Early life determinants of long-run outcomes

Heidi L. Williams
MIT 14.662
Spring 2015

Outline
(1) Preliminaries
(2) Prenatal environments
(3) Early childhood environments
(4) Policy responses

1 Preliminaries

Almond and Currie's 2011 *Handbook of Labor Economics* chapter (Almond and Currie, 2011) gives an excellent synthesis of the literature on early childhood influences on later life outcomes, and of the evidence on various public policies that aim to ameliorate the effects of negative influences. As they discuss, although early childhood conditions have been the focus of study in a wide variety of disciplines, in recent years there has been a flurry of economics research in this area – largely stemming from a growing recognition that early life conditions can have persistent and economically meaningful impacts on later life outcomes. Their take-away in the abstract is: *"Child and family characteristics measured at school entry do as much to explain future outcomes as factors that labor economists have more traditionally focused on, such as years of education. Yet while children can be permanently damaged at this age, an important message is that the damage can often be remedied."*

Almond and Currie begin with a brief overview of the theory that illustrates that evidence of a causal relationship between a shock in early childhood and a future outcome says little about whether the relationship in question is biological or immutable, because parental and social responses are likely to be extremely important in either magnifying or mitigating the effects of a shock. Traditional models of "health capital" - such as the classic model by Michael Grossman (Grossman, 1972) - model health as a stock variable that depreciates over time, and which can increase due to health investments. The structure of depreciation in the model implies that as individuals age, the effects of early childhood health stock and health investments become progressively less important over time. In contrast, the literature on early childhood influences asks whether health and investments in early childhood can have sustained effects on adult outcomes.

Almond and Currie define h as health or human capital at the completion of childhood, but unlike Grossman (1972) leave open the question of whether there is depreciation. They consider a two-period model of childhood production of h:

$$h = A[\gamma I_1 + (1-\gamma)I_2] \tag{1}$$

where

$$I_1 \simeq \text{investments during childhood through age 5} \tag{2}$$
$$I_2 \simeq \text{investments during childhood after age 5} \tag{3}$$

In this set-up, for a given level of investment $(I_1 + I_2)$ the allocation of investment across periods will matter if $\gamma \neq 0.5$. If $\gamma A > 1$, h may respond more than one-for-one with I_1, admitting the possibility that certain childhood periods may exert a disproportionate effect on later life outcomes that does not necessarily decline monotonically with age.

This is a somewhat extreme functional form which gives very sharp predictions on the optimal timing of investments: because investments in the first and second period are perfect substitutes, all investment should be concentrated in one period (up to a discount factor) and no investments should be made during the low-return period. Almond and Currie review a more flexible constant elasticity of substitution (CES) functional form suggested by Heckman (2007):

$$h = A[\gamma I_1^{\phi} + (1-\gamma)I_2^{\phi}]^{\frac{1}{\phi}} \tag{4}$$

In this set-up, for a given level of investment $(I_1 + I_2)$, how the allocation between period 1 and period 2 will affect h depends on the elasticity of substitution $(\frac{1}{1-\phi})$ and the share parameter (γ). This equation simplifies to the more restrictive functional form in the case where $\phi = 1$ (that is, where investments are perfectly substitutable).

Almond and Currie use this framework to consider the effect of exogenous shocks μ_g to health investments that occur during the first childhood period. If investments $\bar{I}_1$ and $\bar{I}_2$ do not respond to this shock, then net investments in the first period are $\bar{I}_1 + \mu_g$. Assume that μ_g is independent of $\bar{I}_1$ and that $\bar{I}_1 + \mu_g > 0$. This thought experiment of holding behavior constant can be thought of as attempting to shed light on a "biological" relationship (*e.g.* what is the effect of air pollution early in life on later life outcomes, holding behavior fixed). In the Heckman framework, the impact of this shock on h is:

$$\frac{\partial h}{\partial \mu_g} = A[\gamma(\bar{I}_1 + \mu_g)^{\phi} + (1-\gamma)\bar{I}_2{}^{\phi}]^{\frac{1}{\phi}-1}\gamma(\bar{I}_1 + \mu_g)^{\phi-1} \tag{5}$$

$$= \gamma A[\gamma(\bar{I}_1 + \mu_g)^{\phi} + (1-\gamma)\bar{I}_2{}^{\phi}]^{\frac{1-\phi}{\phi}}(\bar{I}_1 + \mu_g)^{\phi-1} \tag{6}$$

In the perfect substitutes case ($\phi = 1$), this simplifies to γA - implying that damage to adult human capital h from μ_g is proportional to the share parameter γ on period 1 investments, and is unrelated to $\bar{I}_1$. In the imperfect substitutes case, there is diminishing marginal productivity of investment inputs ($\frac{\partial^2 h}{\partial \mu_g \partial I_1} < 0$) - implying that shocks experienced at different baseline investment levels have heterogeneous effects on h. That is, individuals with higher baseline levels of investment will experience more muted effects in h than will individuals with lower baseline levels of investment. All else equal, this would accentuate the effect of an equal-sized μ_g shock on h among poor families.

Importantly, large damage to h from μ_g says little about the potential effectiveness of remediation in the second period. Almond and Currie model remediation as a second period shock $\mu_g' > 0$, and derive a condition that describes when remediation will be more effective.

A reduced form estimate of $\frac{\partial h}{\partial \mu_g}$ includes any biological effect as well as the effect of any responsive investments. Unless investment responses are costless, damage estimates of $\frac{\partial h}{\partial \mu_g}$ will tend to understate total costs. In the extreme, investment responses could fully offset the effect of early-life shocks on h, but that would not mean that such shocks were costless. More generally, the damage from early-life shocks will be understated if we focus on long-term effects and there are compensatory investments. Investment responses can be either reinforcing or compensatory: Almond and Currie work through special cases of their model that illustrate both. This model is helpful in highlighting the potential role of responsive investments, and as we will discuss a few recent papers use data with measures of parental investments to try to investigate this issue. For example, Royer (2009) investigates whether parental investments soon after birth differ within twin pairs as a function of birth weight differences, and does not find evidence of differences. Almond and Currie review several other papers that have investigated this question, and conclude that as of now there is little evidence that parents in developed countries systematically reinforce or compensate for early childhood events.

2 Prenatal environments

David Barker, a British physician and epidemiologist, popularized the idea that disruptions to the prenatal environment presage chronic health conditions in adulthood, including heart disease and diabetes. Almond and Currie summarize the so-called "Barker hypothesis" as the idea that growth is most rapid prenatally, and that when growth is rapid, disruptions to development can exert long-term effects. They note that this view contrasts with the idea that pregnant mothers serve as an effective buffer for the fetus against environmental insults. See Almond and Currie (forthcoming) for a recent overview of the Barker hypothesis.

Almond and Currie review three sets of prenatal factors: factors affecting maternal (and in turn, fetal) health, economic shocks, and pollution. I focus here on the literature investigating how birth weight affects long-run outcomes, with a particular focus on the recent paper by Black, Devereux and Salvanes (2007).

