Human Capital Development Before Age Five*

Douglas Almond† and Janet Currie‡

November 4, 2009

*We thank Maya Rossin for research assistance.

†Columbia University: da2152@columbia.edu

‡Columbia University: jc2663@columbia.edu

1 Introduction

The last 10 years have seen a boom in research on the long term effects of prenatal and early childhood conditions in a variety of fields. In Economics, the research focuses on the effects on human capital accumulation. A review of the *Journal of Political Economy, Quarterly Journal of Economics,* and *the American Economic Review* (excluding the Papers and Proceedings), suggests that there were no articles on this topic in 2000, but that there have been five or six per year since 2005. This work has been spurred by a growing realization that early life conditions can have a profound impact on later life. Table 1 summarizes several studies using longitudinal data which suggest that characteristics that are measured as of age 7 can explain a great deal of the variation in educational attainment, earnings as of the early 30s, and probability of employment. For example, McLeod and Kaiser (2004) use data from the National Longitudinal Surveys and find that children's test scores and background variables measured as of ages 6 to 8 predict about 12% of the variation in the probability of high school completion and about 11% of the variation in the probability of college completion. Currie and Thomas (1999) use data from the 1958 British Birth Cohort study and find that 4 to 5% of the variation in employment at age 33 can be predicted, and as much as 20% of the variation in wages. To put this later result in context, labor economists generally feel that they are doing well if they can explain 30% of the variation in wages in a human capital earnings function.

This chapter seeks to set out what Economists have learned about the importance of early childhood influences on later life outcomes, and about ameliorating the effects of negative influences. We begin with a brief overview of the theory which highlights the fact that identifying a causal relationship between a shock in early childhood and a future outcome does not imply that the effect in question is either biological or immutable. Parental and social responses are likely to be extremely important in either magnifying or mitigating the effects of a shock. Given that this is the case, it can sometimes be quite difficult to interpret the wealth of evidence that is starting to come in terms of an underlying structural framework.

The theoretical framework is laid out in Section 3 and followed by a brief discussion of methods in Section 4. We do not attempt to cover issues such as identification and instrumental variables methods which are covered in some depth elsewhere (c.f. Angrist and Pischke, 2009). Instead, we focus on several issues that come up frequently in the early influences literature, including estimation using small samples and the potentially high return to better data.

The fifth section of the paper discusses the evidence for long term effects of early life influences in greater detail, while the sixth focuses on the evidence regarding remediation programs. We conclude with a summing up and outstanding questions for future research in Section 7.

2 Conceptual Framework

Grossman [1972] models health as a stock variable that varies over time in response to investments and depreciation. Because some positive portion of the previous period's health stock vanishes in each period (e.g., age in years), the effect of the health stock and health investments further removed in time from the current period tends to fade out. As individuals age, the early childhood health stock and the prior health investments that it embodies become progressively less important.

In contrast, the "early influences" literature asks whether health and investments in early childhood have sustained effects on adult outcomes. The magnitude of these effects may persist or even increase as individuals age because childhood development occurs in distinct stages that are more or less influential of adult outcomes.

Defining h as health or human capital at the completion of childhood, we can retain the linearity of h in investments and the prior health stock as in Grossman [1972], but leave open whether there is indeed "fade out" (i.e. depreciation). For simplicity, we will consider a simple two-period childhood.[1] We can consider production of h:

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

where:

$$\begin{cases} I_1 & \cong & \text{investments during childhood through age 5} \\ I_2 & \cong & \text{investments during childhood after age 5.} \end{cases}$$

For a given level of total investments $I_1 + I_2$, the allocation of investments between period 1 and 2 will also affect h for $\gamma \neq .5$. If $\gamma > .5$, then health at the end of period 1 is more important to h than investments in the second period, and if $\gamma A > 1$, h may respond more than one-for-one with I_1. Thus, (1) admits the possibility that certain childhood periods may exert a disproportionate effect on adult outcomes that does not necessarily decline monotonically with age.This functional form says more than "early life" matters; it suggests that early-childhood events may be more influential than later childhood events.

2.1 Complementarity

The assumption that inputs at different stages of childhood have linear effects is common in economics. While it opens the door to "early origins", perfect substitutability between first and second period investments in (1) is a strong assumption. The absence of complementarity implies that all investments should be concentrated in one period (up to a discount factor) and no investments should be made during the low-return period. In addition, with basic preference assumptions, perfect substitutability "hard-wires" the optimal investment response to early-life shocks to be compensatory, as seen in Section 2.3.

[1]See Zweifel, Breyer, and Kifmann [2009] for a two period version of the Grossman [1972] model.

As suggested by Heckman [2007], a more flexible "developmental" technology is the constant elasticity of substitution (CES) function:

$$h = A\left[\gamma I_1^{\phi} + (1-\gamma) I_2^{\phi}\right]^{1/\phi}, \tag{2}$$

For a given total investment level $I_1 + I_2$, how the allocation between period 1 and 2 will also affect h depends on the elasticity of substitution $1/(1-\phi)$ and the share parameter γ. For $\phi = 1$ (perfect substitutability of investments), (2) reduces to (1).

Heckman [2007] highlights two features of "capacity formation" beyond those captured in (2). First, there may be "dynamic complementarities" which imply that investments in period t are more productive when there is a high level of capability in period $t-1$. For example, if the factor productivity term A in (2) were an increasing function of h_0, the health endowment immediately prior to period 1, this would raise the return to investments during childhood. Second, there may be "self-productivity" which implies that higher levels of capacity in one period create higher levels of capacity in future periods. This feature is especially noteworthy when h is multidimensional, as it would imply that "cross-effects" are positive, e.g. health in period 1 leads to higher cognitive ability in period 2. "Self-productivity" is more trivial in the unidimensional case like Grossman [1972] – even though the effect of earlier health stocks tends to fade out as the time passes, there is still memory as long as depreciation in each period is less than total ($\delta < 1$).

Here, we will use the basic framework in (2) to consider the effect of exogenous shocks μ_g to health investments that occur during the first childhood period.[2] We begin with the simplest case, where investments do not respond to μ_g (and denote these investments $\bar{I}_1$ and $\bar{I}_2$). Net investments in the first period are:

$$\bar{I}_1 + \mu_g.$$

[2] We include the subscript here because environmental influences at some aggregated geographic level g may provide exogenous variation in early childhood investments.

We assume that μ_g is independent of $\bar{I}_1$. While μ_g can be positive or negative, we assume $\bar{I}_1 + \mu_g > 0$. We will then relax the assumption of fixed investments, and consider endogenous responses to investments in the second period, i.e. $\delta I_2^* / \delta \mu_g$, and how this investment response may mediate the observed effect on h.

2.2 Fixed Investments

Conceptually, we can trace out the effect of μ_g while holding other inputs fixed, i.e., we assume no investment response to this shock in either period. Albeit implicitly, most biomedical and epidemiological studies in the "early origins" literature aim to inform us about this *ceteris paribus* relationship. Royer [2009] refers to this partial effect as the "biological" relationship between a shock μ_g and outcomes h.

In this two-period CES production function adopted from Heckman [2007], the impact of an early-life shock on adult outcomes is:

$$\frac{\delta h}{\delta \mu_g} = \gamma A \left[\gamma(\bar{I}_1 + \mu_g)^{\phi} + (1-\gamma)\bar{I}_2^{\phi} \right]^{(1-\phi)/\phi} (\bar{I}_1 + \mu_g)^{\phi - 1}. \tag{3}$$

The simplest production technology is the perfect substitutability case where $\phi = 1$. In this case:

$$\frac{\delta h}{\delta \mu_g} = \gamma A.$$

Damage to adult human capital is proportional to the share parameter on period 1 investments, and is unrelated to the investment level $\bar{I}_1$.

For less than perfect substitutability between periods, there is diminishing marginal productivity of the investment inputs. Thus, shocks experienced at different baseline investment levels have heterogenous effects on h. Other things equal, those with higher baseline levels of investment will experience more muted effects in h than those where baseline investment is low. A recurring empirical finding is that long-term damage due is shocks is more likely among poorer families [Currie and Hyson, 1999]. This is in part due to the fact that children in poorer families are subject to more or larger early-life shocks [Case, Lubotsky, and Paxson, 2002, Currie and Stabile, 2003]. However, it is also possible that the same shock will have a greater impact among children in poorer families if these children have lower period t investment levels to begin with. This occurs because they are on a steeper portion of the production function. *Ceteris paribus*, this would tend to accentuate the effect of an equivalent-sized μ_g shock on h among poor families.[3,4]

2.2.1 Remediation

Is it possible to alter "bad" early trajectories? In other words, what is the effect of a shock $\mu_g^{'} > 0$ experienced during the second period on h? Remediation is of interest to the extent that (3) is substantially less than zero. However, large damage to h from μ_g by itself says little about the potential effectiveness of remediation in the second period as both initial damage and remediation are distinct functions of the three parameters A, γ, and ϕ.

The effectiveness of remediation relative to initial damage is:

$$\frac{\delta h/\delta \mu_g^{'}}{\delta h/\delta \mu_g} = \frac{1-\gamma}{\gamma}\left(\frac{\bar{I}_1 + \mu_g}{\bar{I}_2 + \mu_g^{'}}\right)^{1-\phi}. \tag{4}$$

Thus, for $\bar{I}_1 > \bar{I}_2$ and a given value of γ, a unit of remediation will be more effective at low elasticities of substitution – the lack of $\bar{I}_2$ was the more critical shortfall prior to the shock. If $\bar{I}_1 < \bar{I}_2$ high elastiticites of substitution

[3] i.e. $\delta^2 h/\delta\mu_g \delta I_1 < 0$. On the other hand, $\delta^2 h/\delta\mu_g \delta I_2 > 0$ so lower period two investments would tend to reduce damage to h from μ_g. The ratio of the former effect to the latter is proportional to $\gamma/(1-\gamma)$ [Chiang, 1984]. Thus, damage from a period 1 shock is more likely to be concentrated among poor families when the period-1 share parameter (γ) is high.

