SCHOOL of PUBLIC POLICY

19
Ludwig/Miller Head Start Evaluation

Douglas J. Besharov
Peter Germanis
Caeli A. Higney
and
Douglas M. Call

September 2011

Maryland School of Public Policy
Welfare Reform Academy
www.welfareacademy.org

Part of a forthcoming volume
Assessments of Twenty-Six Early Childhood Evaluations
by Douglas J. Besharov, Peter Germans, Caeli A. Higney, and Douglas M. Call

Note: This report is open to public comments, subject to review by the forum moderator. To leave a comment, please send an email to welfareacademy@umd.edu or fill out the comment form at http://www.welfareacademy.org/pubs/early_education/chapter19.html.

19
Ludwig/Miller Head Start Evaluation

Head Start began in 1965 with the goal of helping low-income children achieve school readiness "by enhancing the social and cognitive development of children through the provision of educational, health, nutritional, social and other services to enrolled children and families."[1] Currently, the program serves approximately 905,000 children at a cost of around $7 billion per year.[2]

Jens Ludwig of the University of Chicago and Douglas Miller of the University of California (Davis) used a regression-discontinuity design to estimate the impact of Head Start funding and participation on the health and educational attainment of children enrolled in the program between 1965 and 1983.[3] The authors report a 50 to 100 percent increase in levels of Head Start participation and funding in the 228 poorest counties in the country as compared to the 349 next poorest counties during this time period.[4] They also find a 33 to 50 percent decrease in mortality rates of children between the ages of five to nine living in the poorest counties due to diseases targeted by Head Start's health program. Finally, they find suggestive evidence for a positive impact of Head Start on high school completion and college attendance.

These findings are often cited as evidence that Head Start can have lasting impacts, however, their lack of generalizability to today's current Head Start population cannot be ignored. The counties included in the study were not only the poorest in the country, but were also

---

[1]U.S. Department of Health and Human Services, Office of Head Start, "About the Office of Head Start," http://www.acf.hhs.gov/programs/ohs/about/index.html (accessed October 20, 2010).

[2] Administration for Children and Families. Head Start Program Fact Sheet Fiscal Year 2010. http://www.acf.hhs.gov/programs/ohs/about/fy2010.html (accessed on October 22, 2010).

[3]Unless otherwise noted all information comes from: Jens Ludwig and Douglas L. Miller, *Does Head Start Improve Children's Life Chances? Evidence From a Regression-Discontinuity Design*, NBER Working Paper 11702 Cambridge, MA: National Bureau of Economic Research (October 2005), Jens Ludwig and Douglas L. Miller, *Does Head Start Improve Children's Life Chances? Evidence From a Regression Discontinuity Design*, IZA Discussion Paper No. 2111 Bonn, Germany: Institute for the Study of Labor (May 2006) and Jens Ludwig and Douglas L. Miller, "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design," *Quarterly Journal of Economics* 122 no.1 (February 2007): 159-208.

[4]While Ludwig and Miller generally refer to the studied counties as the 300 poorest and next 300 poorest counties, we have chosen to use the actual number of counties contained in each group.

primarily located in the South. During the course of the time period studied many other social programs, such as Medicaid and WIC, were implemented that presumably would have influenced health and education outcomes as well. Moreover, the study measures the impacts of the Head Start program as it existed over 40 years ago. There have been many programmatic changes during this time, such as the increasing use of commercially available early childhood curriculums and focus on school readiness.

Additionally, the nonparametric analytic approaches utilized by the authors created large standard errors, required large sample sizes and caused them to choose a large bandwidth in order to minimize bias and increase the precision of their estimates. The size of the bandwidth seems to indicate that there was a lack of data right at the cut-point. This is problematic given that regression discontinuity designs are contingent on this point. This is the point used to approximate randomization and that serves to highlight the differences between the program and comparison groups.

Finally, the authors were also interested in examining whether there were differential impacts on blacks and whites, but the large standard errors of these estimates made the findings non-significant. These methods also require large sample sizes to achieve precision, which leads to questions regarding the use of NELS data as an estimator for educational achievement. The NELS data contained relatively small sample sizes and also produced estimates with large standard errors. There is some indication that this was not an ideal data set to use for the questions being studied by the authors.