2.1 Birth weight and long-run outcomes

The first paper of which I am aware that looked at how birth weight was associated with long-run outcomes is Currie and Hyson (1999). Conditional on a rich set of controls available in the British National Child Development Survey, they documented that low birth weight status is associated with long-term disadvantages in self-reported health status, educational attainment, and labor market outcomes. Although useful, this type of regression seems likely to be problematic given that birth weight is routinely found to be strongly associated with socio-economic background variables, some of which are likely unobserved in any given data set. Thus, it is difficult to ascertain from this type of data whether birth weight has a causal link with long-term outcomes.

One empirical strategy that has been used to address this limitation is to examine twin comparisons, relating within-twin-pair differences in birth weight to differences in twins' long-run outcomes. To the best of my knowledge, Behrman and Rosenweig (2004) was the first paper to pursue this twins approach, finding that the schooling of identical female twins was approximately $\frac{1}{3}$ of a year longer for each pound increase in birth weight (454 grams). While an important advance, the sample size in the Behrman-Rosenweig paper was quite small (402 twin pairs). Royer (2009) took advantage of a much larger sample of twins (constructed from the universe of 1960-1983 California birth records). Royer estimates statistically significant but small associations between birth weight and long-run outcomes: heavier twins obtain more education, give birth to heavier children, and have fewer pregnancy complications, but the estimates tend to be quite small. Royer also tests for evidence of differential investment on the part of health care providers or parents, and does not uncover any evidence for differential investment. Oreopoulos, Stabile, Walld and Roos (2008) is another recent paper with a similar research design.

Almond, Chay and Lee (2005) is an important related paper which used twin estimation and estimated smaller effects of low birth weight on health care costs than had been estimated in prior cross-sectional analyses. The authors emphasize that differences between cross-section and twin fixed effect estimates can support two different interpretations. First, the fixed effects could "solve" an omitted variables bias problem inherent in cross-sectional regressions. Second, different sources of variation in birth weight could have different effects on child outcomes. That is, birth weight itself is not a policy variable, and different policies that affect birth weight may have different effects on other outcomes.[1]

2.2 Black, Devereux and Salvanes (2007)

Black, Devereux and Salvanes (2007) examine short- and long-run effects of birth weight using birth records for the census of Norwegian births from 1967-97 linked to administrative data on infant mortality, APGAR score, height, BMI, IQ, education, labor market outcomes, and birth weight of first child (to test for intergenerational transmission of birth weight).

[1] Almond, Chay and Lee (2005)'s discussion focuses in part on medical evidence that low birth weight is governed by two factors: (1) short gestation (prematurity), and (2) reduced fetal growth at a given gestation length (intrauterine growth retardation, or IUGR). By construction, given that twin births are usually very close together in time, the twin-based variation focuses on the second source of variation.

Table 3 presents estimates from pooled OLS and twins fixed effect estimators. Each coefficient represents the estimate from a separate regression. For mortality, the pooled OLS coefficient of 280 implies that a 10 percent increase in birth weight would reduce 1-year mortality by approximately 28 deaths per 1,000 births. The twin fixed effects coefficient of 41 is statistically significant but only $\frac{1}{6}$ the size of the OLS coefficient. For APGAR scores, the authors similarly find that the OLS estimates overstate the size of the coefficient. However, for later-life outcomes - height, IQ at age 18, earnings, and education - the fixed effects estimates are similar in size to cross-sectional ones. In the conclusion they offer some conjectures about why the OLS and fixed effect estimates might differ in the short-run but converge in the long-term.

TABLE III

Regression Results: Twins Sample Coefficient on LN (Birth Weight)

Dependent variable	Singleton sample: OLS	Singleton sample: Family fixed effects	Twins sample: OLS	Twins sample: Twin fixed effects
One-year mortality	−123.46** (1.71)	−186.71** (.69)	−279.64** (9.12)	−41.10** (7.64)
N	1,253,546		33,366	
Five minute APGAR score	.73** (.01)	1.08** (.01)	1.46** (.06)	.35** (.07)
N	674,577		21,580	
Height (males only)	11.03** (.11)	7.33** (.12)	7.48** (.55)	5.68** (.56)
N	203,741		5,382	
BMI (males only)	−6.19 (7.67)	−22.22 (15.23)	.56** (.23)	1.12** (.30)
N	203,378		5,372	
Underweight	−.09** (.004)	−.07** (.01)	−.07** (.02)	−.11** (.04)
N	203,378		5,372	
Overweight	.08** (.01)	.08** (.01)	.03 (.02)	.09** (.04)
N	203,378		5,372	
IQ (males only)	.91** (.03)	.58** (.04)	.48** (.14)	.62** (.18)
N	184,045		4,920	
High school completion	.16** (.01)	.04** (.01)	.07** (.02)	.09** (.04)
N	536,020		13,106	
Full-time work	.17** (.004)	.21** (.01)	.29** (.02)	.03 (.05)
N	368,582		10,388	
ln(earnings) FT	.09** (.01)	.08** (.01)	.09** (.03)	.12** (.06)
N	239,906		5,952	
ln(birth weight of first child)	.25** (.01)	.13** (.01)	.18** (.04)	.15** (.06)
N	63,842		1,862	

Standard errors are in parentheses. The control variables we use in the OLS estimation are year- and month-of-birth dummies, indicators for mother's education (one for each year), indicators for birth order, indicators for mother's year of birth, and an indicator for the sex of the child. Family fixed effects regressions include all of the above minus mother's education and mother's year of birth. Twin fixed effects regressions include indicators for sex and birth order of the twin (either first born or second born twin). Both cross-sectional and fixed effects regressions for height, BMI, and IQ also include indicator variables for the year the boy was tested by the military. High school completion indicates whether or not the individual has completed at least twelve years of schooling and is restricted to those twenty-one and older. The IQ measure is generated from a composite score from three speeded IQ tests—arithmetic, word similarities, and figures—and is reported in stanine (Standard Nine) units. Earnings are measured as total pension-qualifying earnings reported in the tax registry. These are not topcoded and include labor earnings, taxable sick benefits, unemployment benefits, parental leave payments, and pensions. We restrict attention to individuals aged at least twenty-five. Working full-time indicates whether individuals are full-time, full-year workers. To identify this group, we use the fact that our dataset identifies individuals who are employed and working full time (30+ hours per week) at one particular point in the year (in the second quarter in the years 1986-1995 and in the fourth quarter thereafter). We label these individuals as full-time workers. For ln(birth weight) of child, the sample consists of women born between 1967 and 1988 whose first births occurred by 2004. If the first birth is a twin birth, the woman is dropped from the sample.

** Denotes statistically significant at the 5 percent level.

* Denotes statistically significant at the 10 percent level.

Given that the authors find an impact of birth weight on mortality, when studying long-run outcomes selection becomes important: twin pairs that experience mortality are dropped from the analysis. Because they observe birth characteristics (including birth weight) of twin pairs who subsequently experience mortality events, they can examine characteristics associated with selection into the sample. They make several points:

- Table VI shows that the twin fixed effects estimates of the impact of birth weight on later outcomes have tended to increase over time; over the same time period, infant mortality amongst twins has declined. While not definitive, these patterns are consistent with the idea that later life effects are larger because the sample includes more twins who were on the margin of survival in infancy.
- If there are heterogeneous effects of birth weight across twin pairs and birth weight is more important for twin pairs who subsequently experience mortality, they may be underestimating the average effect of birth weight on later outcomes over all twin pairs. The authors test this theory by estimating the relationship between birth weight and APGAR score for the full sample and separately for the sample of twin pairs where both twins live. When they do this using twin fixed effects, they find that log birth weight has a significantly larger positive effect on the APGAR score for the full sample of twin births. If this relationship is also true for other, later outcomes (not testable), this would imply that the true average effect of birth weight on later outcomes would be understated.
- Taken together, the authors conclude that survival-induced selection bias most likely leads to an understatement of the effects of birth weight on adult outcomes.