[4] The cross-effect $\delta^2 h/\delta\mu_g \delta I_2$ is similar to dynamic complementarity, but Heckman [2007] reserves this term for the cross-partial between the stock and flow, i.e. $\delta^2 h/\delta h_0 \delta I_t$ for t=1, 2 in the example of Section **??**.

increase the effectiveness of remediation – adding to the existing abundance of $\bar{I}_2$ remains effective.

Fortunately, it is not necessary to observe investments and estimate all three parameters in order to assess the scope for remediation. Instead, we merely need to observe how a shock in the second period, $\mu_g^{'}$, affects h. Furthermore, this does not require a distinct shock in addition to μ_g. In an overlapping generations framework, the same shock, $\mu_g = \mu_g{}'$ could affect one cohort in the first childhood period (but not the second) and an older cohort in the second period (but not the first). For a small, "double-barrelled" shock, we would have reduced form estimates of both the damage in (3) and the potential to alter trajectories in (4).[5]

2.3 Responsive Investments

Most analyses of "early origins" focus on estimating the reduced from effect, $\delta h/\delta\mu_g$. Whether this empirical relationship represents a purely biological effect or also includes the effect of responsive investments is an open question. In general, to the extent that "early origins" are important, so too will any response of childhood investments to μ_g. For expositional purposes, we will consider $\mu_g < 0$ and responses that either magnify or attenuate initial damage.

Unless the investment response is costless, damage estimates which monetize $\delta h/\delta\mu_g$ alone will tend to understate total damage. In the extreme, investment responses could fully offset the effect of early-life shocks on h but this would not mean that such shocks were costless [Deschnes and Greenstone, 2007]. More generally, the damage from early-life shocks will be understated if we focus only on long-term effects and there are compensatory investments (i.e. investments that are negatively correlated with the early-life shock ($\delta I_2^*/\delta\mu_g < 0$)). The cost of investments which help remediate damage should be included. But even when the response is reinforcing ($\delta I_2^*/\delta\mu_g > 0$),

[5]How parameters of the production function might be recovered is discussed in Appendix A.

total costs can still be understated by focussing on the reduced form damage to h alone (see below).

To consider correlated investment responses more formally, we assume parents observe μ_g at the end of the first period. The direction of the investment response – whether reinforcing or compensatory – will be shaped by how substitutable period 2 investments are for those in period 1. If substitutability is high, the optimal response will tend to be compensatory, and thereby help offset damage to h.

A compensatory response is readily seen in the case of perfect substitutability. Cunha and Heckman [2007] observed that economic models commonly assume that production at different stages of childhood are perfect substitutes. When $\phi = 1$, (2) reduces to:

$$h = A\left[\gamma(I_1 + \mu_g) + (1-\gamma)I_2\right]. \tag{5}$$

This linear production technology is akin to that in Solon [1999], which also considered parental investments in children's human capital. Further, Solon [1999] assumed parent's utility trades off own consumption against their child's human capital:

$$U_p = U(C, h), \tag{6}$$

where p denotes parents and C their consumption. The budget constraint is:

$$Y_p = C + I_1 + I_2/(1+r). \tag{7}$$

With standard preferences, changes to h through μ_g will "unbalance" the marginal utilities in h versus C.[6] If μ_g is negative, the marginal utility of h becomes too high relative to that in consumption. The technology in (5) permits parents to convert some consumption into h at a constant rate. This will cause I_2^* to increase, which attenuates the effect of the μ_g damage. This attenuation

[6] Obviously, the marginal utilities themselves will not be the same but equal subject to discount factor, preference parameters, and prices of C versus I which have been ignored.

comes at the cost of reduced parental utility. Similarly, if μ_g is positive parents will "spend the bounty" (at least in part), reduce I_2^* and increase consumption. Again, this will temper effects on h, leading to an understatement of effects in analyses that ignore investments (or parental utility). In either case, prefect substitutability hard-wires the response to be compensatory.

The polar opposite technology is perfect complementarity between childhood stages, i.e., a Leontieff production function. Here, a compensatory strategy would be completely ineffective in mitigating changes to h. As h is determined by the minimum of period 1 and period 2 investments, optimal period 2 investments should reinforce μ_g. If μ_g is negative, parents would seek to reduce I_2 and consume more. Despite higher consumption, parents' utility is reduced on net due to the shock (or this bundle of lower h and higher C would have been selected absent μ_g). Again, the full-cost of a negative μ_g shock is understated when parental utility is ignored.

The crossover between reinforcing and compensating responses of I_2^* will occur at an intermediate parameter value of substitutability. (The fixed investments case of Section 2.2 can be seen to reflect an optimized response at this point of balance between reinforcing and compensating responses). The value of ϕ at this point of balance will depend on the functional form of parental preferences in (6), as shown for CES utility in Appendix B.

To take a familiar example, assume a Cobb-Douglas utility function of the form:

$$U_p = (1-\alpha)\texttt{log}C + \alpha\texttt{log}h. \tag{8}$$

If the production technology is also Cobb-Douglas ($\phi = 0$), then no change to I_2^* is warranted. If instead substitution between period 1 and period 2 is relatively easy ($\phi > 0$), compensating for the shock is optimal. If substitution is relatively difficult ($\phi < 0$), then parents should "go with the flow" and reinforce. For this reason, whether conventional reduced form analyses under or over-state "biological" effects (effects with I_2 held fixed) depends on how easy it is to substitute the timing of investments across childhood. If the elasticity of substitution across periods is low, then it may be optimal for parents to reinforce the effect of a shock.

Tension between preferences and the production technology may also be relevant for within-family investment decisions. For example, Behrman, Pollak, and Taubman [1982] considered parental preferences that parameterize varying degrees of "inequality aversion" among (multiple) children. Depending on the strength of parents' inequality aversion relative to the production technology (as reflected by ϕ), parents may reinforce or compensate exogenous within-family differences in early-life health and human capital. If substitutability between periods of childhood is sufficiently difficult (low ϕ), reinforcement of sibling differences may be optimal even in the presence of inequality aversion (see Appendix C).[7] Thus, empirical evidence that some parents reinforce early-life shocks could reveal less about "human nature" than it would reveal about the developmental nature of the childhood production technology.

3 Methods

As discussed above, we confine our discussion to a few issues that seem particularly germane to the early influences literature. One of these is the question of whether sibling fixed effects (or maternal fixed effects) estimation is appropriate. Fixed effects control for characteristics that are shared by siblings, such as aspects of maternal background that may not be observed by the researcher. But fixed effects cannot control for individual-specific factors that may affect

[7] Smaller within family estimates compared with between family estimates may reflect exacerbation of attenuation bias from measurement error or compensatory behavior by parents that promotes sibling equality [Griliches, 1979]. In contrast, Almond, Edlund, and Palme [2009] show that damage to future school outcomes from prenatal exposure to radiation is substantially *stronger* within families than in basic difference-in-differences specifications which may suggest reinforcing behavior by parents.

each sibling. The theory discussed above suggests that it may be optimal for parents to either reinforce or compensate for the effects of early shocks by altering their own investment behaviors. Whether parents do or do not reinforce/compensate obviously has implications for the interpretation of models estimated using family fixed effects. If on average, families compensate, then fixed effects estimates will understate the true "biological" effect of the shock. In some circumstances, such a bias might be benign in the sense that any significant coefficient could then be interpreted as a lower bound on the "true" biological effect. It is likely to be more problematic if parents systematically reinforce shocks, because then any effect that is observed results from a combination of underlying biological effects and parental reactions rather than the shock itself. In the extreme, if parents seized on a characteristic that was unrelated to ability and systematically favored children who had that characteristic, then researchers might wrongly conclude that the characteristic was in fact linked to success.

The issue of how parents allocate resources between siblings has received a good deal of attention in Economics, starting with Becker and Tomes (1976) and Behrman et al. (1982, 1989).[8] Some empirical studies from developing countries find evidence of reinforcing behavior (see Rosenzweig and Schultz, 1982; Rosenzweig and Wolpin, 1988; Pitt, Rosenzweig and Hassan, 1990). Empirical tests of these theories in developed countries such as the United States and Britain generally use adult outcomes such as completed education as a proxy for parental investments (see for example, Griliches, 1979; Behrman, Rosenzweig and Taubman 1995; Ashenfelter and Rouse, 1998).

Several recent studies have used birth weight as a measure of the child's endowment and assessed whether explicit measures of parental investments during early childhood are related to birth weight. For example, Datar, Kilburn, and Loughran [2010] use data from the National Longitudinal Survey of Youth-Child and show that low birth weight children are less likely to be

[8] See also, Becker (1992), Behrman, Pollak and Taubman (1995) and Mulligan (1997).

breastfed, have fewer well-baby visits, are less likely to be immunized, and are less likely to attend preschool than normal birth weight siblings. However, all of these differences could be due to poorer health among the low birth weight children. For example, if a child is receiving many visits for sick care, they may receive fewer visits for well care and this will not say anything about parental investment behaviors. Hence, Datar, Kilburn, and Loughran [2010] also look at how the presence of low birth weight siblings in the household affects the investments received by normal birth weight children. They find no effect of having a low birth weight sibling on breastfeeding, immunizations, or preschool. The only statistically significant interaction is for well-baby care. This could however, be due to transactions costs. It may be the case that if the low birth weight sibling is getting a lot of medical care, it is less costly to bring the normal birth weight child in for care as well, for example. Del Bono, Ermisch, and Francesconi [2008] also estimate a model that allows endowments of other children to affect parental investments in the index child. They find, however, that the results from this dynamic model are remarkably similar to those of mother fixed effects models in most cases. Moreover, although they find a positive effect of birth weight on breastfeeding, the effect is very small in magnitude.

