0	-0.5	0.5

From Table 4, we can tell the correlation coefficients between all the variables. First, all the variables are correlated with other variables. Only WICpreg is negatively associated with mathraw97 (-0.1870) and only bthwht is positively associated with WICpreg (0.1059). AGE97 is highly correlated to mathraw97 as the absolute value of the correlation coefficient (0.9191) is close to 1, while the absolute values of other correlation coefficients are all less than 0.5 (except the values on major diagonal of Table 4).

Table 4

Correlation Matrix

	mathraw97	AGE97	faminc97	bthwht	WICpreg	HOME97	Race
mathraw97	1.0000	0.9191	0.1569	0.1378	-0.1870	0.3107	0.1080
AGE97	0.9191	1.0000	0.0526	0.2214	-0.0890	0.1921	-0.0161
faminc97	0.1569	0.0526	1.0000	-0.0839	-0.3863	0.3958	0.3643
bthwht	0.1378	0.2214	-0.0839	1.0000	0.1059	-0.0879	-0.1672
WICpreg	-0.1870	-0.0890	-0.3863	0.1059	1.0000	-0.4160	-0.4905
HOME97	0.3107	0.1921	0.3958	-0.0879	-0.4160	1.0000	0.4484
Race	0.1080	-0.0161	0.3643	-0.1672	-0.4905	0.4484	1.0000

The Regression Analyses

The Regression Analyses have 3 subsections including assumption diagnostics and violation corrections, outliers and multicollinearity and the regression analyses of main effects model and interaction models.

Assumption Diagnostics and Violation Corrections

Normality of Continuous Random Variables and Assumptions of the General Linear Model. If the normality assumption is not justified, the significant test result may well be entirely spurious. To get the Best Linear Unbiased Estimates (B.L.U.E.), it is necessary to check if the assumptions of the general linear model are violated or not. This study examines the normality and linearity of the model.

From the normal Q-Q plots below, because the points do not follow red lines, family income in 1997 (faminc97) and child age in 1997 (AGE97) do not conform to the normality assumptions (see Figure 1). This requires proper transformations of these two variables (see the Data Transformations and Re-specification part later). According to the loess curves of the Figure 2, faminc97 and AGE97 are not linearly associated with mathraw97.

Figure 1

Histograms and Normal Q-Q Plots of faminc97, AGE97 and HOME97

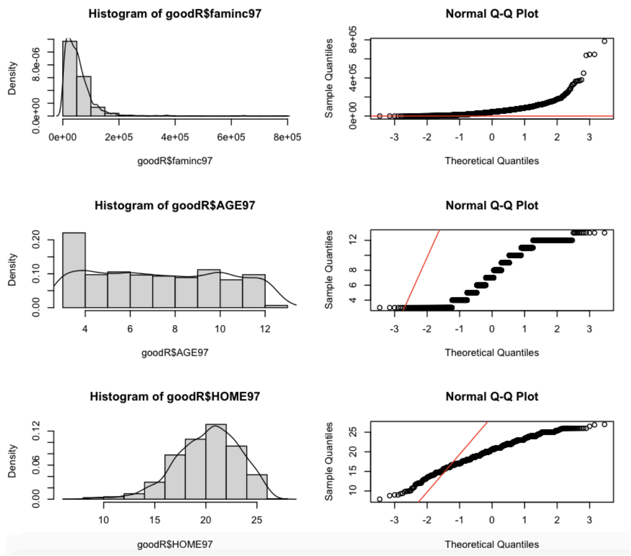
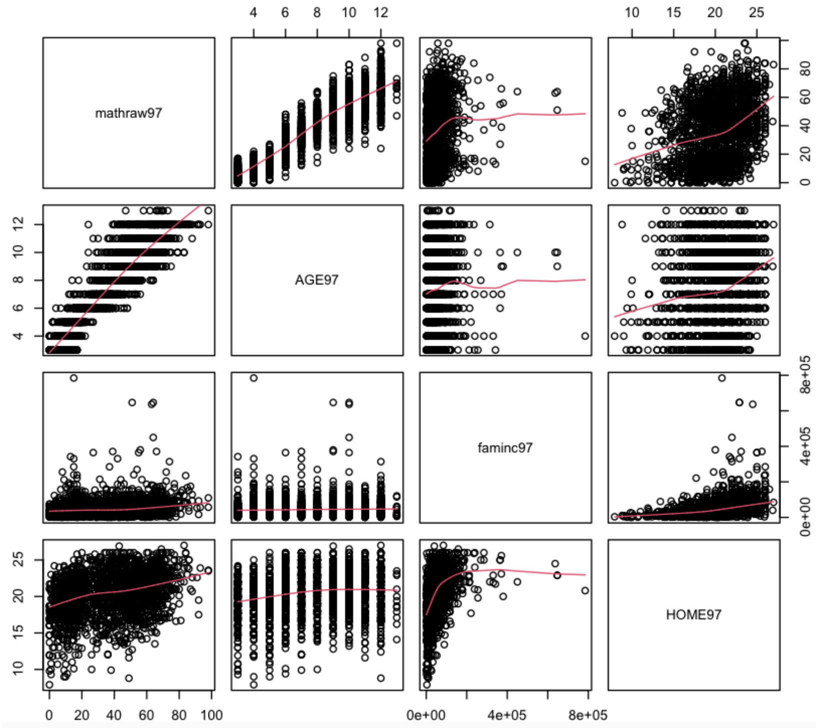