Program Design

Program group. Before Head Start began, in 1965, there was a push to publicize the program and to promote the participation of those areas of the country that were most in need. The Office of Economic Opportunity (OEO) sought to achieve these goals by providing assistance to the 300 poorest counties in the country in securing Head Start funding. In the spring of 1965, the OEO sent interns to each of the targeted counties with the directive of helping would be local Head Start staff members with their funding proposals.

The OEO used data from the 1960 Census in order to determine which counties were the 300 poorest and found that the poorest county had a poverty rate of about 59 percent.[5] In the studies done by Ludwig and Miller, those counties with poverty rates 20 percentage points higher than the cutoff of 59.198 were considered the poorest, while the next poorest counties had poverty rates within 20 percentage points below the cutoff. This resulted in a sample of 228 "program" (or poorest counties) and 349 "comparison" (or next poorest counties). Additionally,

[5]The authors note, "'poverty' had not yet been officially defined by the federal government at the time of the 1960 Census. OEO apparently relied on a special 1964 reanalysis of the 1960 Census data by the Census Bureau using the then-new federal poverty definition." Ludwig and Miller, 2007, 167.

the majority of the poorest counties included in the study were located in the South.[6] As reported by the 1990 Census, the 228 program counties were 1.7 percent urban, 26.6 percent Black and had a per capita income of $8488. The 349 comparison counties were 2.5 percent urban, 16.3 percent Black and had a per capita income of $9520. Since the data in this study are aggregated at the county level, there is little reported regarding the demographic or baseline characteristics of the children living in these counties who would have been enrolled in Head Start at the time.

Services. When Head Start was first established, it consisted of a few hours a day of center-based education during the summer just before Kindergarten. The education services offered in those early years varied greatly from site to site and were not based on any set curriculum. According to Louise B. Miller, formerly a professor of psychology at the University of Louisville who wrote about the program in the 1970's, "There has never been a specific curriculum model that could properly be called the Head Start curriculum […] No administrator or researcher has ever been in a position to answer the question of what most Head Start centers were like."[7]

In the early years of the program, Head Start teachers were provided with a pamphlet outlining expectations for students such as, "Learn to work independently and play independently, able to accept both help and direction from adults […] play with, and come to comprehend, mathematical concepts like sequence […] grow in ability to express inner creative impulses."[8] This is in contrast to Head Start, as it exists today, which is more focused on the goal of school readiness. The curriculum used in centers now must "support each child's cognitive and language development, including emergent literacy skills" and must adhere to Head Start's performance standards.[9] Additionally, Head Start centers now have access to published curriculums, such as The Creative Curriculum® or HighScope.

Head Start also had a health component, which "included a medical and dental evaluation

---

[6]The authors specify that "one-third of the 300 poorest counties in 1960 were in Mississippi, Kentucky, or Georgia. Almost all of the 300 poorest counties were in just ten states (the others are Alabama, Arkansas, Louisiana, North and South Carolina, Tennessee, and Texas). These ten states account for over two-thirds of the 300 'control' counties (1960 poverty rates that rank from three hundred first to six hundredth in the United States), with most of the rest in Florida, Oklahoma, Virginia, or West Virginia." Ludwig and Miller, 2007, 164.

[7]Louise B. Miller, "Development of Curriculum Models in Head Start," *Project Head Start: A Legacy of the War on Poverty*, eds. Edward Zigler and Jeanette Valentine (New York: The Free Press, 1979), 195.

[8]Jeannette Galambos Stone, "General Philosophy: Preschool Education Within Head Start," *Project Head Start: A Legacy of the War on Poverty*, eds. Edward Zigler and Jeanette Valentine (New York: The Free Press, 1979), 172.

[9]United States General Accounting Office, "Head Start: Curriculum Use and Individual Child Assessment in Cognitive and Language Development," Report to Congressional Requesters (September 2003), 1.

for every child as well as the provision of a psychological evaluation when deemed appropriate."[10] In 1967, the health component was expanded to include care for any health problems found in medical or dental evaluations, and also "preventive services such as immunizations and dental fluoride applications."[11] Also included as a health component was the provision of nutritional meals and snacks. Additional elements of Head Start included parent involvement, social services, and mental health services.

The Evaluation. Ludwig and Miller used a regression discontinuity design to investigate the impacts of OEO's grant-writing assistance to the 228 program counties. They used the large discontinuity in Head Start funding and participation across the program and comparison counties to investigate the program's long-term effects on health and education outcomes. Specifically, they compared rates of child mortality, high school completion and college attendance between the 228 program counties and the 349 comparison counties.