We conducted our own investigation of this issue using data on twins from the Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), using twin differences to control for potential confounders. At the same time, twins routinely have large differences in endowment in the form of birth weight. Table 2 presents estimates for all twins (with and without controls for gender), same sex twins, and identical twins. Overall, there are very few significant differences in the treatment of these twins: Parents seemed to be more concerned about whether the low birth weight twin was ready for school, and to delay introducing solid food (but this is only significant in the identical twin pairs). We see no evidence that parents are more likely to praise, caress, spank or otherwise treat children differently, and despite their worries about school readiness, parents have similar expectations regarding college for both twins. This table largely replicates the basic finding of Royer [2009], who also considered parental investments and birth weight differences in the ECLS-B data. In particular, Royer [2009] focussed on investments soon after birth, finding that breastfeeding, NICU admission, and other measures of neonatal medical care did not vary with within twin pair birth weight differences.

The parental investment response has also been explored in the context of natural experiments. Kelly [2009] asked whether observed parental investments (e.g., time spent reading to child) were related to flu-induced damages to test scores in the 1958 British birth cohort study but did not detect an investment response. As noted above, Almond, Edlund, and Palme [2009] compare the effect of prenatal exposure to radiation on test scores estimated in a conventional difference-in-differences specification to a fixed effect specification that restricts comparisons to siblings. Estimated damage from Chernobyl was substantially stronger within families, suggesting that cognitive damage may have been reinforced (which could occur regardless of whether initial damage was attributed to Chernobyl.) In contrast, no difference was found in the likelihood or timing of having a subsequent sibling when an earlier child was more exposed prenatally to radiation, i.e., whether the quantity-quality relationship was observed with this shock to quality.[9]

In an interesting contribution to this literature, Hsin [2009] looks at the relationship between children's endowments, measured using birth weight, and maternal time use using data from the Child Supplement of the Panel Study of Income Dynamics. She finds that overall, there is little relationship between low birth weight and maternal time investments. However, she argues that this masks important differences by maternal socioeconomic status. In particular, she finds that in models with maternal fixed effects, less educated women spend less time with their low birth weight children, while more ed-

[9]For earlier work "running the quantity-quality experiment in reverse", see Rosenzweig and Wolpin [1988].

ucated women spend more time. This finding is based on only 65 sibling pairs who had differences in the incidence of low birth weight, and so requires some corroboration. Still, one interpretation of this result in the context of the Section 2 framework is that the elasticity of substitution between C and h varies by socioeconomic status. In particular, if $\varphi_{poor} > \varphi_{rich}$, low income parents tend to view their consumption and children's h as relatively good substitutes. This would lead low-income parents to be more likely to reinforce a negative shock than high-income parents (assuming that the developmental technology, captured by γ and ϕ, does not vary by socioeconomic status). A second possible interpretation of the finding is that parents' responses may reflect their budget constraint more than their preferences. If parents would like to invest in both children, but have only enough resources to invest adequately in one, then they may be forced to choose the more well endowed child (see Appendix C – TBA by dva). Interventions that relaxed resource constraints would have quite different effects in this case than in the case in which parents preferred to maximize the welfare of a favored child. More empirical work on this question seems warranted. For example, the PSID-CDS in 1997 and 2002 has time diary data for several thousand sibling pairs which have not been analyzed for this purpose.

Parent's choices are determined in part by the technologies they face, and these technologies may change over time, with potential implications for the potential biases in fixed effects estimates.[10] For example, Currie and Hyson [1999] asked whether the long term effects of low birth weight differed by various measures of parental socioeconomic status in the 1958 British birth cohort. They found little evidence that they did (except that low birth weight women from higher SES backgrounds were less likely to suffer from poor health as adults). But it is possible that this is because there were few effective

[10] For example, the effectiveness of remedial investments would change over time if γ varied with the birth cohort. Remediation would be more effective for later cohorts if $\gamma_t > \gamma_{t+1}$ in equation (4).

interventions for low birth weight infants in 1958. Currie and Moretti [2007] looked at Californian mothers born in the late 1960s and 70s and find that women born in low income zip codes were less educated and more likely to live in a low income zipcode than sisters born in better circumstances. Moreover, women who were low birth weight were more likely to transmit low birth weight to their own children if they were born in low income zip codes, suggesting that early disadvantage compounded the initial effects of low birth weight.

To the extent that behavioral responses to early-life shocks are important empirically, they will affect estimates of long-term effects whether family fixed effects are employed or not. These behaviors will be particularly important for fixed effects estimates if parents make comparisons across their children when deciding how to invest. Our conclusion is that users of these fixed effects designs should be particularly careful to consider any evidence that may be available about whether parents are reinforcing or compensating for the particular early childhood event at issue. Such parental behaviors will inform the appropriate interpretation of the estimates. There is relatively little evidence at present that parents in developed countries systematically reinforce or compensate for early childhood events, but more research is needed on this question.

3.1 Power

Given that there are relatively few data sets with information about early childhood influences and future outcomes, economists may be tempted to make use of relatively small data sets that happen to have the requisite variables. Power calculations can be helpful in determining whether these data sets are likely to yield any interesting findings. Table 3 provides two sample calculations. The first half of the table considers the relationship between birth weight and future educational attainment as in Black, Devereux and Salvanes (2007). Their key result was that a 1% increase in birth weight increased high school completion by .09 percentage points. The example shows that under reasonable assumptions about the distribution of birth weight and schooling attainment, it requires a sample of about 4000 children to be able to detect this effect in an OLS regression. We can also turn the question around and ask, given a sample of a certain size, how large would an effect have to be before we could be reasonably certain of finding it in our data? The second half of the table shows that if we were looking for an effect of birth weight on a particular outcome in a sample of 1,300 children, the coefficient on log(birth weight) would have to be at least .15 before we could detect it with reasonable confidence. If we have reason to believe that the effect is smaller, then it is not likely to be useful to estimate the model without more data.

3.2 Data Constraints

The lack of large-scale longitudinal data (i.e. data that follows the same persons over time) has been a frequent obstacle to evaluating the long-term impacts of early life influences. Yet, the answer may not always be to collect more longitudinal data. Drawbacks to longitudinal data collection include the fact that it is costly to collect; the fact that long term outcomes cannot be immediately assessed; and the fact that attrition often poses a serious problem, and increases costs.

In many cases, existing data can offer a potential solution to this problem. First, it may be possible to add retrospective questions to existing data collections. Second, it may be possible to merge new information to existing data sets. Third, it may be possible to merge several administrative data sets in order to address previously unanswerable questions. The major issue with each of these approaches is often data security, and each approach places different demands on the security of the data, as described below.

Smith (2009) and Garces, Thomas, and Currie (2002) are examples of adding retrospective questions to existing data collections. Smith had retrospective questions about health in childhood added to the Panel Study of Income Dynamics (PSID). The PSID began in the 1960s with a representative national sample, and has followed the original respondents and their family members every since. Using this data, Smith is able to show that adult respondents who were in poor health during childhood have lower earnings than their own siblings who were not in poor health. Such comparisons are possible because the PSID has data on large numbers of sibling pairs. Garces, Thomas, and Currie added retrospective questions about Head Start participation to the PSID, and were able to show that young adults who had attended Head Start had higher educational attainment, and were less likely to have been booked or charged with a crime than siblings who had not attended.

This method offers a powerful way to address questions about long-term impacts but it is not without its drawbacks. First, retrospective data may be reported with error, although it may be possible to assess the extent of reporting error using data from other sources. Second, only outcomes that are already in the data can be assessed. Still, the method is promising enough to suggest that on-going government funded data collections should be required to have some mechanism for researchers to suggest the addition of questions to waves of the surveys.

A second way to address long-term questions is to merge new information to existing data sets. The merge generally requires the use of geocoded data. For some purposes, such as exploring variations in policies across states, only a state identifier is required. For other purposes, such as examining the effects of traffic patterns on asthma, ideally the researcher would have access to exact latitude and longitude. There are many examples in which this approach has been successfully employed. For example, Ludwig and Miller (2007) study the long term effects of Head Start by exploiting the fact that the Office of Economic Opportunity initially offered the 300 poorest counties in the country assistance in applying for Head Start. They show using data from the National Educational Longitudinal Surveys that children who were in counties just poor enough to be eligible for assistance were much more likely to have attended Head Start than children in counties that were just ineligible. They go on to show that child mortality rates in the relevant age ranges were lower in counties whose Head Start enrollments were higher due to the OEO assistance. Using Census data they find that education is higher for people living in areas with higher former Head Start enrollment rates. Unfortunately however, the Census does not collect county of birth, so they cannot identify people who were born in these counties (there is obviously a good deal of measurement error involved in using county of residence as a proxy for county of birth, or county where someone went to school). A great deal of research on long-term outcomes could be facilitated by adding questions about county of birth and county at key ages (5, 14) to Census.