Figure 2

Scatter Plot Matrix of each IV against DV and other IVs



Data Transformations and Re-specification. I took a logarithmic transformation on faminc97, centered AGE97 and created a squared AGE97 variable.

After the data transformations, the re-specified model becomes:

$$mathraw97=\beta_0+\beta_1*AGE97c+\beta_2*AGE97c^2+\beta_3*logfaminc+\beta_4*bthwht+\beta_5*WICpreg+\beta_6*HOME97+\beta_7*Race+\varepsilon \quad (4)$$

The p- values for the coefficient of $AGE97c^2$ in the equation (4) is less than 0.05 (see Table 5), which indicates that there is indeed a curvilinear relationship between age and mathraw97 and we need to keep the re-specified model above.

Table 5

Regression Results for Model after Violation Corrections

	Model after Violation Corrections

	Estimate	p-value
(Intercept)	20.3384	<.05***
WICpreg	-1.3031	<.05**
AGE97c	6.9590	<.05***
AGE97c^2	-0.1018	<.05***
logfaminc	0.5548	<.05**
bthwht	-1.4794	<.05***
Race	2.4919	<.05***
HOME97	0.5792	<.05***
F-statistic	1796	<.05***
R^2	0.8723	

Note. ***,**,* denote the significance level 0.1%, 1%, 5%.

Homoscedasticity and Normality of Residuals. By plotting the residual values against the fitted values, we can assess the homoscedasticity assumption to see how the variance of the residuals change with the independent variables. The loess curve of the residuals in Figure 3 basically follows a horizontal line, which shows the assumption of homoscedasticity is met. Residuals that are not normally distributed may yield biased standard errors. From the normal Q-Q plot of the residuals (see Figure 4), the residuals meet the normality assumption.