The authors used a variety of data sets, all of which were aggregated at the county level. This meant that the county, rather than individuals, served as the primary unit of analysis. Data on Head Start and federal expenditures were obtained from OEO files of the National Archives and Records Administration (NARA). The authors relied on county-level data from the Vital Statistics Compressed Mortality Files (CMF) to estimate the impact of Head Start funding on child mortality rates due to causes addressed in the health services component of the program. Specifically, they assessed the mortality rates of children between the ages of five to nine during the years 1973 to 1983. The logic of using this particular cohort is that they would have been eligible for participation in Head Start after the program had begun. Additionally, it should be noted that the deaths attributable to Head Start-related causes accounted for less than 10% of all deaths and did not include deaths due to accidents or injuries.

The authors relied on Census data from 1960 through 2000 to examine the effects of Head Start participation and funding on high school graduation and college attendance. The 1990 Census included a special tabulation that provided more detailed information on educational attainment by gender, age and race. They estimated schooling attainment for different birth cohorts: those that would have been directly treated by Head Start, those that might have been indirectly impacted by Head Start due to having an enrolled sibling or child, and those that would have been too old for any type of participation in Head Start. The main cohorts of interest were those that would have been directly treated by Head Start (age 18-24 in 1990).

In order to control for the possible effects of selective migration across counties, the

[10]Julius B. Richmond, "Head Start, A Retrospective View: The Founders," *Project Head Start: A Legacy of the War on Poverty*, eds. Edward Zigler and Jeanette Valentine (New York: The Free Press, 1979), 121.

[11]A. Frederick North, Jr., "Health Services in Head Start," *Project Head Start: A Legacy of the War on Poverty*, eds. Edward Zigler and Jeanette Valentine (New York: The Free Press, 1979), 233.

authors sought to replicate their Census findings using data from the National Education Longitudinal Study (NELS), which allowed them to "identify county of residence for respondents in eighth grade."[12] NELS began in 1988 with a sample of nationally representative 8th graders. Students were administered a range of surveys and cognitive tests and were tracked with periodic follow-ups through the year 2000. This allowed for data collection during high school, after high school and beyond.[13] The NELS sampled students from 1,052 schools, selecting 26 students from each school. For the purposes of their study, Ludwig and Miller aggregated the NELS data to the county level by calculating the average outcome within the county for each county and NELS respondent.

Since Ludwig and Miller were only interested in examining county effects using individuals residing in 568 counties, the number of students available for inclusion in their evaluation was limited to 649 students from the 228 program counties and 674 students from the 349 comparison counties. Notably, only 17 of the program counties and 28 of the comparison counties participated in the NELS. It is not clear how these counties were distributed across the range of poverty rates reported. The authors used measures of academic achievement obtained from the cognitive tests administered in 1988 and measures of schooling attainment and job outcomes obtained from data collected during the final NELS follow-up in 2000. Additionally, the NELS provided data on child enrollment and participation in Head Start obtained from parent reports.

Major Findings

The authors found a 50-100 percent difference in Head Start funding and participation between the 228 program and 349 comparison counties. This amounted to an increase in spending of more than $100 per four-year-old child in the poorest counties as compared to the next poorest counties ($288 vs. $134) and a participation rate of about 40% in the program counties. This implies that three-fourths of the poor children living in the program counties participated in Head Start. Moreover, they reported a 33 to 50 percent decrease in mortality rates for children living in the program counties. Importantly, the discontinuity in Head Start funding also had a positive effect on educational attainment.

Cognitive. The only cognitive measures studied were eighth-grade reading and math assessments completed during the base year 1988 NELS. The authors did not find any significant difference in these scores between the program counties and the comparison counties.

[12]Ludwig and Miller, 2007, 192.

[13]Institute of Education Sciences, National Center for Education Statistics. National Education Longitudinal Study of 1988, NELS:88 Design. http://nces.ed.gov/surveys/nels88/design.asp (accessed on October 21, 2010).

School Performance. Ludwig and Miller sought to address the question of whether Head Start had any long-term impact on educational attainment by examining high school graduation and college attendance rates.