In another example, Currie and Gruber (1996) were able to examine the effects of the Medicaid expansions on the utilization of care among children by merging state-level information on Medicaid policy to data from the National Health Interview Survey (NHIS). At the time, this was only possible because one of the authors had access to the NHIS state codes through his work at the Treasury Department. It has since become easier to access geocoded health data either by traveling to Washington to work with the data, or by using it in one of the secure data centers that Census and the National Centers for Healh Statistics (NHCS) support. However, it remains a source of frustration to health researchers that NCHS does not make state codes and/or codes for large counties available on its public use data sets.

A third way of using existing data is to merge administrative data bases from several sources. Merging these data bases requires personal identifiers such as names and birth dates or social security numbers, and access to these data is a sensitive issue. Nevertheless, it remains a powerful way to address many questions of interest. Because such data is more readily available outside the United States, many examples of this approach use data from other countries. For example, Black Devereux and Salvanes (2005, 2007) use Norwegian data on all twins born over 30 years to look at long-term effects of birth weight, birth order, and family size on educational attainment. Currie, Stabile, Manivong, and Roos (forthcoming) use Canadian data on siblings to examine the effects of health shocks in childhood on future educational attainment and welfare use. Almond, Edlund, and Palme [2009] use Swedish data to look at the long term effects of low-level radiation exposure from the Chernobyl disaster on children's educational attainment.

In the U.S., Doyle (2008) uses administrative data from child protective services and the criminal justice system in Illinois to examine the effects of foster care. He shows first that there is considerable variation between foster care case workers in whether or not a child will be sent to foster care. Moreover, whether a child is assigned to a particular worker is random, depending on who is on duty at the time a call is received. Using this variation, Doyle shows that the marginal child assigned to foster care is significantly more likely to be incarcerated in future. These examples exploit large sample sizes, objective indicators of outcomes, sibling or cohort comparisons, as well as a long follow up period. Some limitations of using existing data include the fact that administrative data sets often contain relatively little background information, and that outcomes are limited to those that are collected in the data bases.

The major challenge to research that involves either merging new information to existing data sets, or merging administrative data sets to each other, is that privacy concerns are making it increasingly difficult to obtain data just as it is becoming more feasible to link them. In some cases, access to public use data has deteriorated. For example, for many years, individual level Vital Statistics Natality data from birth certificates included the state of birth, and the county (for counties with over 100,000 population). Since 2005, however, these data elements have been suppressed and it is now necessary to ask for special permission to obtain Vital Statistics data with geocodes.

There are several potential solutions to these problems. First, creators of large data sets need to be sensitive to the fact that their data may well be useful for addressing questions that they have not envisaged. In order to preserve the ability to use data to answer future questions, it is essential to retain information that can be used for linkage. At a minimum, this should include geographic identifiers at the smallest level of disaggregation that is feasible (for example a Census tract). Ideally, personal identifiers would also be preserved.

Second, more effort needs to be expended in order to make sensitive data available to researchers. Several approaches are feasible:

1. Suppress small cells or to merge small cells in public use data files. For example, NCHS data sets such as NHIS could be released with state identifiers for large states, and with identifiers for groups of smaller states.

2. Add small amounts of "noise" to public use data sets, or do data swapping in order to prevent identification of outliers. For example, Cornell University is coordinating the NSF-Census Bureau Synthetic Data Project which seeks to develop public-use "analytically valid synthetic data" from micro datasets customarily accessed at secure Census Research Data Centers.

3. Create model servers. In this approach, users log in to estimate models using the true data, but get back output that does not allow individuals to be identified.

4. Data use agreements. The National Longitudinal Survey of Youth and the National Educational Longitudinal Survey have successfully employed data use agreements with qualified users for many years, and without any documented instances of data disclosure.

5. Creation of de-identified merged files. Currie, Neidell, and Schmieder [2009] asked the state of New Jersey to merge birth records with infor-

mation about the location of pollution sources, and create a de-identified file. This allows them to study the effect of air pollution on infant health.

6. Secure data facilities. The Census Research Data Centers have facilitated access to much confidential data, but do impose large costs on researchers who are not located close to the facilities.

These ideas have been explored in the statistics literature for more than 20 years (see Delenius, and Reiss, 1982), and have been much discussed at Census (see for example, Reznek, 2007).

In summary, there are many secrets currently locked in existing data that researchers do not have access to. We need to explore ways to make these data available. In many cases, this will be a more cost effective and timely way to answer important questions than carrying out new data collections.

4 Empirical Literature: Evidence of Long Term Consequences

> *What is of importance is the year of birth of the generation or group of individuals under consideration. Each generation* ***after the age of 5 years*** *seems to carry along with it the same relative mortality throughout adult life, and even into extreme old age.*
> Kermack, McKendrick, and McKinlay [1934] in *The Lancet* (emphasis added).

In this section, we summarize recent empirical research finding that experiences before five have persistent effects, shaping human capital in particular. A hallmark of this work is the attention paid to identification strategies that seek to isolate causal effects of the early childhood environment. An intriguing sub-current is the possibility that some of these effects may remain latent during childhood (at least from the researcher's perspective) until manifested in either adolescence or adulthood. Recently, economists have begun to ask how parents or other investors in human capital (e.g. school districts) *respond* to early-life shocks, as suggested by the conceptual framework in Section 2.3.

As the excerpt above from Kermack, McKendrick, and McKinlay [1934] indicates, the idea that early childhood experiences may have important, persistent effects did not originate recently, nor did it first appear in economics. An extensive epidemiological literature has focussed on the early childhood environment, nutrition in particular, and its relationship to health outcomes in adulthood. For a recent survey, see Gluckman and Hanson [2006]. This literature has been criticized, even within epidemiology, for credulous empirical comparisons (see, e.g. Lan [2001], Rasmussen [2001]). Absent clearly-articulated identification strategies, health determinants that are difficult to observe and are therefore omitted from the analysis (e.g., parental concern) are presumably correlated with the treatment and can thereby generate the semblance of "fetal origins" linkages, even when fetal effects do not exist.

4.1 Prenatal Environment

In the 1990s, DJ Barker popularized and developed the argument that disruptions to the prenatal environment presage chronic health conditions in adulthood, including heart disease and diabetes [Barker, 1992]. Growth is most rapid prenatally and in early childhood. When growth is rapid, disruptions to development caused by the adverse environmental conditions may exert life-long health effects. Barker's "fetal origins" perspective contrasted with the view that pregnant mothers functioned as an effective buffer for the fetus against environmental insults.[11]

In Table 4, we categorize prenatal environmental exposures into three

[11] For example, it has been argued that nausea and vomiting in early pregnancy (morning sickness) is an adaptive response to prevent maternal ingestion of foods that might be noxious to the fetus.

groups. Specifically, we differentiate among factors affecting maternal and thereby fetal health (e.g. nutrition and infection), economic shocks (e.g. recessions), and pollution (e.g. lead).

Currie and Hyson [1999] broke ground in economics by exploring whether "fetal origins" (FO) effects were confined to chronic health conditions in adulthood, or might extend to human capital measures. Using the British National Child Development Survey, low birth weight children were more than 25% less likely to pass English and math O-level tests, and were also less likely to be employed. The finding that test scores were substantially affected was surprising as epidemiologists routinely posited fetal "brain sparing" mechanisms, whereby adverse *in utero* conditions were parried through a placental triage that prioritized neural development over the body, see, e.g., Scherjon, Oosting, de Visser, de Wilded, Zondervan, and Kok [1996]. Furthermore, Stein, Susser, Saenger, and Marolla [1975]'s influential study found no effect of prenatal exposure to the Dutch Hunger Winter on IQ [Stein, Susser, Saenger, and Marolla, 1975].

Currie and Hyson [1999] were followed by a series of papers that exploited differences in birthweight among siblings and explored their relationship to sibling differences in completed schooling. In relatively small samples (approximately 800 families), Conley and Bennett [2001] found negative but imprecise effects of low birth weight on educational attainment. Statistically significant effects of low birth weight on educational attainment were found when birth weight was interacted with being poor, but in general sample size prevented detection of all but the largest effects (see Section 3.1). Using a comparable sample size, Behrman and Rosenzweig [2004] found the schooling of identical female twins was nearly one-third of a year longer for a pound increase in birth weight (454 grams), with relatively imprecise effects on adult BMI or wages.

In half a million birth records for California, Currie and Moretti [2007] matched mothers to their sisters. Here, low birth weight was found to have statistically significant negative impacts on educational attainment and the likelihood of living in a wealthy neighborhood. However, the estimated magnitudes of the main effects were more modest: low birth weight increased the likelihood of living in a poor neighborhood by 3% and reduced educational attainment approximately one month on average. Like Conley and Bennett [2001], the relationship was substantially stronger for the interaction between low birth weight and being born in poor neighborhoods.

In a sample of Norwegian twins, Black, Devereux, and Salvanes [2007] also found long-term effects of birth weight, but did not detect any heterogeneity in the strength of this relationship by parental socioeconomic status.[12] Oreopoulos et al. [2008] find similar results for Canada and Lin and Lui (2007) find positive long term effects of birth weight in Taiwan. Royer [2009] found long-term health and educational effects within California twin pairs, with a weaker effect of birth weight than several other studies, esp. Black, Devereux, and Salvanes [2007]. Responsive investments could account for this discrepancy if they differed between California and elsewhere (within twin pairs). Alternatively, there may be more homogeneity with respect to socioeconomic status elsewhere than in California. As described in Section 3, Royer [2009] analyzed investment measures directly with the ECLS-B data, concluding that her estimates of long-term effects may indeed represent biological ones (see Section 2.2).