Figure 3

Scatter Plot of Fitted Values against Residuals

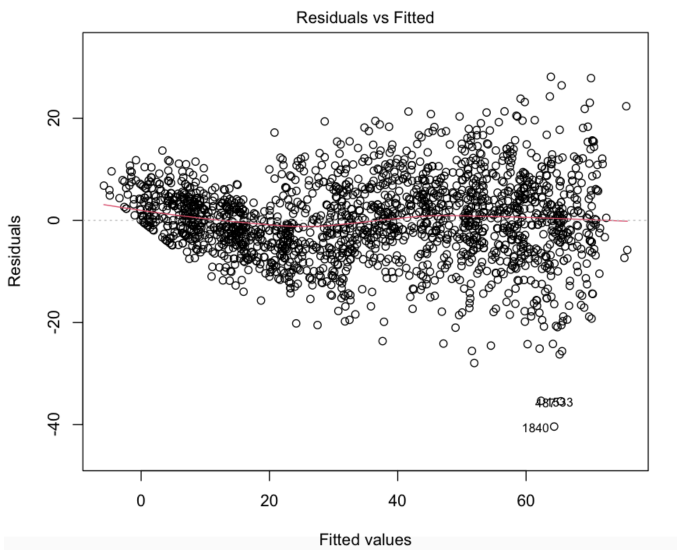
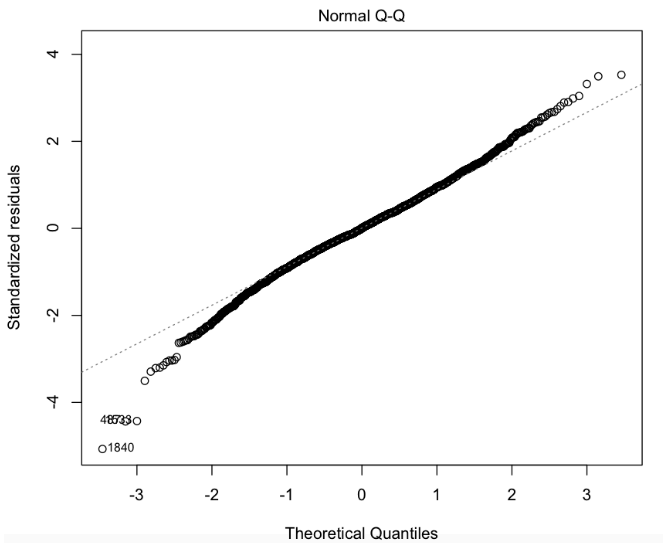


Figure 4

Normal Q-Q Plot of Residuals



Outliers and Multicollinearity

Outliers. This study uses the rules of thumb of Cook’s Distance identify outliers. 5 outlier observations were deleted which are greater than 95th percentile of the extreme values for Cook's D (4/N, where N=1848).

Multicollinearity. Here the Variance Inflation Factor (VIF) is used to detect multicollinearity which would influence the regression coefficient and inflate the standard errors of the collinear variables and potentially other predictors. The VIFs for the independent variables are all less than 10 (see Table 6), so the re-specified model doesn't suffer from multicollinearity problems.

Table 6

VIFs for Independent Variables

Variable	VIF
WICpreg	1.497643
AGE97c	1.116801

AGE97c2	1.107665
logfaminc	1.444223
bthwht	1.182956
HOME97	1.537022
Race	1.536842

The Regression Analyses of Main Effects Model and Interaction Models

First, new centered variables for all of the continuous variables are created by subtracting the mean value from each observation, which makes the interpretation easier. The three interaction terms are also created based on the centered variables. All the new centered variables are shown in Table 7.

Table 7

New Centered Variables

Variable	Definition
chome	HOME97-mean
cinc	logfaminc-mean
cincWIC	cinc*WICpreg
agecWIC	AGE97c*WICpreg
agec2WIC	AGE97c2*WICpreg
raceWIC	Race*WICpreg

After centering, we got the main effects model (Model 1):

$$mathraw97=37.74+6.97*AGE97c-0.10*AGE97c^2+0.70*cinc-1.44*bthwht-1.11*WICpreg+0.60*chome+2.34*Race \quad (5)$$

In the main effects model, as shown in the Table 7, the p-value for the F-statistic is less than 0.05, showing that at least one of the independent variables helps explain the variation in mathraw97. Approximately 88% of the variance in mathraw97 can be explained by the independent variables. The regression coefficients are all statistically significant different from zero because p- values for the independent variables are all less than 0.05.

- Intercept term: The coefficient 37.74 means that the expected math score for an "average" black child is 36.57 (37.74-2.34*0.5=36.57) and the expected math score for an "average" white child is 38.91 (37.74+2.34*0.5=38.91). An "average" child is a child who has the average age (about 7 years old), whose mother did not take part in the WIC program, who comes from a family with the average income (logged family income) and who was born with a normal weight.
- WICpreg: The coefficient -1.11 means that on average children with family participation in WIC program during pregnancy are predicted to have 1.11 points lower mathematics achievement scores than children whose families don't participate in the program, holding all other independent variables constant.
- AGE97c & $\mathrm{AGE97c}^2$: The coefficient 6.97 for AGE97c tells us that math scores increase on average 6.97 points for each year increase in age above the average age (which is approximately 7 years of age). The coefficient -0.10 for $\mathrm{AGE97c}^2$ indicates that this increase in math scores decelerates by 0.20 (-0.10 * 2) points for each additional year. That is, math scores increase at a decreasing rate with age.
- cinc: The coefficient 0.70 means that on average, a child's mathematics achievement score will increase by about 0.007 (0.70/100=0.007) points for every one percent increase in the family income above the average income, holding all other independent variables constant.
- bthwht: The coefficient -1.44 means that on average the mathematics achievement scores of children with low birth weight are 1.44 points lower than children born with a