High School Completion. Overall, discontinuities in Head Start spending led to higher rates of educational attainment for persons living in the program counties. The 1990 Census results show a 3 to 4 percentage point increase in high school completion rates for those in the program counties who would have been enrolled in Head Start. The impacts are fairly equal across blacks and whites, and males and females. However, the findings for blacks are a bit diminished due to the large standard errors. Ludwig and Miller posit that part of Head Start's influence might have been to increase the number of students graduating from high school as opposed to completing high school with a GED. The data from the NELS presents a similar pattern, showing a discontinuity of one half-year of schooling between the comparison and program counties.

College Attendance. The higher level of Head Start spending in the poorest counties also had a positive impact on college attendance, leading to a 15 percent increase in the likelihood of attending some college.[14] The estimates for college attendance were around the same magnitude as they were for high school completion. The NELS data for college attendance was similar to the results for high school completion, with an overall increase of about five years of schooling. The authors did not measure college completion rates for the Census or NELS cohorts. Presumably, this would have been a stronger measure of Head Start's effects on educational attainment.

Socialemotional development. Relevant tests apparently not administered or results not reported.

Health. Ludwig and Miller examined deaths due to Head Start "susceptible" causes (e.g. diseases that could have been prevented due to the immunizations or health screenings provided in the health services component of the program). They report a decrease of 33-50 percent in mortality rates for children age five to nine who would have attended Head Start during the period from 1973 to 1983. This equates to 1 or 2 fewer deaths per 100,000 four-year-old children.

Given that the difference in mortality rates could have been due to factors unrelated to Head Start, the authors also examined mortality rates for injuries and accidents and found no differences between the program and comparison counties. There were also no significant differences in mortality rates for the periods before the inception of Head Start.

[14]Jens Ludwig and Deborah Phillips, "The Benefits and Costs of Head Start," *Social Policy Report* Volume XXI, 3 (2007): 4.

Ludwig and Miller were not able to make any definitive conclusions regarding the impacts of Head Start on health outcomes across different populations of children as the standard errors of the estimates for blacks were very large, which could make the estimates inaccurate.

Behavior. Data apparently either not collected or not reported.

Crime/delinquency. Data apparently either not collected or not reported.

Early/nonmarital births. Data apparently either not collected or not reported.

Economic outcomes. Data apparently either not collected or not reported.

Effects on parents. Program effects for parents were not a primary interest of the study, but their educational attainment was measured using Census data. Parents were included in the cohorts of those who could have been indirectly treated by Head Start. The authors present findings from the 1970 Census and break the indirectly treated group into two birth cohorts, age 18 to 24 and age 25 and older. They estimate that about one-third of poor Southern people in these age groups would have had Head Start eligible children in their households during this time. In 1970, there was a positive discontinuity in high school completion for the younger cohort (age 18-24) but not for the older cohort (25 or older). Additionally, the 2000 Census results reveal a positive jump in college attendance of about 2-3 percentage points for those between the ages of 45 and 64-years-old.

Benefit-cost findings. Ludwig and Miller provide very rough estimates on the potential benefits provided by Head Start participation. They estimate the extra Head Start spending at $400 (in 2003 dollars) per four-year-old child and a $120 per child benefit due to "mortality improvements." This increased spending could amount to a 2 percent increase in lifetime earnings or $15,000 in present value.[15] Additionally, the authors suggest that the additional 0.5 years of education obtained by those in the program counties could amount to a 5 to 10 percent increase in earnings. The authors also suggest that if their estimates were accurate, then the program's benefits would exceed its costs.

Overall Assessment

The studies conducted by Ludwig and Miller attempted to estimate the impacts of Head Start on poor children's educational and health outcomes through the use of a regression discontinuity design. The discontinuity arose in the higher level of Head Start funding provided to the 228 program counties as compared to the 349 comparison counties. A clear point of concern

[15]Jens Ludwig and Deborah Phillips, "The Benefits and Costs of Head Start," *Social Policy Report* Volume XXI, 3 (2007): 5.

is selection bias, given that there could have been a myriad of county-level differences that could have accounted for the reported differences. Additionally, the data sources used to estimate educational attainment may have led to issues with selective migration (Census) and to weakened estimates based on small sample sizes (NELS).