Following a literature in demography on seasonal health effects, Doblhammer and Vaupel [2001] and Costa and Lahey [2005] focused on the potential long-term health effects of birth season. A common finding is that in the northern hemisphere, people born in the last quarter of the year have longer life expectancies than those born in the second quarter. Both the availability of nutrients can vary seasonally (particularly historically), as does the likelihood of common infections (e.g., pneumonia). Therefore, either nutrition or

[12] Royer [2009] notes that Black, Devereux, and Salvanes [2007] find a "negligible effect of birth weight on high school completion for the 1967-1976 birth cohort, but for individuals born between 1977 and 1986, the estimate is nearly six times as large."

infection could drive this observed pattern. Almond [2006] focused on prenatal exposure to the 1918 Influenza Pandemic, estimating that children of infected mothers were 15% less likely to graduate high school and wages were between 5 and 9% lower. Kelly [2009] found negative effects of prenatal exposure to 1957 "Asian flu" in Britain on test scores, though the estimated magnitudes were relatively modest. Interestingly, while birth weight was reduced by flu exposure, this effect appears to be independent of the test score effect. Like Royer [2009], Kelly [2009] considered whether observed parental investments (e.g., time spent reading to child) were related to flu-induced damages to test scores, but did not detect an investment response. Finally, Field, Robles, and Torero [2009] found that prenatal iodine supplementation raised educational attainment in Tanzania by half a year of schooling, with larger impacts for girls.

A second set of papers considers economic shocks around the time of birth. Here, health in adulthood tends to be the focus (not human capital), and findings are perhaps less consistent than in the studies of nutrition and infection described above. Berg et al. [2006]'s basic result is that adult survival in the Netherlands is reduced for those born during economic downturns. In contrast, Cutler et al. [2007] detected no long term morbidity effects in the Health and Retirement Survey data for cohorts born during the Dustbowl era of 1930s. Banerjee, Duflo, Postel-Vinay, and Watts [2009] found that shocks to the productive capacity of French vineyards did not have detectable effects on life expectancy or health outcomes, but did reduce height in adulthood. Baten, Crayen, and Voth [2007] related variations in grain prices in the decade of birth to numeracy using an ingenious measure based on "age heaping" in the British Censuses between 1851 and 1881. Persons who are more numerate are less likely to round their ages to multiples of 5 or 10. They find that children born in decades with high grain prices were less numerate by this index.

The third strand of the literature examines the effect of pollution on fetal health. Epidemiological studies have demonstrated links between very severe pollution episodes and mortality: one of the most famous focused on a "killer fog" in London, England and found dramatic increases in cardiopulmonary mortality (Logan and Glasg, 1953). Previous epidemiological research on the effects of moderate pollution levels on prenatal health suggests negative effects but have produced inconsistent results. Cross-sectional differences in ambient pollution are usually correlated with other determinants of fetal health, perhaps more systematically than with nutritional or disease exposures considered above. Many of the pollution studies have minimal (if any) controls for these potential confounders. Banzhaf and Walsh [2008] found that high-income families move out of polluted areas, while poor people in-migrate. These two groups are also likely to provide differing levels of (non-pollution) investments in their children, so that fetuses and infants exposed to lower levels of pollution may tend to receive, e.g., better quality prenatal care. If these factors are unaccounted for, this would lead to an upward bias in estimates. Alternatively, certain pollution emissions tend to be concentrated in urban areas, and individuals in urban areas may be more educated and have better access to health care, factors that may improve health. Omitting these factors would lead to a downward bias, suggesting the overall direction of bias from confounding is unclear.

Two studies by Chay and Greenstone [2003a,b] address the problem of omitted confounders by focusing on "natural experiments" provided by the implementation of the Clean Air Act of 1970 and the recession of the early 1980s. Both the Clean Air Act and the recession induced sharper reductions in particulates in some counties than in others, and they use this exogenous variation in levels of pollution at the county-year level to identify its effects. They estimate that a one unit decline in particulates caused by the implementation of the Clean Air Act (recession) led to between five and eight (four and seven) fewer infant deaths per 100,000 live births. They also find some evidence that the decline in TSPs led to reductions in the incidence of low birth weight. However, the levels of particulates studied by Chay and Greenstone are much higher than those prevalent today; for example, PM10 levels have fallen by nearly 50 percent from 1980 to 2000.

Several recent studies consider natural experiments at more recently-encountered pollution levels. For example, Currie, Neidell, and Schmieder [2009] use data from birth certificates in New Jersey in which they know the exact location of the mothers residence, and births to the same mother can be linked. They focus on a sample of mothers who live near pollution monitors and show that variations in pollution from carbon monoxide (which comes largely from vehicle exhaust) reduces birth weight and gestation. Currie and Walker [2009] exploit a natural experiment having to do with introduction of electronic toll collection devices (E-ZPass) in New Jersey and Pennsylvania. Since much of the pollution produced by automobiles occurs when accelerating and decelerating, electronic toll collection greatly reduces auto emissions in the vicinity of a toll plaza. Currie and Walker [2009] compare mothers near toll plazas to those who live near busy roadways but further from toll plazas and find that E-ZPass increased birth weight and gestation. They show that they obtain similar estimates following mothers near toll plazas over time and estimating mother fixed effects models. These papers are notable in part because it has proven more difficult to demonstrate effects of pollution on fetal health than on infant health, as discussed further below. Hence, it appears that being *in utero* may be protective against at least some forms of toxic exposure (such as particulates).

This literature on the effects of air pollution is closely related to work on the effects of smoking on infant health. Smoking is, afterall, the most important source of indoor air pollution. Medical research has shown that nicotine constricts the oxygen supply to the fetus, so there is an obvious mechanism for smoking to affect infant health. Indeed, there is near unanimity in the medical literature that smoking is the most important preventable cause of low birth weight. Economists have focussed on ways to address heterogeneity in other determinants of birth outcomes that is likely associated with smoking. Tominey [2007] found that relative to a conventional multivariate control specification, roughly one-third of the harm from smoking to birth weight is explained by unobservable traits of the mother. Moreover, the reduction in birth weight from smoking was substantially larger for low-SES mothers. In a much larger sample, Currie, Neidell, and Schmieder [2009] showed that smoking significantly reduced birth weight, even when comparisons are restricted to within-sibling differences. Moreover, Currie, Neidell, and Schmieder [2009] document a significant interaction effect between exposure to carbon monoxide exposure and infant health in the production of low birth weight, which may help explain the heterogeneity in birth weight effects reported by Tominey [2007]. Aizer and Stroud [2009] note that impacts of smoking on birth weight are generally much smaller in sibling comparisons than in OLS and matching-based estimates. Positing that attenuation bias is accentuated in the sibling comparisons, Aizer and Stroud [2009] use serum cotinine levels as an instrument for measurement error in smoking and find that sibling comparisons yield similar birth weight impacts (around 150 grams). Lien and Evans [2005] use increases in state excise taxes as an instrument for smoking and find large increases in birth weight (182 grams) as a result. Using propensity score matching, Almond, Chay, and Lee [2005] document a large decrease in birth weight from prenatal smoking (203 grams), but argue that this weight decrease is weakly associated with alternative measures of infant health, such as prematurity, APGAR score, ventilator use, and infant mortality.

Some recently-released data will enable new research on smoking's short and long-term effects. In 2005, twelve states began using the new U.S. Standard Certificate of Live Birth (2003 revision). Along with other new data elements (e.g., on surfactant replacement therapy), smoking behavior *by trimester* is reported. It will be useful to consider whether smoking's impact on birth weight varies by trimester, and also whether smoking is more closely tied to other measures of newborn health if it occurs early versus late pregnancy. Second, there is relatively little research by economists on the long-term effects of prenatal exposure to smoking. Between 1990 and 2003, there were 113 increases in state excise taxes on cigarettes [Lien and Evans, 2005].[13] Since 2005, the American Community Survey records both state and quarter of birth, permitting linkage of these data to the changes in state excise taxes during pregnancy.

To summarize, the recent "fetal origins" literature in economics finds substantial effects of prenatal health on subsequent human capital and health. As we discuss in Section 5, this suggests a positive role for policies that improve human capital by affecting the birth endowment. That is, despite being congenital (i.e. present from birth), this research indicates that the birth endowment is malleable in ways that shape human capital. This finding has potentially important implications for public policy since it suggests that one of the more effective ways to improve children's long term outcomes might be to target women of child bearing age in addition to focusing on children after birth.

[13] Some states enacted earlier excise taxes: the "average state tax rate increased from 5.7 cents in 1964 to 15.5 cents in 1984" [Farrelly, Nimsch, and James, 2003]; high 1970s inflation can be an additional potential source of identification as excise taxes were set nominally.

4.2 Early Childhood Environment

The absence of detectable long-term effects from a very severe early childhood shock (e.g., a head injury, or emotional trauma) would be surprising. Therefore, a more interesting set of questions considers how developmental linkages operating at the individual level affect human capital formation in the aggregate. For this, we need to know how many children are affected by negative early childhood experiences that could plausibly exert persistent effects. In addition, how big and long-lasting are the effects of less severe early childhood shocks relative to more severe shocks? Taken together, how much of the differences in adult attainments might be accounted for by things that happen to children between birth and age five? Furthermore, how are these linkages between shocks and outcomes mediated or moderated by third factors? For example, is the effect of childhood lead exposure on subsequent test scores stronger for families of lower socioeconomic status (i.e. is the interaction with SES an important one) and if so why? Alternatively, is the effect of injury mediated by health status, or is the causal pathway a direct one to cognition?