normal weight, holding all other independent variables constant.

- chome: The coefficient 0.60 means that on the mean of child's mathematics achievement scores will increase by about 0.60 points for every additional score of parenting practices at home above the average score (which is approximately 20 points), holding all other independent variables constant.
- Race: The coefficient 2.34 means that on average the mathematics achievement scores of black children are 2.34 points lower than white children, holding all other independent variables constant.

Interaction effect of WICpreg*cinc

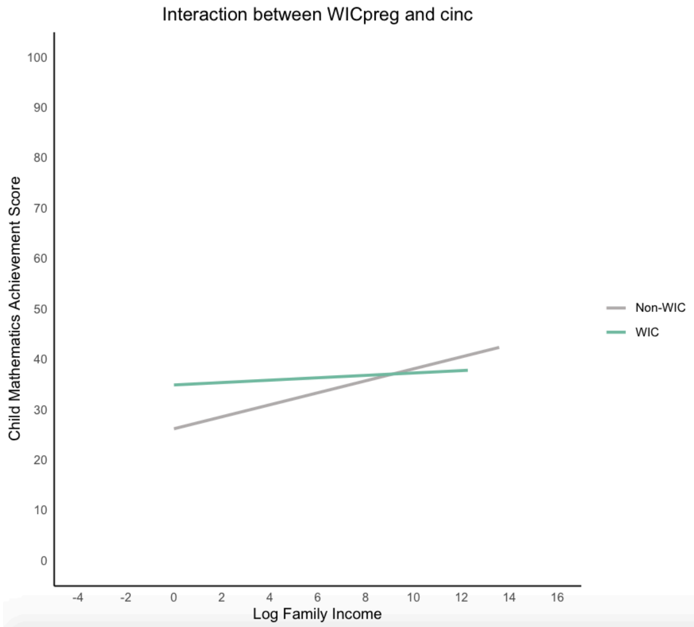
After containing an interaction term between WICpreg and cinc, we got the Model 2:

$$mathraw97=37.57+6.97*AGE97c-0.10*AGE97c^2+1.19*cinc-1.43*bthwht-1.27*WICpreg+0.59*chome+2.24*Race-0.95*cincWIC \quad (6)$$

In this interaction model, the p-value for the cincWIC is significant ($p<0.05$), meaning that there is evidence that the impact of WICpreg on children's math scores depends on the level of family income. The parameter estimate for WICpreg (-1.27) can be interpreted as the effect of WIC participation on math scores when cincWIC=0. On average, children whose mothers participated in the WIC program during pregnancy have math scores that are 1.27 points lower than children whose mothers did not participate in the program. The parameter estimates for cinc and cincWIC can explain the moderator effect. For every 1-unit increase in the logged family income, WIC participants' children will only earn an additional 0.24 points, while a child whose mother was not a participant will earn 1.19 points. We can see this disparity in Figure 1 through the fact that the slope of the Non-WIC line is bigger than the slope of the WIC line. In Figure 5, the x-axis shows the normal scale of the logged family income. The existence of the intersection point is because the coefficient for cincWIC and the values of cinc at the beginning are all negative, thus the predicted mean for the WIC group is larger.

Figure 5

Plot of Interaction Effect between WICpreg and cinc



Note. Figure 5 controls for all the other covariates: AGE97c=0, $AGE97c^2$=0, chome=0, Race=0, bthwht=0.