Program theory. Ludwig and Miller describe Head Start a source of early intervention for low-income children, providing them and their parents with educational, health and social services. They suggest that "interest" in Head Start is based on its potential to address some of the racial and socioeconomic disparities observed in young children's "cognitive and non-cognitive skills."[16] The authors posit that the program "could affect parent outcomes and change the stream of investments they make in their children over the life course. And the nutrition and medical services provided by Head Start could improve child health […] and may also affect later economic outcomes."[17] Beyond this, no specific theory is detailed beside the general expectation that early intervention programs promote school readiness and improve developmental outcomes for children.

Program implementation. Ludwig and Miller do not give a description of how Head Start programs were implemented at funded sites. Rather, they provide data detailing the ways in which programs were first funded, with the OEO providing grant-writing assistance to the poorest 228 counties in the country. This assistance translated into a 50-100 percent increase in Head Start funding between the poorest counties and the next poorest counties.

Assessing randomization. Random assignment was not utilized in this evaluation.

Assessing statistical controls in experimental and nonexperimental evaluations. There are a number of methodological issues with the Ludwig and Miller study. Ludwig and Miller use fairly large confidence intervals (reflected in the large standard errors in the model) to achieve statistical significance.[18] These large intervals create a lot of variation in the model and increase the likelihood of some scores not falling in the band of estimates. In fact, some of the bands are so wide that they include data points that would not be significantly different than the comparison group. The 228 program counties show much greater variability in their scores of educational attainment and in some cases, mortality rates. This could be due to the fact that this was a smaller sample of counties, as the variation was not as obvious for the 349 comparison counties. This variability and sensitivity to sample size points to one potential flaw in using

[16]Ludwig and Miller, 2007, 160.

[17]Ludwig and Miller, 2007, 160.

[18]For depictions of their wide confidence intervals, see Jens Ludwig and Douglas L. Miller, *Does Head Start Improve Children's Life Chances? Evidence From a Regression Discontinuity Design*, IZA Discussion Paper no. 2111 (Bonn, Germany: Institute for the Study of Labor, May 2006), figures 1, 2, 4, and 6.

aggregate data rather than individual data.

The large standard errors observed by Ludwig and Miller are largely due to the study's design. Regression discontinuity assigns subjects to treatment or control conditions on the basis of a cutoff score. In order to get a large sample, researchers often have to increase the scope of scores surrounding the cutoff and this process increases the standard errors of the estimated treatment effects.[19] Additionally, "RD designs produce estimates of program effects that are less efficient (i.e., the standard error of the estimate of the program effect is larger) than estimates of program effects resulting from true experiments."[20] This is one methodological weakness of natural experiments.

There also seems to be a lack of data right at the cut-point, evidenced by the graphs showing a range of poverty scores that is quite large, between 40 and 80 percent. The core component of regression discontinuity analysis is the cut-point. This is the point that determines whether groups or individuals will be assigned to treatment or control, a process that approximates random assignment. According to Howard S. Bloom of the MDRC, "as one approaches the cut-point, the resulting treatment group and control group become increasingly similar in all ways except for receipt of treatment. Hence, at the cut-point, assignment to treatment by ratings is like random assignment to treatment."[21]

Ludwig and Miller use two nonparametric approaches to analyze their data, local linear regression and weighted-kernel regressions. There are a few issues to consider about the latter type of nonparametric estimate: it can cause biased estimates of treatment, is contingent on large sample sizes, and is sensitive to the size of chosen bandwidths.[22] There is no formalized method for choosing bandwidths, however, bandwidths that are too wide introduce bias and bandwidths that are too narrow decrease precision. The authors chose to use a bandwidth range of 9 to 36, with a "focal preferred" bandwidth of 18. This range of bandwidths, allowed them to stay close to the cutoff and to yield enough data to obtain meaningful estimates. However, the results are reported using the bandwidth of 18.

---

[19]Donald P. Green, Terence Y. Leong, Holger L. Kern, Alan S. Gerber and Christopher W. Larimer, "Testing the Accuracy of Regression Discontinuity Analysis Using Experimental Benchmarks," *Political Analysis* 17 (2009):400-417.

[20]Marvin B. Mandell, "Having One's Cake and Eating it, Too: Combining True Experiments with Regression Discontinuity Designs," *Evaluation Review*, 32, 5 (October, 2008): 416.

[21]Howard S. Bloom, *Modern Regression Discontinuity Analysis*, MDRC Working Papers on Research Methodology (New York, MDRC, December 2009): 7.