We might also wish to know how parents respond to early childhood shocks. To date, there has been less focus on this question in the early childhood period than in the prenatal period, perhaps because it seems less plausible to hope to uncover a "pure" biological effect of a childhood shock given that children are embedded in families and in society. However, this opens the possibility to a richer set of behavioral responses – of the kind considered by economists – might be at play. Furthermore, early childhood admits a wider set of environmental influences than the prenatal period. For example, abuse in early childhood can be distinguished from malnutrition, a distinction more difficult for the *in utero* period, and these may have quite different effects.

We define early childhood as starting at birth and ending at age five. From an empirical standpoint, early childhood so defined offers advantages and disadvantages over analyses that focus on the prenatal period. Mortality is substantially lower during early childhood than *in utero*, which reduces the scope for selective attrition caused by environmental shocks to affect the composition of survivors. On the other hand, it is unlikely that environmental sensitivity during early childhood tapers discontinuously at any precise age (including age five). From a refutability perspective, we cannot make sharp temporal comparisons of a cohort "just exposed" to a shock during early childhood to a neighboring cohort "just unexposed" by virtue of its being too old to be sensitive. Moreover, it will often be difficult to know *a priori* whether prenatal or postnatal exposure is more influential.[14] Thus, studies of early childhood exposures tend to emphasize cross-sectional sources of variation, including that at the geographic and individual level. The studies reviewed in this section focus on tracing out the relationships between events in early childhood and future outcomes, and are summarized in Table 5.

4.2.1 Infections

Insofar a specific health shocks are considered, infections are the most commonly studied. In epidemiology, long-term health effects of infections – and the inflammation response they trigger – has been explored extensively, e.g. Crimmins and Finch [2006]. Outcomes analyzed by economists include height, health status, educational attainment, test scores, and labor market outcomes. The estimated impacts tend to be large. Using geographic differences in hookworm infection rates across the US South, Bleakley [2007] found that eradication after 1910 increased their subsequent literacy rates but did not increase the amount of completed schooling, except for Black children. The literacy improvement was much larger among Blacks than Whites, and stronger among women then men. The return to education increased substantially, and Bleakley [2007] estimated that hookworm infection throughout childhood reduced wages in adulthood by as much as 40%. Case and Paxson [2009] focussed

[14] For example, early postnatal exposure to Pandemic influenza apparently had a larger impact on hearing than did prenatal flu exposure [Heider, 1934].

on reductions in U.S. childhood mortality from typhoid, malaria, measles, influenza, and diarrhea during the first half of the 20th Century. They found that improvements in disease environment in one's state of birth were mirrored in improved cognitive performance at older ages, but like Bleakley [2007], this effect did not seem to operate through increased years of schooling. However, the estimated cognitive impacts in Case and Paxson [2009] were not robust to the inclusion of state-specific time trends in their models. Chay, Guryan, and Mazumder [2009] found that reduced exposure to pneumonia and diarrhea in early childhood during the late 1960s raised subsequent AFQT and NAEP scores towards those of Whites. Changes in postneonatal mortality rates (dominated by infections) explained between 50 and 80 percent or the (large) reduction in the Black-White AFQT gap. Finally, Bozzoli, Deaton, and Quintana-Domeque [2009] highlight that in developing countries, high average mortality rates cause the selection effect of early childhood mortality to overwhelm the "scarring" effect. Thus, the positive relationship between early childhood health and subsequent human capital may be absent in analyses that do not account for selective attrition in high mortality settings.

4.2.2 Health Status

Many of the studies reviewed in Table 5 investigate the link between health in childhood and future cognitive or labor market outcomes. These studies can be viewed as a subset of a broader literature asking whether income affects health, and how health affects income? For example, using cross-sectional U.S. data, Case, Lubotsky and Paxson (2002) find a striking relationship between family income and a child's reported health status, which becomes stronger as children age. Their motivation for looking at children is that the child's health is unlikely to have a large direct effect on family income, so that the direction of causality is relatively clear. Currie and Stabile (2003) investigate this relationship using Canadian panel data and argue that one reason the relationship between income and child health increases over time is that poorer children are subject to many more negative health shocks. In fact, in Canada, this is the dominant mechanism driving the relationship (which is not surprising given that all Canadian children have public health insurance so that gaps in treatment rates are small).[15]

The question we focus on here is how much poor health in childhood, in turn, affects future outcomes. One of the chief ambiguities in answering this question is what we mean by health in childhood. While it has become conventional to measure fetal health using birth weight (though there may be better measures, see Almond, Chay and Lee, 2005) there are a wide variety of different possible measures of child health, ranging from maternal reports about the child's general health status through questions about diagnoses of specific chronic conditions, to the occurrence of "adverse events." Case and Paxson (2008a,b) do not have a direct measure of child health, but argue that adult health is a good proxy for child health. One useful distinction that is emerging in the literature is between mental and physical health conditions. A second problem is that it is often unclear whether the ill health dates from a particular period (e.g. an injury) or whether it might reflect a continuing, perhaps a congenital, condition. For example, Smith (2009) uses data from the Panel Study of Income Dynamics which asked young adults a retrospective question about their health status before age 16. In models with sibling fixed effects, he finds that the sib with the worse health had significantly lower earnings, although educational attainment was not significantly affected. He also finds using data from the Health and Retirement Survey that reports of

[15] Conliffe and Link (2008) argue that in the U.S. differential access to care also plays a role in the steepening of the relationship between income and child health with age. A number of studies have investigated this relationship, dubbed "the gradient," in other countries (c.f. Currie, Shields, and Price, 2007; Case, Lee and Paxson, 2008; Doyle Harmon and Walker, 2005; Khanam Nghiem, and Connelly, forthcoming). Kahn et al. (2005) and Propper et al. (2007) are particularly interesting because they find that when maternal mental health is controlled, the relationship disappears, suggesting that it is mediated largely by factors that affect maternal mental health.

general poor health in childhood do tend to be correlated in the expected way with the presence of specific health conditions. However, it is not possible to ascertain that the negative effects are due to poor health at any particular "critical" window. Salm and Schunk (2008) attempt to deal with this problem using detailed health information from a medical examination of young German children entering school. In models with sibling fixed effects, they find a significant relationship between poor mental health and asthma on the one hand, and measures of cognitive functioning on the other. They control for the child's birth weight in an effort to distinguish between the effects of health at birth and health after birth (though to the extent that birth weight is an imperfect measure of health at birth, it is possible that the other health measures partly capture congenital conditions).

Currie et al. (2009) use administrative data from the Canadian public health insurance system to follow children from birth through young adulthood. Using information about all contacts with medical providers, they construct measures of whether children suffered injuries, asthma, mental health problems or other health problems at ages 0 to 3, 4 to 8, 9 to 13, and 14 to 18. It is interesting that even in a large sample, there were relatively few children with specific health problems other than injuries, asthma, or mental health problems, so that it was necessary to group the remaining problems together. They then look at the relationship between health at various ages, educational attainment, and use of social assistance as a young adult in sibling fixed effects models that also control flexibly for birth weight and the presence of congenital anomalies. The results are perhaps surprising in view of the conceptual framework developed in Section 3. When entered by themselves, early childhood health conditions (at age 0-3 and at age 4-8) are predictive of future outcomes, conditional on health at birth. However, when early physical health conditions are entered along with later ones, generally only the later ones matter. This result suggests that physical health in early childhood affects future outcomes largely because it affects future health (i.e., subsequent health mediates the relationship), and not because there is a direct link between early physical health status and cognition. In contrast, mental health conditions at early ages seem to have significant negative effects on future outcomes even if there are no intermediate report of a mental health condition. This result suggests that common mental health problems such as Attention Deficit Hyperactivity Disorder (ADHD, also called Attention Deficit Disorder or ADD) or Conduct Disorders (i.e. disorders usually involving abnormal aggression and anti-social behavior) may impair the process of human capital accumulation even if they do not lead to diagnoses of mental health disorders in adulthood.

Several recent papers focus specifically on measures of mental health conditions. Currie and Stabile (2006) use questions similar to those on mental health "screeners" which were administered to large samples of children in the U.S. and Canada in two national surveys. They find that children whose scores indicated mental health problems in 1994 had worse outcomes as of 2002-4 than siblings without such problems. They controlled for birth weight (among other variables) and estimated models with and without including children with diagnosed learning disabilities. In all specifications, they found negative effects of high ADHD scores on test scores on schooling attainment. Smith and Smith (2008) report similar results using data from the PSID which includes retrospective questions about mental health problems before age 16. Like Smith (2009) and Currie et al. (2009) they estimate models with sibling fixed effects, and find significant long term effects of mental health conditions which are much larger than those of physical health conditions. Vujic et al. (2008) focus on conduct disorders using a panel of Australian twins and find that conduct disorder before age 18 has strong negative effects on the probability of high school graduation as well as positive effects on the probability of criminal activity. None of these three papers focus specifically on measures of mental health conditions before age 5 but "externalizing" mental health conditions such as ADHD and Conduct Disorder typically manifest themselves at early ages. Finally, although they are conceptually distinct, many survey measures of mental health resemble measures of "non-cognitive skills." Hence, one might interpret evidence that non-cognitive skills in childhood are important determinants of future outcomes as further evidence of the importance of early mental health conditions (Blanden, Gregg and McMillan, 2007; Heckman and Rubinstein, 2001).