3 Early childhood environments

Almond and Currie also review the literature on how "early childhood environment" (birth to age five) affects long-term outcomes. They discuss papers estimating the impact of infectious diseases, health status, home environment (including maternal employment), and pollution/toxins. I here focus on a recent paper by Bharadwaj, Løken and Nielson (2011) that examines how early life health interventions affect academic achievement.

3.1 Bharadwaj, Løken and Nielson (2011)

A significant amount is spent on medical care for at-risk newborns (preterm and low-birth weight infants), and yet the benefits of this spending is not well understood – in part because very few randomized trials are conducted on infant populations. In addition, medical studies typically track the impact of health spending on 1-year mortality, but not on other outcomes; if early life health interventions have benefits on the "intensive margin" (that is, generating gains for infants beyond those on the margin of survival), valuing these benefits is important from a health and social policy perspective.

In order to investigate the effects of neonatal care on long-term outcomes, Bharadwaj, Løken and Nielson (2011) take advantage of a set of rules and recommendations that generate a discontinuity in the receipt of medical care around the "very low birth weight" threshold at 1500 grams. This idea follows up an earlier paper by Almond, Doyle, Kowalski and Williams (2010), who identified a discontinuity in medical expenditures around 1500 grams for US infants and used this variation to estimate the returns to medical spending in terms of reduced mortality. The key idea is that infants born at 1490 grams should be similar to infants born at 1510 grams, but that by falling below the 1500 gram threshold the former group of infants receive additional medical inputs - inputs which appear to translate into a reduced probability of mortality.

Bharadwaj, Løken and Nielson (2011) apply this research design to data from Chile and Norway. Both Chile and Norway publish official medical recommendations that appear to be binding around these thresholds. The institutional structure of these recommendations also offer a nice 'placebo check': because the medical recommendations apply to all infants that are either less than 1500 grams *or* less than 31 weeks of gestation, the authors can check that there is no discontinuity across 1500 grams in the sample of newborns less than 31 weeks of gestation (since all infants in that sample should be eligible for 'intensive' care independent of their birth weight). Their results are easily summarized in a series of graphs: they find evidence that additional medical care (Figure 1) translates into lower mortality (Figure 2) as well as improved school performance (Figure 3).