[22]Howard S. Bloom, *Modern Regression Discontinuity Analysis*, MDRC Working Papers on Research Methodology (New York, MDRC, December 2009): 16-17.

The nature of regression discontinuity limits its external validity. As Greg J. Duncan, of the University of California (Irvine) points out:

> "the highly selective nature of natural experiments […] may reduce bias, but they suffer from a damaging loss in external validity. By burrowing deep within population data, these techniques sacrifice their population-wide perspective on causal effects […] 'Natural experiments' may generate arguably unbiased causal estimates, but often only for small and idiosyncratic subgroups of the population, such as twins or, in the case of Ludwig and Miller (2007), individuals residing in counties with poverty rates around 59%."[23]

This issue is of concern given the amount of attention the findings of this study has generated. It is often cited in reviews of Head Start, which point to the main conclusion that the program has a positive effect on health and educational outcomes.

In a related point, one potential limitation of Ludwig and Miller's study was their use of aggregate, rather than individual data to estimate Head Start's long-term impacts. During the fifteen-year period between 1965 and 1980, only between about 3 and 7 percent of all poor children were enrolled in Head Start.[24] Even if coverage was higher in the poorest counties, it is likely that a substantial number of poor children in these counties did not receive Head Start services. And coverage rates in even the poorest counties most likely declined as Head Start enrollment declined in the early 1970s, as the program shifted from being mainly a summer program to a full-year program and from being a part-time to full-time program. As Michael Rutter of the Institute of Psychiatry and King's College in London writes: "The limitations stem from reliance on countywide measures, from reliance on residence at follow up rather than at program initiation, and from lack of satisfactory data on in and out migration."[25] Ludwig and Miller address this final point in their discussion of the Census findings.

The authors acknowledge the methodological weakness found in the Census data, "One concern with these results comes from the possibility of migration across counties between when people were Head Start age (three or four) and when they are observed as adults in the decennial

[23]Greg J. Duncan, "When to Promote and When to Avoid a Population Perspective," *Demography* 45, 4 (2008): 778.

[24] Author's calculations from U.S. Census Bureau, "Table 3: Poverty Status of People, by Age, Race, and Hispanic Origin: 1959 to 2006," http://www.census.gov/hhes/www/poverty/histpov/hstpov3.html (accessed March 17, 2010); and U.S. Department of Health and Human Services, Head Start Bureau, "Head Start Fact Sheet," various years (Washington, DC: U.S. Department of Health and Human Services, various years).

[25]Michael Rutter, "Proceeding From Observed Correlation to Causal Inference: The Use of Natural Experiments," *Perspectives on Psychological Science* 2, no. 4: 390.

census."[26] As mentioned, the authors used data from the NELS base year of 1988, 1990 follow-up and 2000 follow-up to try and control for the problem of selective migration. However, the NELS data was very "noisy" and had small sample sizes at the county and individual level. Only 17 (7%) of the poorest counties and 28 (8%) of the next poorest counties participated in the NELS. Additionally, the authors note that the NELS estimates are inhibited by large standard errors. Relying on the Panel Study of Income Dynamics (PSID), there were modest rates of migration across counties for people between the ages of three and thirteen, with 86% of respondents residing in the same county at the age of three and the age of 13. However, it is not clear how many people are represented in this estimate since the sample was restricted to people with known addresses at the ages of three and thirteen and with information on adult education outcomes.

The data sources that Ludwig and Miller use are not very informative about other things that might have been happening in the poorest counties to influence the positive outcomes on child mortality and educational attainment. One possibility that Greg J. Duncan presents is that "the Washington-lawyer-led Head Start proposal development process in the 300 poorest counties led to a more general mobilization of community leaders and resources that somehow benefitted the children in ways that would not be reflected in other social spending."[27] The authors do attempt to address this concern by examining voter registration records during the period of 1968 to 1980 and do not find any discontinuities. Whether voter registration is a good measure of community mobilization is unclear. However, Duncan makes an important point, especially given the variability in the confidence intervals surrounding the estimates of the poorest counties. It seems plausible that there could have been other factors at play beyond the discontinuity in Head Start funding.