4.2.3 Home Environment

The home is one of the most important environments affecting a young child and there is a vast literature in related disciplines investigating the relationship between different aspects of the home environment and child outcomes. We do not attempt to summarize this literature here, but pick three aspects that may be most salient: Maternal mental health and/or substance abuse, maternal employment, and child abuse/foster care (which may be considered to be an extreme result of bad parenting). Given the importance of child mental health and non-cognitive skills, it is interesting to ask how maternal mental health affects child outcomes? Frank and Meara (2009) examine this question using data from the National Longitudinal Survey of Youth. They include a rich set of control variables (mother's cognitive test score, grandparent's substance abuse, permanent income); estimating models with mother fixed effects; and propensity scores. Their estimates suggest large effects (relative to the effects of income) of contemporaneous maternal depression on the quality of the home environment and on children's behavioral problems, but little effect on math and reading scores. Estimates of the effects of maternal substance abuse are mixed, which echoes the findings of Chatterji and Markowitz (2001) using the same data. Unfortunately, the authors are not able to look at the long term effects of maternal depression experienced by children aged 0 to 5 because the depression questions in the NLSY have been added only recently. As these panel data are extended in time, further investigation of this issue is warranted.

There is also a large literature, including some papers by Economists, examining the effect of maternal employment at early ages on child outcomes. Much of this literature suffers from the lack of an appropriate conceptual framework. If we think of child outcomes being produced via some combination of inputs, then the important question is how maternal employment affects the inputs chosen? This will evidently depend on how much her employment income relaxes the household budget constraint, and the price and quality of the child care alternatives that are available. Some of the literature on maternal employment seems to implicitly assume that the mother's time is such an important and unique input that no purchased input can adequately replace it. This may possibly be the case but is a strong assumption. If the mother's time is replaceable at some price, then one might expect maternal employment to have quite different effects on women with different levels of household income (moreover, mother's time may not all be of equal quality, so that it is easier to replace some mother's time than others with the market). This argument suggests that it is extremely important to consider explicitly the quality of the mother's time inputs and the availability of potential substitute inputs in models of maternal employment, something that is difficult to do in most available data sets. Studies that rely on regression methods and propensity score matching (see Hill et al. 2005; Ruhm 2003) often find small negative effects of maternal employment (especially in the first year) on children's cognitive development. However, two recent studies using variation in maternity leave provisions find that while more generous maternity leave policies are associated with increased maternal employment, there is little effect on children's outcomes (Baker and Milligan, forthcoming; Dustmann and Schoenberg, 2008). Dustmann and Schoenberg (2008) have data that permit cohorts affected by expansions in German maternity leave laws to be followed for many years. They see no effect of maternal employment on educational attainment or wages.

Finally, there are a few papers examining the effects of child abuse/foster care on child outcomes. This is a difficult area to investigate because it is hard to imagine that abuse (or neglect) can be divorced from other characteristics of the household. Currie and Widom (2009) use data from a "prospective longitudinal study in which abused children (the treatments) were matched to controls. After following these children until their mid 40s, they found that the abused children were less likely to be employed, had lower earnings, and fewer assets, and that these patterns were particularly pronounced among women. It is possible that these results are driven by unobserved differences between the treatments and controls although focusing on various subsets of the data (e.g. children whose mothers were on welfare; children of single mothers) produced similar results. Currie and Tekin (2009) use data from the National Longitudinal Study of Adolescent Health to examine the effect of having been abused before age 7 on the propensity to commit crime. They find strong effects which are quite similar in OLS, sibling, and twin fixed effects models. It is possible that these results reflect a characteristic of an individual child (such as difficult temperament) which makes it both more likely that they will be abused and more likely that they will commit crime. However, controlling directly for measures of temperament and genetic endowments does not alter the results. The Doyle (2008) study of the effects of foster care on the marginal child is also summarized in Table 5.

4.2.4 Toxic Exposures

Epidemiological studies of postnatal pollution exposure and infant mortality have yielded mixed results and are likely to suffer from omitted variables bias (Woodruff et al., 2009). Currie and Neidell [2005] examine the effect of more recent (lower) levels of pollution on infant health, along with the role of specific pollutants in addition to particulates (only TSPs were measured during the time periods analyzed by Chay and Greenstone [2003a,b]). Using within-zip code variation in pollution levels, they find that a one unit reduction in carbon monoxide over the 1990s in California saved 18 infant lives per 100,000 live births. However, unlike Currie, Neidell, and Schmieder [2009] they were unable to find any consistent evidence of pollution effects on health at birth, probably because of the crudeness of their measure of maternal location.

Reyes [2007] found large effects of banning leaded gasoline on crime in the U.S., but results were not robust to state-specific time trends despite a relatively long panel of state-level lead measurements. Nilsson [2009] considered reductions in ambient lead levels in Sweden following the banning of lead in gasoline and measure possible exposures using the concentrations of lead in 1,000 moss (bryophyte) collection sites that have been maintained by the Swedish environmental protection agency since the early 1970s. Nilsson [2009] found that early childhood exposure reduced human capital, as reflected by both grades and graduation rates. These effects persisted when comparisons were restricted within siblings, and were substantially larger for low-income families.

4.2.5 Summary re: Long Term Effects of Fetal and Early Childhood Environment

The last 10 years have seen an upsurge of empirical work on the long-term effects of early childhood. As a result, much has been learned. We can state fairly definitively that at least some things that happen before age five have long-term consequences for health and human capital. Moreover, these effects are sufficiently large and general to shape outcomes at the population level. On balance, effects of fetal exposure tend to be somewhat larger than postnatal effects, but there are important exceptions. Mental health is a prime example. Mental health conditions and non-cognitive skills seem to have large, persistent effects independent of those exerted by the prenatal environment.

5 Empirical Literature: Policy Responses

The evidence discussed above indicates that prenatal and early childhood often have a critical influence on later life outcomes. However, by itself this evidence says little about the effectiveness of remediation. Hence, this section discusses evidence about whether remediation in the zero to five period can be effective in shaping future outcomes. In so doing, we take a step away from explicit consideration of an early-childhood shock u_g as in Section 2. Instead, we focus on the specific public policies that may be able to alter an improve developmental trajectories, usually in disadvantaged sub-populations. We begin with programs that raise income, and then move on to programs that target specific domains. The emphasis is on recent studies with credible research designs, though given how quickly the research base is growing, we will inevitably have neglected some worthy studies.

5.1 Income Enhancement

In the model sketched above, there are many ways for poverty to affect child outcomes. Even with identical preferences, poorer parents will make different investment choices than richer ones. In particular, poor families will optimize at lower investment (and consumption) levels and thereby have children with lower health and human capital, other things equal. Further, parents may find input prices higher prices for certain goods. Poorer parents may also have access to different production technologies, that is, they may be less able to produce good outcomes given the same inputs, if for example, they are less educated. For example, lack of income could lead to stress and conflict among family members, which could impair children's development (Yeung et al., 2002). Providing cash transfers addresses the budgetary problems without necessarily changing the production technology. Hence, it is of interest to see whether cash transfers, in and of themselves, can improve outcomes. It is however, remarkably difficult to find examples of policies that increase incomes without potentially having a direct effect on outcomes. For example, many studies of cash welfare programs have demonstrated that children who are or have been on welfare almost always remain worse off than other children. This does not necessarily mean, however, that welfare has failed them. Without welfare, their situation might have been even worse. Berger, Paxson, and Waldfogel (in press) explore the relationship between family income, home environments, child mental health outcomes, cognitive test scores using data from the Fragile Families and Child Well-being Study which follows a cohort of five thousand children born in several large U.S. cities between 1998 and 2000. They show that all of the measures of the home environment they examine (which include measures of parenting skills as well as physical aspects of the home) are highly related to income and that controlling for these measures reduces the effects of income on outcomes considerably. Levine and Zimmerman (2000) showed that children who spent time on welfare scored lower than other children on a range of tests, but that this difference disappeared when the test scores of their mothers were controlled for, suggesting that welfare had little effect either positive or negative. Similarly, Zimmerman and Levine (1996) argue that children of welfare mothers were more likely to grow up to be welfare mothers, mainly because of other characteristics of the household they grew up in. Currie and Cole (1993) compare siblings in families where the mother received welfare while one child was *in utero*, but not while they other child was in utero, and find no difference in the birth weight of the siblings. Given that research has shown little evidence of positive effects of cash welfare on children, it is not surprising that the literature evaluating welfare reform in the United States has produced similarly null findings. The National Research Council (Smolensky and Appleton 2003) concluded that "no strong trends have emerged, either negative or positive, in indicators of parent well-being or child development across the years just preceding and following the implementation of [welfare reform]." However, U.S. welfare reform was a complex intervention that changed many parameters of daily life by, for example, imposing work requirements on recipients.

Conditional tax credits represent an alternative approach to providing income to poor families, and hence to poor children. The early years of the Clinton administration in the United States saw a huge expansion of the Earned Income Tax Credit (EITC), while in the U.K., the Working Families Tax Credit approximately doubled in 1999. These are tax credits available to poor working families. Their essential feature is that they are "refundable"—in other words, a family whose credit exceeds its taxes receives the difference in cash. The tax credits are like welfare in that they give cash payments to poor families. But life welfare reform, the tax credits are a complex intervention in that recipients need to work and file tax returns in order to be eligible, and a great deal of research has shown that such tax credits affect maternal labor supply and marriage patterns (Eissa and Leibman 1996; Meyer and Rosenbaum 2001, Blundell, 2006). This is because the size of the payment increases with earnings up to a maximum level before being phased out, so that it creates an incentive to work among the poorest households but a work disincentive for households in the phase-out range. In the U.S., the number of recipients grew from 12.5 million families in 1990 to 19.8 million in 2003, and the maximum credit grew from \$953 to \$4,204. The rapid expansion of this formerly obscure program run through the tax system has resulted in cash transfers to low-income families that were much larger than those that were available under welfare. Gunderson and Ziliak (2004) estimate that the EITC accounted for half of the reduction in after-tax poverty that occurred over the 1990s (the other half being mainly accounted for by strong economic growth).