Interaction effect of WICpreg*Race

After containing an interaction term between WICpreg and Race, we got the Model 3:

$$mathraw97=37.56+6.97*AGE97c-0.10*AGE97c^2+0.69*cinc-1.42*bthwht+-1.32*WICpreg+0.59*chome+3.17*Race-2.15*raceWIC \quad (7)$$

In this interaction model, the p-value for the raceWIC is significant ($p<0.05$), meaning that there is evidence that the impact of WICpreg on children's math scores depends on race. Since WICpreg and Race are all categorical variables with 2 levels each, we have four predicted means (see Table 8). Specifically, Black children whose mothers participated in the WIC program will have math scores that are 0.25 points lower than black children whose mothers did not participate in the program. White children whose mothers were WIC participants will have math scores that are 2.40 points lower than white children whose mothers were not WIC participants. Among the WIC participants' children, the math scores of the black are 1.02 points lower than the white. Among the children whose mothers did not participate in the program, the mathematics learning performance of the black is relatively worse with 3.17 scores lower than the white.

Table 8

Regression Results

	WIC	Non-WIC	Difference
Black	35.72531	35.97336	-0.24805
White	36.74558	39.14352	-2.39794
Difference	-1.02027	-3.17016	

Interaction effect of WICpreg*age

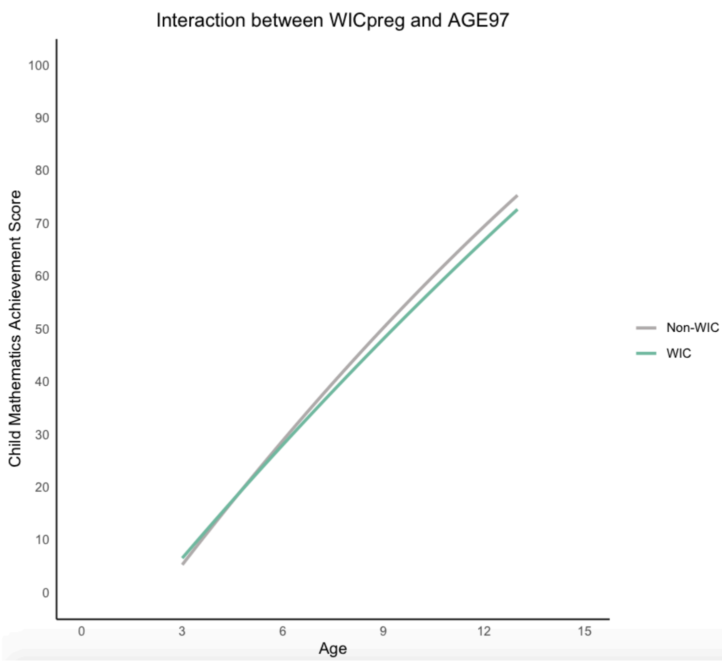
After containing two interaction terms between WICpreg and age, we got the Model 4:

$$mathraw97=37.89+7.15*AGE97c-0.12*AGE97c^2+0.73*cinc-1.41*bthwht-1.50*WICpreg+0.60*chome+2.33*Race-0.44*agecWIC \quad (8)$$

The interaction effects between WICpreg and the two age variables are assumed to be curvilinear interactions. However, only the p-value for the agecWIC is significant ($p<0.05$), the p-value for agec2WIC is greater than 0.05. But it can still indicate that the impact of WICpreg on children's math scores depends on the child age. The parameter estimate for WICpreg (-1.50) means the effect of WIC participation on math scores when agecWIC=0. On average, children whose mothers participated in the WIC program during pregnancy have math scores that are 1.50 points lower than children whose mothers did not participate in the program. The parameter estimates for AGE97c and agecWIC can explain the moderator effect. For every additional year in age above the average age, WIC participants' children should only expect to earn an additional 6.71 points, while a child whose mother was not a participant should expect to earn 7.15 points. We can see this disparity in Figure 2 through the fact that the slope of the Non-WIC line is bigger than the slope of the WIC line. In Figure 6, the x-axis shows the normal scale of the child age. The existence of the intersection point is because the coefficient for agecWIC and the values of AGE97c at the beginning are all negative, thus the predicted mean for the WIC group is larger.