Figure 1: Treatments around 1500 grams

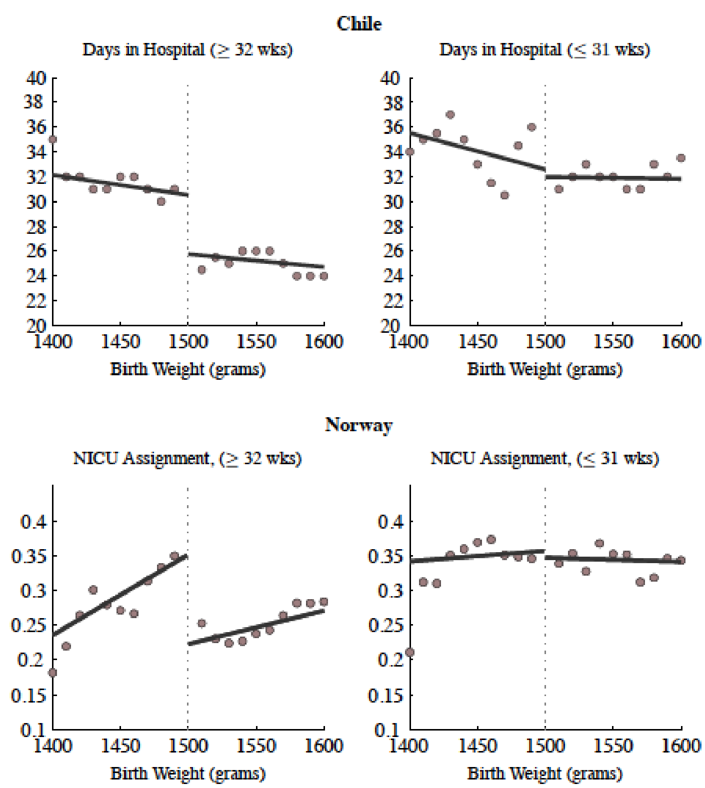


Figure 2: Infant Mortality

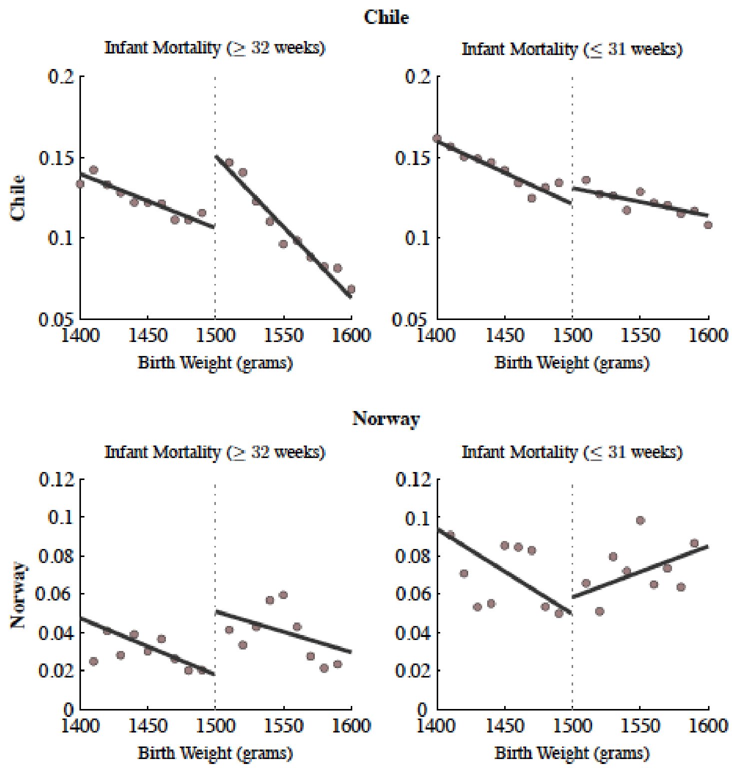
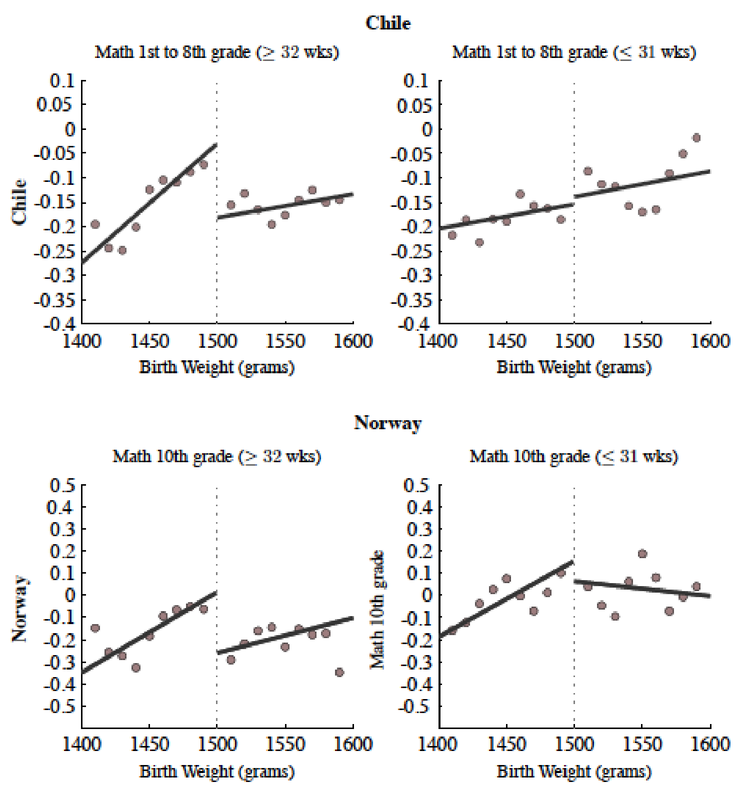
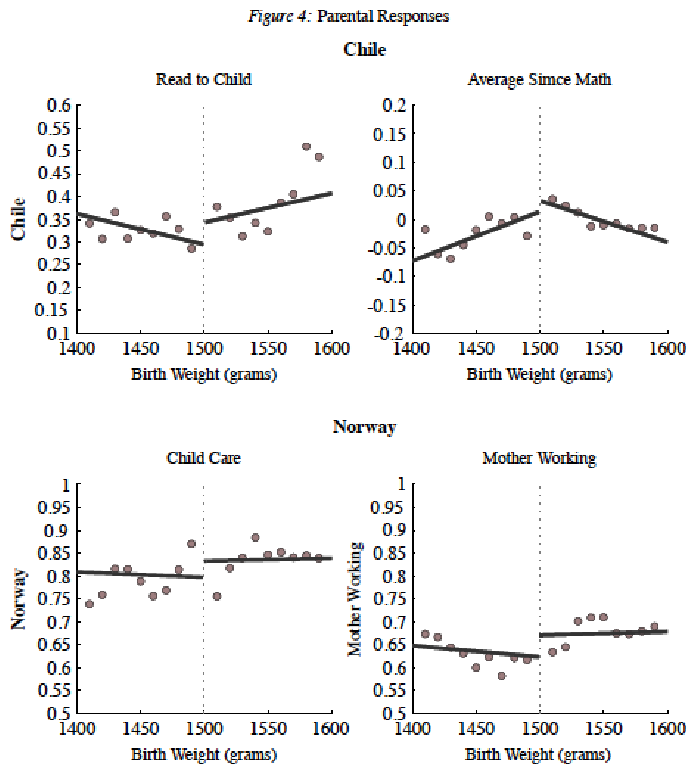


Figure 3: School Performance



Using data from time use surveys and other sources, they find no evidence that these results are driven by differential parental investments around the cutoff (Figure 4).



Courtesy of Prashant Bharadwaj, Katerine V. Løken, Christopher Neilson, and the American Economic Association. Used with permission.

4 Policy responses

The evidence above - reviewed more thoroughly in Almond and Currie (2011) - suggests that prenatal and early childhood factors can have important influences on later life outcomes. However, on its own this evidence has little to say about the effectiveness of remediation. Almond and Currie review the evidence on income transfer programs (such as payments from the Earned Income Tax Credit), "near cash" programs (such as food stamps), early intervention programs (such as home nurse visiting programs), and health insurance. I here focus on briefly reviewing empirical evidence from some recent analyses of the Head Start program (Currie and Thomas, 1995; Deming, 2009; Ludwig and Miller, 2007), and then discuss in more detail Doyle (2007)'s investigation of the effects of foster care.

4.1 Head Start

In an influential paper, Currie and Thomas (1995) used sibling fixed effect models to estimate how Head Start attendance affected children's outcomes. The data reveal some differences between participant and non-participant children within families which suggest that the Head Start sibling typically attended when the family was relatively disadvantaged, but there were no within-family differences in individual characteristics such as birth weight. The data suggested significant positive effects of Head Start on subsequent educational attainment. Deming (2009) uses the same research design and data to examine child outcomes at older ages, and argues that the projected gain in earnings are enough to offset the cost of the program.

4.1.1 Ludwig and Miller (2007)

Ludwig and Miller (2007) evaluate Head Start via a different research design, looking at the roll-out that occurred at the program's inception. When initially established in 1965, Head Start provided assistance to the 300 poorest counties in the US to develop Head Start proposals. As with the Bharadwaj, Løken and Nielson (2011) paper, their results are easily summarized in a series of graphs. Using new data on historical Head Start program expenditures dug out of the National Archives, they document in Figures 1 and 2 that this led to a substantial and persistent discontinuity in Head Start funding rates (and participation rates, in Figure 1 - not shown).

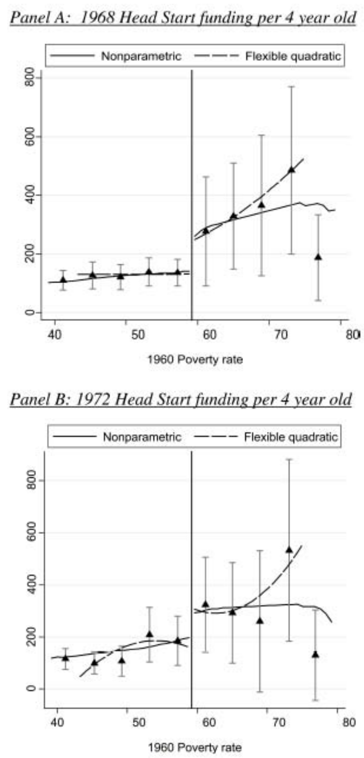


FIGURE II

Estimated Discontinuity in Head Start Funding per Four-Year-Old, National Archives. (A) 1968 Head Start funding per four-year-old and (B) 1972 Head Start funding per four-year-old

Notes: Each panel shows the nonparametric estimate (solid line) for the function relating 1960 county poverty rate to the dependent variable [$m(P_c)$ from (3) in the text] as well as the implied discontinuity (α) using a bandwidth of 18, a parametric estimate (dashed line) that uses a quadratic to model $m(P_c)$, and raw cell means (triangles) and their 95 percent confidence intervals (bars) from grouping the data into five categories on each side of the cutoff for counties with 1960 poverty rates from 40 to 80 percent. Panel A Estimated nonparametric discontinuity = 114.71 T-stat = 1.19, bandwidth = 18. Panel B Estimated nonparametric discontinuity = 89.96 T-stat = 0.83, bandwidth = 18.

Comfortingly, Figure 3 (not shown) suggests there are no differences in other federal per-capita social spending across this threshold. Categorizing causes of death to those that are more versus less likely to be affected by Head Start, Figure 4 documents a large drop in mortality rates from "Head Start causes" at the cutoff.