Finally, there is the consideration presented by William Gormley of Georgetown University:

> "The study's focus—on the impact of Head Start at its inception—limits our ability to draw inferences about the Head Start program today. For example, substantial increases in the child immunization rates and preventive health access since 1965 mean that the same substantial impacts on health outcomes are less likely today (the control group today receives better care, thanks to Medicaid, the Early Periodic Screening, Diagnosis and Treatment program, and the State-Children's Health Insurance Program). Similarly, child care subsidies available to children not participating in Head Start today could limit the academic benefits experienced by Head Start children when compared to control group

[26]Ludwig and Miller (2007): 191.

[27]Greg J. Duncan, "When to Promote and When to Avoid a Population Perspective," *Demography* 45, 4 (2008): 778.

children."[28]

There were other major social initiatives that were implemented along with Head Start, most notably Medicaid, which increased poor children's access to health care and routine screenings. The Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) was established in 1972, providing eligible recipients with food vouchers, nutrition, education, and health-care referrals.[29] The introduction of these two programs does not seem to be accounted for in Ludwig and Miller's analysis.

Sample Size. The authors used counties as an aggregate measure for estimating long-term impacts of Head Start. The number of counties in the program group and comparison group is relatively large, however, there are about 100 more counties in the latter group.

The sample obtained from NELS: 88 is relatively small, with only 649 individuals from the 17 program counties and 647 from the 28 comparison counties. Since these data were aggregated, the sample size is even smaller than this, with only 17 of the program counties and 28 of the comparison counties represented in the NELS data. These counties reported Head Start participation rates of 38% and 23% respectively at the time of initial data collection (1988). Given this information, it is likely that the sample obtained from NELS was not representative of the populations of the hundreds of counties included in the Census data estimates. The authors acknowledge that the small sample size did weaken the statistical power of their tests.

Attrition. Attrition is not a factor in this study since the authors investigated county-level impacts, rather than individual impacts.

Data Collection. Data was not collected for this study; rather it was compiled from large sources, such as the OEO files from NARA, 1990 Census and the 1988 NELS. There is some question as to whether the data sources used were appropriate for the study. The NELS data was particularly problematic because it relied on such a small sample of counties and, in the case of some measures, produced very "noisy" estimates. The authors also acknowledge that the NELS data was not useful in making predictions about the impacts of Head Start on adult employment. Moreover, many of the estimates derived from the NELS data had large standard errors.

Measurement issues. One of the primary measures for determining educational attainment is college attendance. Using the available data, the authors were only able to estimate impacts on college attendance and not on college completion, which would be a stronger measure

[28]William T. Gormley, Jr., "Early Childhood Care and Education: Lessons and Puzzles," *Journal of Policy Analysis and Management* 26, 3 (2007): 648.

[29]U.S. Department of Agriculture, Food and Nutrition Service. "About WIC," http://www.fns.usda.gov/wic/aboutwic/mission.htm (accessed July 7, 2011).

of educational attainment.

Head Start participation was measured by parent reports taken during the base year of data collection for the NELS, when their children were enrolled in eighth grade. The main issue with this data is that parents might have poor recall of events that occurred 10 years earlier and that parental reports can be inaccurate. The authors were not able to match parent reports with administrative records but did compare the NELS data with other national data sources and concluded they were consistent.

Generalizability. The study reports patterns exhibited in the 228 poorest and 349 next poorest counties in the country through the period of 1960 through 2000. The measured impacts of the discontinuity in Head Start funding were through the mid 1960s to late 1970s, making the findings hard to generalize to today's Head Start population. Ludwig and Miller acknowledge this in their discussion, "Our study in this sense suffers from the generic problem of trying to generalize estimates of long-term effects to current policies -- long-term impacts can only be estimated for cohorts treated a long time ago."[30] Additionally, the counties studied suffered from very high poverty rates, had a disproportionate sample of blacks and were mostly located in the South. This is paired with the programmatic changes that have happened within Head Start since its inception in 1965 and with overall increases in rates of immunization and decrease in deaths related to causes addressed by the program (such as meningitis, measles, anemia).

As Greg J. Duncan points out, "Ludwig and Miller may have produced an elegant estimate of Head Start impacts, but it is hardly a general one: the exogenous variation in Head Start spending they exploited applies to Head Start centers operating more than 30 years ago, in counties with poverty rates just above or below 59%. Whether their estimated impacts hold true in more affluent areas, or in more recent years, is a matter of considerable speculation."[31]

Replication. There have not been any other studies that have replicated these findings.