Figure 6

Plot of Interaction Effect between WICpreg and AGE97c



Note. Figure 6 controls for all the other covariates: chome=0, cinc=0, Race=0, bthwht=0.

Table 9

Regression Results

	Model 1	Model 2	Model 3	Model 4
	Estimate	Estimate	Estimate	Estimate
(Intercept)	37.7440 (<.05***)	37.5701 (<.05***)	37.5584 (<.05***)	37.8927 (<.05***)
WICpreg	-1.1139 (<.05*)	-1.2666 (<.05**)	-1.3230 (<.05**)	-1.4984 (<.05*)
AGE97c	6.9653	6.9705	6.9662	7.1506

	(<.05***)	(<.05***)	(<.05***)	(<.05***)
$AGE97c^2$	-0.0963 (<.05***)	-0.0961 (<.05***)	-0.0954 (<.05***)	-0.1207 (<.05***)
cinc	0.7033 (<.05***)	1.1906 (<.05***)	0.6908 (<.05***)	0.7252 (<.05***)
bthwht	-1.4383 (<.05***)	-1.4288 (<.05***)	-1.4250 (<.05***)	-1.4102 (<.05***)
Race	2.3411 (<.05***)	2.2427 (<.05***)	3.1702 (<.05***)	2.3294 (<.05***)
chome	0.5995 (<.05***)	0.5913 (<.05***)	0.5906 (<.05***)	0.6010 (<.05***)
cincWIC	——	-0.9531 (<.05**)	——	——
agecWIC	——	——	——	-0.4426 (<.05***)
agec2WIC	——	——	——	0.0418 (>.05)
raceWIC	——	——	-2.1499 (<.05*)	——
F-statistic	1841 (<.05***)	1617 (<.05***)	1616 (<.05***)	1441 (<.05***)
R^2	0.8753	0.8758	0.8758	0.8762

Note. P-value is shown inside the parentheses; ***,**,* denote the significance level 0.1%, 1%, 5%.

Conclusion

This study attempted to evaluate the effect of the WIC Program participation during pregnancy on child mathematics achievement, focusing on if the effect of the program is moderated by (1) family income, (2) race, and (3) the child age. Results indicated that the three interactions are all significant, which means that we need to take into account the joint effect of WICpreg and cinc, Race or AGE97c over and above their separate effects.

According to the multiplicative effects of family income and age on the WIC program participation, we can find that the children whose mothers did not participate in the WIC program will earn more additional points than the children whose mothers participated in the program with the every 1-unit increase in income or age. As the family income or age of WIC participants' children increases, the potential of marginal increase in the math scores is less than those children whose mother were not WIC participants. From the previous reports, we have already known that the overall conditions of the WIC applicant families are worse than average. For example, their gross income must fall at or below 185 percent of the U.S. Poverty Income Guidelines (USDA, 2019). In this situation, we can't expect the children in such families to have high learning abilities and qualities to support their improvement of mathematics achievement. Besides, it is possible that nutrition provided by the WIC program mainly related to the improvement of child's other abilities, rather than mathematical abilities which can be measured by mathematics achievement scores. Thus, the disparity of math scores between WIC and non-WIC groups across income levels or age may not well reflected the real effect of the WIC program.

As for the interaction effect between WICpreg and Race, the effect of the WIC program on black children seems to be more effective than the effect on white children. This is shown in statistical results that the difference between WIC and non-WIC groups of black children is smaller than the difference between WIC and non-WIC groups of white children. The underlying mechanisms need further study.

Although this report gives out a rigorous regression model with interaction terms , there are still many problems to be solved. Firstly, it does not cover the assumption diagnostics about reliability of predictors and independence of residuals (including auto-correlation and clustered data). Secondly, the control variables for the analysis models are limited. What's more, the interactions between the WIC program participation and other social and demographic characteristics might be also necessary to be captured.

References

University of Michigan Institute for Social Research [ISR]. (n.d.). Retrieved October 7, 2020, from: https://psidonline.isr.umich.edu/CDS/cdsi_userGD.pdf

U.S. Department of Agriculture [USDA]. (2019, July 1). Retrieved October 7, 2020, from: https://www.fns.usda.gov/wic