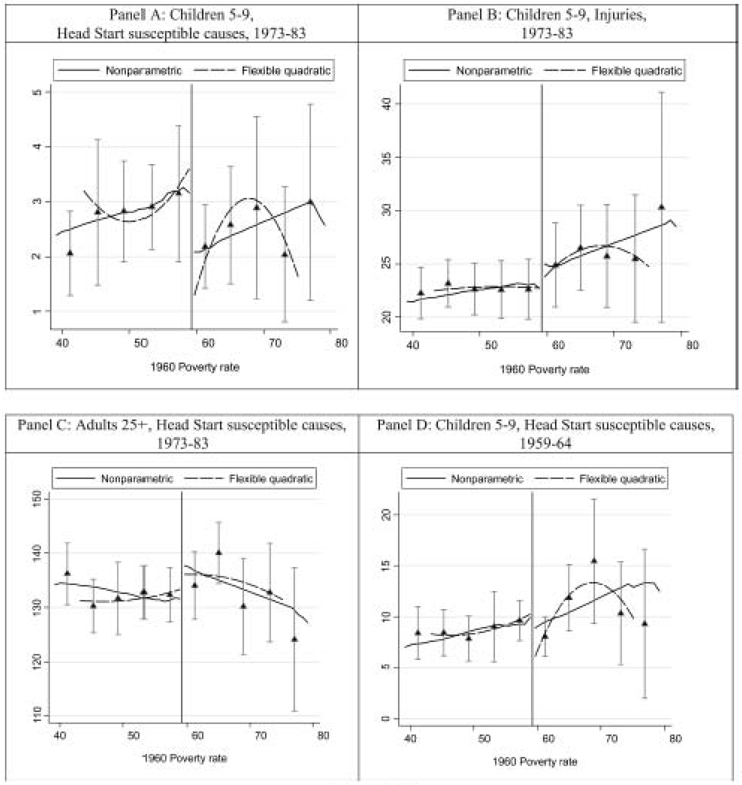


FIGURE IV

Estimated Discontinuities at OEO Cutoff in Mortality Rates per 100,000 for Children and Adults, from Causes Affected by Head Start and from Injuries

Note: Each panel shows the nonparametric estimate (solid line) for the function relating 1960 county poverty rate to the dependent variable [$m(P_c)$ from (3) in the text] as well as the implied discontinuity (α) using a bandwidth of 18, a parametric estimate (dashed line) that uses a quadratic to model $m(P_c)$, and raw cell means (triangles) and their 95 percent confidence intervals (bars) from grouping the data into five categories on each side of the cutoff for counties with 1960 poverty rates from 40 to 80 percent. Panel A, Estimated nonparametric discontinuity = −1.198 T-stat = 1.42, bandwidth = 18. Panel B, Estimated nonparametric discontinuity = 2.246 T-stat = 0.86, bandwidth = 18. Panel C, Estimated nonparametric discontinuity = 6.016 T-stat = 1.31, bandwidth = 18. Panel D, Estimated nonparametric discontinuity = −1.076 T-stat = 0.52, bandwidth = 18.

Figures 5 and 6 (not shown) provide more suggestive evidence for a positive effect of Head Start on educational attainment.

4.2 Foster care: Doyle (2007)

Joe Doyle has an important and thoughtful series of papers investigating child welfare service decisions over whether to leave abused or neglected children in their home, or to place them in foster care. As discussed in his paper, although an abusive home environment undoubtedly harms child development, removing children from home and placing them in a potentially unstable foster family relationship may be harmful as well.

The key empirical idea in Doyle's paper is to generate a measure of the removal tendency of child protection investigators in order to develop an instrumental variables strategy to identify the causal effects of foster care placement on child outcomes. The basis of the identification is that cases are distributed to investigators on a rotational basis within geographic field teams that effectively randomizes families to investigators. He imbeds this identification in a marginal treatment effects framework, which is a useful empirical strategy for you to have in your toolkit (one of the reasons I wanted to spend the time needed to go over this particular paper in detail). Doyle combines this strategy with a new, unique dataset that links children in the Illinois child welfare system with a variety of administrative datasets in order to track child outcomes such as juvenile delinquency, teen motherhood, employment, and earnings.

4.2.1 Empirical framework

The set-up of Doyle's empirical framework considers how the benefit or harm of the decision to remove a child from home can vary across children. Consider a random coefficient model for an outcome Y (earnings), observable case characteristics X, and an indicator R for removal from home for child i:

$$Y_i = X_i\beta + \alpha_i R_i + \varepsilon_i \tag{7}$$

Here, α_i will be positive for children where the placement is associated with higher earnings, but may be negative for children where the disruption of placement is associated with lower earnings. Rewriting this question to reflect the standard single coefficient model reveals two error terms:

$$Y_i = X_i\beta + \bar{\alpha} R_i + R_i(\alpha_i - \bar{\alpha}) + \varepsilon_i \tag{8}$$

Doyle discusses two econometric issues that can arise when estimating this equation. First, R may be correlated with ε if an omitted variable (for example, a family characteristic) leads to both an increased likelihood of removal and a decreased earnings capacity. Second, R will be correlated with α_i if agents select treatment based on gains (in a Roy model sense). Doyle notes that for foster care placement the treatment is not chosen by the child, but by the child protection system; although the placement decision is likely not explicitly based on the returns to earnings, if earnings were indicative of child well-being in general than we expect such a correlation to exist.

Doyle's estimation uses an instrument Z that can both overcome the endogeneity concern, and allow estimation of marginal treatment effects as Z varies. Consider two types of investigators, tough and lenient. The difference in outcomes across these investigators could be used to measure local average treatment effects (LATEs): the children induced into foster care on the basis of the investigator assignment. Letting $Z = 1$ if the family is assigned to a tough investigator and $Z = 0$ if assigned to a lenient investigator, the LATE estimand is:

$$\alpha^{\text{LATE}} = \frac{E(Y|Z=1) - E(Y|Z=0)}{P(R=1|Z=1) - P(R=1|Z=0)} \tag{9}$$

The usual conditions are required in order to interpret this estimate as a LATE:

1. First stage: the instrument is associated with foster care placement
2. Exclusion restriction: Z is not in the outcome equation
3. Monotonicity: any child removed by a lenient investigator would also be removed by a strict one, and a child not removed by a strict case manager would not be removed by a lenient one

Doyle conceptualizes his instrument within the context of a placement decision model where investigators observe cases along a distribution of abuse levels θ, as in Figure 1:

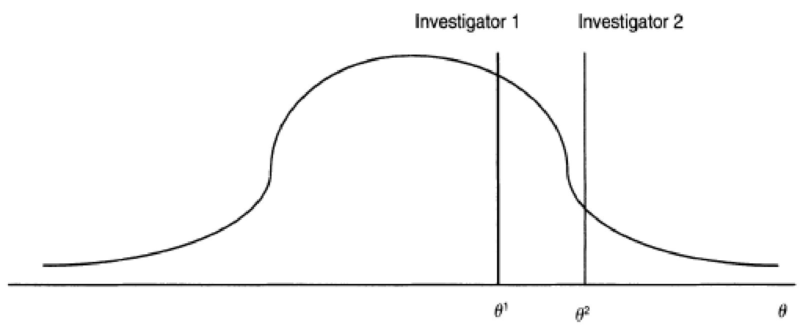


FIGURE 1. ABUSE THRESHOLDS FOR REMOVAL

The two types of investigators are defined by the threshold of abuse required to recommend placement. Each type observes the same abuse levels, so investigator types can be described by the fraction of children they recommend for placement (Z). The comparison of outcomes across investigator types focuses on variation in placement among marginal cases: for very high levels of abuse both types would recommend removal, for very low levels of abuse both types would recommend keeping the child at home, and the only empirical identification comes from children on the margin of placement. While not measuring the effects for all children, this is a very policy relevant group.