Evaluator's description of findings. In their 2005 paper, the authors conclude, "If our study population of poor children in the South in the 1960s and 1970s is more disadvantaged than the average Head Start child today, then our estimates may provide an upper bound for the effects of expanding HS funding for more recent national cohorts."[32] The estimates reported in this version of the study were a bit larger than those reported in their 2006 and 2007 papers. Specifically, they report a 2-5 percentage point difference in high school completion, a 3-6

[30]Ludwig and Miller, 2007, 199.

[31]Greg J. Duncan, "When to Promote and When to Avoid a Population Perspective," *Demography* 45, 4 (2008): 778.

[32]Ludwig and Miller, 2005, 35.

percentage point difference for college attendance, and from the NELS data a discontinuity of .5 and 1.0 years of schooling.[33]

In his 2006 David N. Kershaw Award lecture, Dr. Ludwig suggested that these findings went against the notion "… that only very intensive, expensive interventions can achieve long-term impacts."[34] This is a very strong statement given that their findings showed a decrease in short-term health outcomes and were only able to provide suggestive evidence for educational attainment.

The conclusions made by Ludwig and Miller in later versions of the paper are more tempered. The authors conclude that their study contributes a new source of data that can be used to assess long-term impacts of Head Start and also shows how differential funding of Head Start programs had a significant impact on child mortality rates for causes attributable to Head Start participation. They frame their findings on educational attainment as "suggestive" due largely in part to the limitations in the Census and NELS data. Additionally, they conclude that their findings are "suggestive" for positive impacts on blacks and whites, but this seems a bit of a stretch given that the standard errors around their estimates for blacks were too large for any clear conclusions to be drawn.

Moreover, the authors conclude that their findings "seem to argue against the claim that Head Start has been a failure from the beginning"[35] and lend support to the notion that the program's benefits exceeded its costs, at least for the population and time period that they studied. They also posit that test-scores and similar measures may not be reliable predictors of the long-term impacts of Head Start as they only found suggestive evidence for educational attainment using measures of funding discontinuity.

The impact of their conclusions regarding the costs and benefits of the Head Start program were expanded upon by Ludwig and his former colleague Deborah Phillips: "the available evidence suggests to us that the Head Start program as it currently operates probably passes a benefit-cost test. Changing the program in various ways that have figured prominently in recent policy discussions may not make the program any better, and could make things worse."[36] This statement is based, in part, on the findings of Ludwig and Miller and of Garces, Thomas and

---

[33]Ludwig and Miller, 2005, 24, 26.

[34]Jens Ludwig, "Does Head Start Improve the Life Chances of Poor Children? New Evidence from Natural and Randomized Experiments," David N. Kershaw Award Lecture, Association for Public Policy Analysis and Management Fall Research Conference (November 3, 2006).

[35]Ludwig and Miller, 2007, 199.

[36]Jens Ludwig and Deborah Phillips, "Long-term effects of Head Start on Low Income Children," *Annals of the New York Academy of Sciences* 1136 (2008): 267.

Currie (see Chapter 5 of this volume). These studies examined Head Start as it was during the 1960s through the 1980s and seemingly demonstrated the program's long-term impacts. The assumption here is that the program of the past is comparable to the program of today and therefore, one would expect the same type of cost-benefit effects. This is a very strong assertion and considering the findings of recent studies of Head Start (see Chapter 13 of this volume) seems to be misplaced.

Evaluator's independence. The evaluator's are independent researchers from the University of Chicago and the University of California (Davis).

Statistical significance/confidence intervals. Statistical significance is measured and reported at the 1 percent and 5 percent level

Effect sizes. Effect sizes were not reported.

Sustained effects. This study examined the sustained effects of Head Start participation on education achievement through college. They were not able to measure college graduation rates and did not investigate Head Starts effects on related variable such as employment and crime. Additionally, they measured mortality rates a few years after participation in Head Start but did not look at any long-term health impacts.

Benefit-cost analysis. Apparently not performed.

Cost-effectiveness analysis. Apparently not performed.

Note: This report is open to public comments, subject to review by the forum moderator. To leave a comment, please send an email to welfareacademy@umd.edu or fill out the comment form at http://www.welfareacademy.org/pubs/early_education/chapter19.html.