Doyle summarizes this by a latent index model for child i:

$$R_i^* = -Z_i\gamma + \theta_i \tag{10}$$

$$R_i = 1 \text{ if } R_i^* > 0 \tag{11}$$

Here, Z_i can be thought of as characterizing the threshold the investigator assigned to child i must observe before she decides to recommend foster care placement, and γ represents the influence that such a recommendation will actually result in a placement. A child with abuse level θ will be placed in care if that level is greater than the investigator's threshold for removal multiplied by the effectiveness of that recommendation.

The conditions for identification here are that there is a first stage ($\gamma \neq 0$), and three conditions that are implied by quasi-random assignment:

1. $E(Z\theta) = 0$

2. $E(Z\varepsilon) = 0$

3. $E(Z(\alpha - \bar{\alpha})) = 0$

Here, the monotonicity assumption is imbedded in the common coefficient γ. An MTE is the limit of the LATE as the difference in the probability of treatment, given the instrument, goes to zero. In Figure 1, this means comparing outcomes for children across case managers whose thresholds are close together. Letting $P(Z)$ equal $P(R = 1|Z = z)$, the marginal treatment effect is the derivative: $\alpha^{\text{MTE}} = \frac{\partial E(Y)}{\partial P(z)}$. The MTE estimates are here of interest because they describe whether outcomes improve or become worse as different types of children are induced into foster care based on different values of the investigator propensity.

4.2.2 Results

Table 1. Table 1 reports summary statistics for the "delinquency sample" (individuals in Cook County, for whom delinquency data is available). Throughout the paper, Doyle shows results from three samples: this delinquency sample, the teen motherhood sample (females only), and the employment sample (all children linked to employment outcomes).

TABLE 1—SUMMARY STATISTICS: DELINQUENCY SAMPLE

Variable		*Mean*	*Standard deviation*	*Maximum*	*Minimum*
	Foster care placement	0.27	0.44	0	1
Initial reporter	Physician	0.12	0.33	0	1
	School	0.13	0.33	0	1
	Police	0.13	0.34	0	1
	Family	0.29	0.46	0	1
	Neighbor	0.06	0.23	0	1
	Other government	0.09	0.29	0	1
	Anonymous	0.15	0.35	0	1
	Other reporter	0.03	0.16	0	1
Age at report	Age	11.3	2.5	5	15
Sex	Boy	0.47	0.50	0	1
Race	White	0.11	0.31	0	1
	African American	0.76	0.43	0	1
	Hispanic	0.12	0.32	0	1
	Other race/ethnicity	0.01	0.10	0	1
Allegation	Physical abuse	0.17	0.38	0	1
	Substantial risk of harm	0.24	0.43	0	1
	Other abuse	0.02	0.15	0	1
	Lack of supervision	0.37	0.48	0	1
	Environmental neglect	0.15	0.36	0	1
	Other neglect	0.04	0.20	0	1
Location	Cook County	1.00	0.00	1	1
Outcome	Delinquency	0.17	0.38	0	1
	Observations	15,039			

Tables 2 and 3. If families are effectively randomized to investigators, then child characteristics should be similar across investigators and should not predict the case manager's removal differential. To test this, Table 2 regresses the instrument on child characteristics. The estimates suggest that baseline observable child characteristics are not related to the instrument. This conclusion is confirmed by an F-test for joint significance between child characteristics and the case manager removal rate (Table 3).

TABLE 2—CHILD CHARACTERISTICS AND CASE MANAGER ASSIGNMENT: DELINQUENCY SAMPLE

Dependent variable: Case manager removal differential Variable		Coefficient	t	p-value
Initial reporter (Other reporter excluded)	Physician	−0.006	−0.81	0.416
	School	−0.005	−0.74	0.457
	Police	−0.008	−1.11	0.269
	Family	−0.003	−0.52	0.605
	Neighbor	−0.005	−0.73	0.464
	Other government	−0.007	−0.96	0.339
	Anonymous	−0.007	−1.07	0.287
Age at report (Youngest age excluded)	Age 6	0.005	0.41	0.679
	Age 7	0.012	1.07	0.284
	Age 8	0.009	0.90	0.367
	Age 9	0.015	1.42	0.156
	Age 10	0.008	0.72	0.470
	Age 11	0.009	0.94	0.346
	Age 12	0.010	0.99	0.324
	Age 13	0.013	1.26	0.207
	Age 14	0.009	0.91	0.366
	Age 15	0.009	0.89	0.373
Sex	Boy	−0.002	−1.20	0.232
Race/ethnicity (Other race excluded)	White	−0.014	−1.32	0.186
	African American	−0.015	−1.22	0.224
	Hispanic	−0.012	−0.88	0.377
Allegation (Other neglect excluded)	Physical abuse	−0.002	−0.43	0.668
	Substantial risk of harm	−0.006	−0.94	0.348
	Other abuse	0.003	0.43	0.670
	Lack of supervision	−0.005	−0.98	0.325
	Environmental neglect	−0.007	−1.29	0.199
	Mean of dependent variable	0.0001		
	Standard deviation	0.0921		
	F-statistic of joint significance	0.84		
	p-value	0.75		
	Number of case managers	409		
	Observations	15,039		

Note: t-statistics and F-statistic are calculated using standard errors clustered by case manager.

TABLE 3—CHILD CHARACTERISTICS AND CASE MANAGER ASSIGNMENT

Dependent variable: Case manager removal differential

Sample:	Delinquency (1)	Teen motherhood (2)	Employment (3)
F-statistic of joint significance	0.84	1.07	0.96
p-value	0.75	0.34	0.54
Mean of dependent variable	0.0001	−0.0007	−0.0007
Standard deviation of dependent variable	0.0921	0.1035	0.0729
Number of case managers	409	705	815
Observations	15,039	20,091	30,415

Notes: All models include full controls (individual year, month, and age indicators). F-statistics are calculated using standard errors clustered by case manager.

Tables 4 and 5. Table 4 reports the first stage estimates for the juvenile delinquency sample: the case manager removal differential is positively associated with foster care placement. The estimated marginal effect is 0.3 implies that an increase in the removal differential from one standard deviation below the mean to one standard deviation above would be associated with a 6 percentage point increase in the likelihood of removal, 22 percent of the mean removal rate. The addition of controls does not change the estimates very much, as expected by the estimates in Tables 2 and 3. Table 5 reports the first stage estimates for other samples.

TABLE 4—CASE MANAGER ASSIGNMENT AND FOSTER CARE PLACEMENT: JUVENILE DELINQUENCY SAMPLE

Dependent variable: Case manager removal differential

	Model:	Probit			Probit		
		Coefficient	S.E.	*p*-value	Coefficient	S.E.	*p*-value
Key explanatory variables	Case manager removal differential	0.30	0.07	0.00	0.27	0.05	0.00
Initial reporter (Other reporter excluded)	Physician				0.10	0.03	0.00
	School				−0.02	0.03	0.43
	Police				0.14	0.03	0.00
	Family				0.05	0.03	0.06
	Neighbor				0.02	0.03	0.53
	Other government				0.07	0.03	0.03
	Anonymous				−0.06	0.03	0.02
Age at report (Youngest age excluded)	Age 6				0.05	0.05	0.21
	Age 7				0.05	0.04	0.18
	Age 8				0.02	0.04	0.66
	Age 9				0.03	0.04	0.44
	Age 10				0.03	0.04	0.42
	Age 11				0.02	0.04	0.55
	Age 12				0.00	0.04	0.97
	Age 13				−0.02	0.04	0.63
	Age 14				−0.04	0.04	0.32
	Age 15				−0.07	0.03	0.08
Sex	Boy				−0.01	0.01	0.14
Race/ethnicity (Other race excluded)	White				0.00	0.05	0.95
	African American				0.11	0.04	0.02
	Hispanic				−0.03	0.05	0.50
Allegation (Other neglect excluded)	Physical abuse				−0.07	0.02	0.00
	Substantial risk of harm				0.00	0.02	0.88
	Other abuse				−0.09	0.02	0.00
	Lack of supervision				0.00	0.02	0.89
	Environmental neglect				−0.08	0.02	0.00
	Chi-squared (1) stat.	17.9			27.8		
	Mean of dep. var.	0.27					
	Observations	15,039					

Note: Marginal effects and standard errors clustered at the case manager level are reported.

TABLE 5—CASE MANAGER ASSIGNMENT AS A PREDICTOR OF REMOVAL

Dependent variable: Foster care placement

	Delinquency sample		Teen motherhood sample		Employment sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Case manager removal differential	0.301	0.274	0.231	0.204	0.327	0.288
	(0.071)	(0.052)	(0.050)	(0.035)	(0.060)	(0.039)
Mean of dependent variable	0.27		0.21		0.23	
Chi-squared (1) statistic	17.9	27.8	21.5	34.2	29.3	55.0
Observations	15,039		20,091		30,415	
Full controls	No	Yes	No	Yes	No	Yes

Note: Results of probit models, with marginal effects and standard errors clustered at the case manager level, are reported.

Figure 2. Figure 2 presents local linear regressions of an indicator for foster care placement on the case manager removal differential for each of the three samples. This graph is helpful in graphically illustrating that there is a fairly monotonic increase in foster care placement with the case manager removal differential.

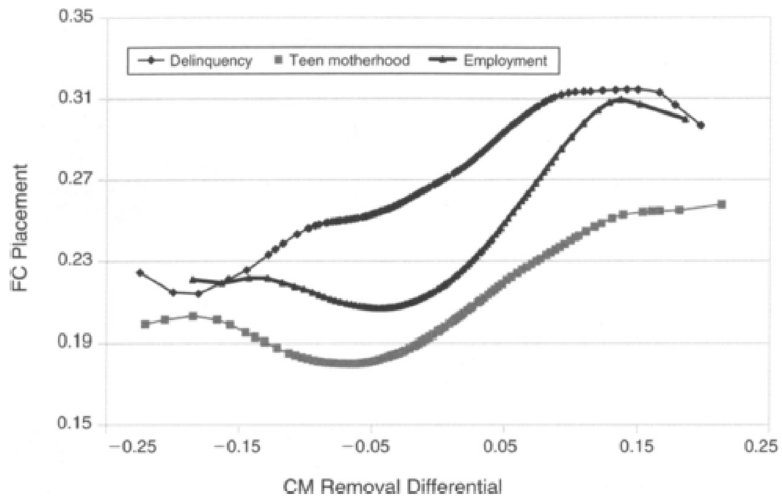


FIGURE 2. FC PLACEMENT VERSUS CASE MANAGER REMOVAL DIFFERENTIAL

Notes: Local linear regressions for the three samples. Bandwidth = 0.05.

Tables 6, 7, and 8. Tables 6, 7, and 8 report IV estimates for juvenile delinquency, teen motherhood, and employment outcomes (respectively). The estimated IV effects are quite large, although also quite imprecise.

TABLE 6—FOSTER CARE PLACEMENT AND JUVENILE DELINQUENCY

Dependent variable: Juvenile delinquency

	Model:	Probit				IV Probit			
		Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
	FC placement	0.01	0.01	0.00	0.01	0.26	0.14	0.35	0.14
Initial reporter	Physician			0.00	0.02			−0.02	0.02
(Other reporter excluded)	School			0.00	0.02			0.00	0.02
	Police			0.02	0.02			−0.01	0.03
	Family			0.00	0.02			−0.01	0.02
	Neighbor			0.01	0.03			0.00	0.03
	Other government			0.03	0.02			0.01	0.02
	Anonymous			0.01	0.02			0.03	0.02
Age at report	Age 5			—	—			—	—
(Youngest age excluded)	Age 6			0.06	0.05			0.04	0.05
	Age 7			0.10	0.05			0.08	0.05
	Age 8			0.13	0.05			0.12	0.05
	Age 9			0.13	0.05			0.12	0.05
	Age 10			0.17	0.06			0.15	0.05
	Age 11			0.19	0.06			0.18	0.05
	Age 12			0.22	0.06			0.21	0.05
	Age 13			0.23	0.06			0.23	0.06
	Age 14			0.23	0.06			0.23	0.06
	Age 15			0.12	0.05			0.14	0.05
Sex	Boy			0.19	0.01			0.19	0.01
Race/ethnicity	White			−0.07	0.03			−0.07	0.03
(Other race excluded)	African American			−0.02	0.04			−0.05	0.04
	Hispanic			−0.07	0.03			−0.07	0.03
Allegation	Physical abuse			−0.01	0.02			0.01	0.02
(Other neglect excluded)	Substantial risk of harm			−0.03	0.01			−0.03	0.02
	Other abuse			−0.02	0.02			0.01	0.03
	Lack of supervision			−0.02	0.02			−0.03	0.02
	Environmental neglect			−0.02	0.02			0.00	0.02
	Mean of dep. var.	0.17							
	Observations	15,039							

Note: Marginal effects and standard errors clustered at the case manager level are reported.

TABLE 7—FOSTER CARE PLACEMENT AND TEEN MOTHERHOOD

Dependent variable	Teen pregnancy			
Model	Probit (1)	Probit (2)	IV Probit (3)	IV Probit (4)
Foster care placement	0.106 (0.009)	0.090 (0.010)	0.171 (0.158)	0.291 (0.171)
Mean of dependent variable	0.35			
Full controls	No	Yes	No	Yes
Observations	20,091			

Note: Marginal effects and standard errors clustered at the case manager level are reported.

TABLE 8—FOSTER CARE PLACEMENT AND EMPLOYMENT & EARNINGS

Dependent variable	Fraction of quarters working in 2002				Average quarterly earnings in 2002			
Model	OLS (1)	OLS (2)	2SLS (3)	2SLS (4)	OLS (5)	OLS (6)	2SLS (7)	2SLS (7)
Foster care placement	−0.023 (0.006)	0.002 (0.006)	−0.110 (0.120)	−0.154 (0.113)	−82.8 (29.5)	−50.4 (30.6)	−851 (597)	−1,296 (626)
Mean of dep. var.	0.320				1,044			
Full controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	30,415							

Notes: Standard errors clustered at the case manager level are reported. Average quarterly earnings include those with zero earnings.

Figure 3. As the propensity to be placed in foster care increases with the case manager placement differential, the probability of adverse outcomes also rises and the employment measures tend to fall. The first derivative of each of these relationships represents the marginal treatment effect function. To calculate the MTE function, Doyle estimates the predicted probability of placement using a probit model in which the only explanatory variable was the case manager removal differential. The relationship between each outcome variable and the predicted probability of placement was then estimated using a local quadratic estimator and evaluated at each percentile of the predicted probability of placement. These estimates are reported in Figure 3.

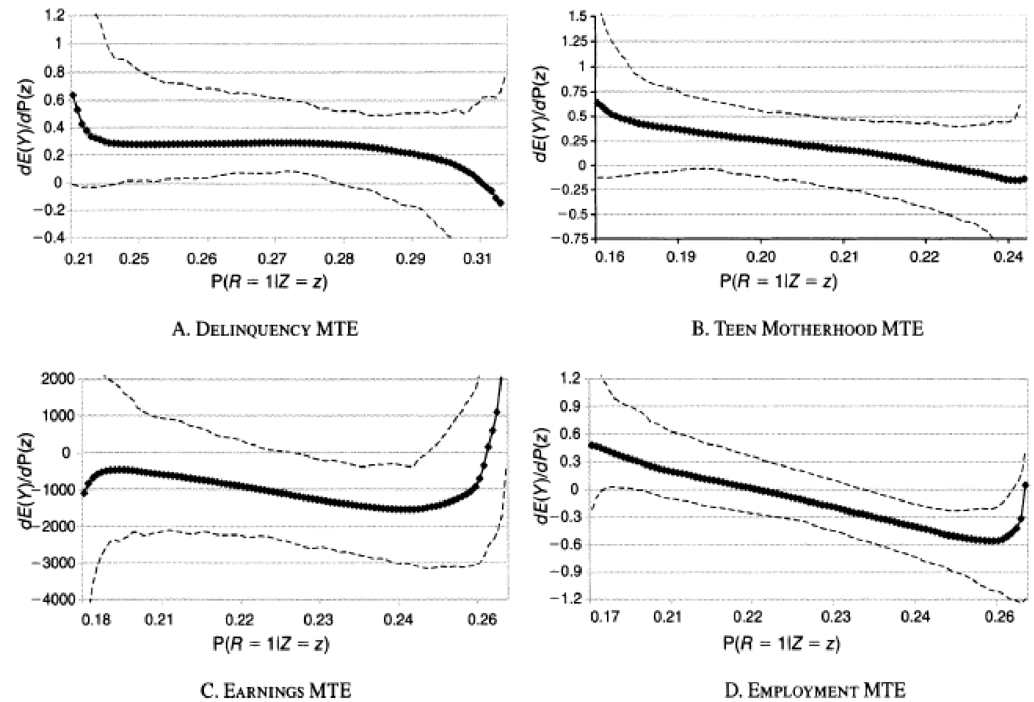


Figure 3

Notes: Figures report the results of a local quadratic estimator evaluated at each percentile of P(z). Confidence intervals of 5 to 95 percent reported, calculated using a bootstrap with 250 replications, clustered at the case manager level. Bandwidth = 0.037.

Table 9. Table 9 (not shown) reports heterogeneous treatment effects across different types of children, with a goal if clarifying which types of children are driving the results.

Table 10. Table 10 (not shown) reports some specification checks.

4.2.3 Take-aways

I think this is an important, thoughtful, and well-written paper. The child welfare system in the US is an important set of institutions serving a relatively large and very at-risk population, but prior to Doyle's research there was little policy-relevant research available to guide decisions for children on the margin of placement in foster care. The novel data collection took advantage of a large number of administrative data sets in order to track several different types of meaningful outcome variables. The marginal treatment effects framework is integrated very thoughtfully into the paper – blending nice econometrics with qualitative research on the child welfare system in a way that fits together very well. While the resulting estimates are somewhat imprecise, the point estimates suggest that large gains from foster care placement are unlikely for children at the margin of placement.

References

Almond, Douglas and Janet Currie, "Human capital development before age five," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics Volume 4B*, 2011, pp. 1487–1541.

__ **and** __ , "Killing me softly: The fetal origins hypothesis," *Journal of Economic Perspectives*, forthcoming.

__ **, Joseph Doyle, Amanda Kowalski, and Heidi Williams**, "Estimating marginal returns to medical care: Evidence from at-risk newborns," *Quarterly Journal of Economics*, 2010, *125* (2), 591–634.

__ **, Kenneth Chay, and David Lee**, "The costs of low birth weight," *Quarterly Journal of Economics*, 2005, *120* (3), 1031–1083.

Behrman, Jere and Mark Rosenweig, "Returns to birthweight," *Review of Economics and Statistics*, 2004, *86* (2), 586–601.

Bharadwaj, Prashant, Katrine Løken, and Christopher Nielson, "Early life health interventions and academic achievement," 2011. Mimeo, UCSD.

Black, Sandra, Paul Devereux, and Kjell Salvanes, "From the cradle to the labor market? The effect of birth weight on adult outcomes," *Quarterly Journal of Economics*, 2007, *122* (1), 409–439.

Currie, Janet and Duncan Thomas, "Does head start make a difference?," *American Economic Review*, 1995, *85* (3), 341–364.

__ **and Rosemary Hyson**, "Is the impact of health shocks cushioned by socioeconomic status? The case of low birth weight," *American Economic Review*, 1999, *89* (2), 245–250.

Deming, David, "Early childhood intervention and life-cycle skill development," *American Economic Journal: Applied Economics*, 2009, *1* (3), 111–134.

Doyle, Joseph, "Child protection and child outcomes: Measuring the effects of foster care," *American Economic Review*, 2007, *97* (5), 1583–1610.

Grossman, Michael, "On the concept of health capital and the demand for health," *Journal of Political Economy*, 1972, *80* (2), 223–255.

Heckman, James, "The economics, technology, and neuroscience of human capability formation," *Proceedings of the National Academy of Sciences*, 2007, *104* (33), 13250–13255.

Ludwig, Jens and Douglas Miller, "Does Head Start improve children's life chances? Evidence from a regression discontinuity design," *Quarterly Journal of Economics*, 2007, *122* (1), 159–208.

Oreopoulos, Philip, Mark Stabile, Randy Walld, and Leslie Roos, "Short, medium, and long term consequences of poor infant health: An analysis using siblings and twins," *Journal of Human Resources*, 2008, *43* (1), 88–138.

Royer, Heather, "Separated at girth: US twin estimates of the effects of birth weight," *American Economic Journal: Applied Economics*, 2009, *1* (1), 49–85.