Table 4 provides an overview of some of the research on the effects of income on children. Dahl and Lochner (2008) use variation in the amount of the EITC households are eligible for over time and household type to identify the effects of household income and find that each \$1,000 of income improves childrens' test scores by 2 to 4 percent of a standard deviation. An attractive feature of the changes in the EITC is that households may well have regarded them as permanent, so this experiment may approximate the effects of changes in permanent rather than transitory income. Their result implies, though, that it would take on the order of a \$10,000 transfer to having an educationally meaningful effect on test scores.

Milligan and Stabile (2008) take advantage of a natural experiment resulting from changes in Canadian child benefits. These benefits vary across provinces and were reformed at different times. An advantage of their research is that the changes in income were not tied to other changes in family behavior, in contrast to programs like the EITC. They find that an extra \$1,000 of child benefits leads to an increase of about 0.07 of a standard deviation in the math scores and in the Peabody Picture Vocabulary Test, a standardized test of language ability for four to six year old children. If we think of a change of a third or a half a standard deviation in test scores as a meaningful educational effect, then these results suggest that an increase of as little as \$5000 in family income has a meaningful effect. Milligan and Stabile (2008) go beyond Dahl and Lochner by examining effects on other indicators. They find that higher child benefits lower aggression in children and decrease depression scores for mothers. They do not find much impact on physical health measures, though they do find a decrease in families reporting that their children went hungry. There is some evidence of gender differences, with girls showing greater responsiveness to income on the mental health and behavioral scores while boys show greater responsiveness on test scores. These findings are extremely intriguing, but raise several questions. First, do the effects of income vary depending on the child's age? Smith et al. (1997) argue that income is more important at younger ages, though persistent poverty is worst of all. Second, are there really gender effects in the impact of income, and if so why? Third, the effects that Milligan and Stabile find are roughly twice those find by Dahl and Lochner. Is this because the former study a pure income transfer while the later study a tied transfer? Fourth, will the effects last, or will they be subject to “fade out” as the children grow older?

Table 4 also includes examples from growing literature analyzing ”conditional cash transfer programs” (CCTs). These are programs that tie transfers to specific behavior on the part of the family. For example, the parents may be required to make sure that the children attend school or get medical care in return for the transfer. These programs have become increasingly popular in developing countries, and have also been implemented to a limited extent in rich countries (for example, there is a program in New York City which is being evaluated by MDRC). By their nature, CCTs are complex programs that cannot tell us about the pure impacts of income. Still, these programs have attracted attention because randomized controlled trials have shown at least short-term results. It is difficult however to compare across programs, given that they all tend to focus on different outcomes.

Given this positive evidence about the effects of income, it is a puzzle why so much aid to poor families is transferred in kind. Currie and Gahvari (2008) survey the many reasons for this phenomena that have been offered in the Economics literature and conclude that the most likely reasons aid is offered in kind are agency problems, paternalism, and politics. In a nutshell, policy makers and the voters they represent may be more concerned with ensuring that children have medical care than with maximizing their parents utility, even if the parent’s utility is assumed to be affected by the children’s access to health care. Politics come in because coherent lobby groups (such as doctors, teachers, or farmers) may have incentives to advocate for various types of in kind programs. In any case, in kind programs are an important feature of aid policies in all Organization for Economic Cooperation and Development states, accounting for over 10% of GDP if health care and educational programs are included. In what follows, we first discuss “near cash programs” and then programs whose benefits are less fungible with cash.

5.2 Near-Cash Programs

Programs such as the U.S. Food Stamp Program (FSP, now renamed the Supplemental Nutrition Assistance Program, or SNAP) and housing assistance are often referred to as "near cash" programs because they typically offer households benefits that are worth less than what the household would have spent on food or housing in any case. Hence, canonical microeconomic theory suggests that households should think of them as equivalent to cash and that they should have the same impact as the equivalent cash transfer would have. In the case of food stamps, it has proven difficult to test this prediction because the program parameters are set largely at the national level, so that there is only time series variation. Currie (2003) provides an overview of the program, and the research on its effects that had been conducted up to that point. Schanzenbach (forthcoming) uses data from a food stamp cash out experiment to examine the effect on food spending. She finds that a minority of households actually received more in food stamps than they would otherwise spent on food. In these constrained households, families did spend more on food than they would have otherwise, while in other households, food stamps had the same effect as cash. Unfortunately, there is little evidence that constrained households bought foods that were likely to have beneficial effects; they seem, for example, to have spent some of the "extra" food money on products such as soda.

Hoynes and Schanzenbach (2009) use variation from the introduction of the FSP to identify its effects on food spending. The FSP began as a small pilot program in 1961, and gradually expanded over the next 13 years: In 1971, national eligibility standards were established, and all states were required to inform eligible households about the program. In 1974, states were required to extend the program statewide if any areas of the state participated. Using data from the PSID, the introduction of the FSP was associated with an 18%

increase in food expenditures in the full sample, with somewhat larger effects in the most disadvantaged households. They find that the marginal propensity to consume (MPC) food out of food stamp income was .16 compared to .09 for cash income. Thus, it does seem that many households were constrained to spend more on food than they otherwise would have (or alternatively, that the person receiving the food stamps had a stronger preference for food than the person controlling cash income in the household). From a policy makers point of view, this means that the FSP has a bigger impact on food spending than an equivalent cash transfer. Still, it is a leaky bucket as only 16 cents on every dollar transferred goes to food.

Bingley and Walker (2009) conduct an investigation of the Welfare Milk Program in the U.K.. They identify the effect of the program on household milk expenditures using a large change in eligibility for the program that had differential effects by household type. They find that about 80% of a transfer of free milk is crowded out by reductions in milk purchases by the household. This estimate is quite similar to that of Hoynes and Schanzenbach, though it still suggests that the in kind transfer is having some effect on the composition of spending. Details of these two studies are shown in Table 7.

Given that these programs appear to have some effect on food expenditures, it is reasonable to ask what effect they have on child outcomes. There is a substantial older literature examining this question (see Currie, 2003 for a summary). The modal study compares eligible participants to eligible non-participants using a multiple regression model. The main problem with drawing inferences about the efficacy of the FSP from this exercise is that participants are likely to differ from eligible non-participants in ways that are not observed by the researcher. Thus, for example, Basiotis et al. (1998) and Butler et al. (1996) both find that participation in the FSP reduces consumption of some important nutrients. Since it is hard to imagine how giving people food coupons could do this, one suspects that these results are driven by negative selection into the FSP program.

Several recent papers examining the effects of the FSP on young children are summarized in Table 7. Currie and Moretti (2008) were the first to try to use variation in the timing of the introduction of the food stamp program to look at effects on birth outcomes. Using Vital Statistics Natality data from California, they find that the introduction of the FSP increased the number of births, particularly in Los Angeles County. They also find some evidence that the FSP increased the probability of fetal survival among the lightest white infants, but the effect is very small, and only detectable in Los Angeles (L.A.). Notably, the FSP increased (rather than decreased) the probability of low birth weight but the estimated effect is small, and concentrated among teenagers giving birth for the first time. Thus, it appears that in California, the FSP increased fertility and infant survival (in some groups) with overall zero or negative effects on the distribution of birth weight.

Almond et al. (forthcoming) examine the same question using national data, and focus on receipt of the FSP during the third trimester, when the fetus typically puts on most of the weight the baby will have at birth. In contrast to Currie and Moretti, they find that the introduction of the FSP increased birth weights for whites and had even larger effects of blacks. The percentage reductions in the incidence of low birth weight were greater than the percentage increases in mean birth weight, suggesting that the FSP had its largest effects at the bottom of the birth weight distribution. Almond et al. find no effect of food stamp receipt in the first trimester of pregnancy and much weaker evidence for effects of receipt in the second. This suggests that one reason for the contrast between their results and those of Currie and Moretti is that the latter did not focus narrowly enough on the relevant part of pregnancy. Moreover, Almond et al. find larger effects in the South than in other regions, raising the possibility that overall effects were smaller in California than in other regions. Finally, it is possible that the effects in California are obscured by the substantial in-migration that the state experienced over this period.

Baum (2008) examines the effects of the FSP on weight gain among pregnant women, with particular attention to whether women gained either less than the recommended amount or greater than the recommended amount given their pre-pregnancy body mass index. He estimates a simultaneous equations model in which weight gain and FSP participation are jointly determined. FSP participation is assumed to be affected by various state-level rules about eligibility, outreach and so on. One difficulty is that these rules may be affected by other characteristics of states (such as overall generosity of social programs) which have direct effects on weight gain (e.g. through superior access to health care during pregnancy). Baum finds that FSP participation reduces the probability that women experienced inadequate weight gain during pregnancy, but has no effect on the probability that they gained too much weight. Since inadequate maternal weight gain is an important risk factor for low birth weight, it is likely that FSP had a positive effect on birth weights among affected mothers.

As discussed above, the other large category of "near cash" programs encompasses programs that offer subsidized housing. Many OECD countries have large housing assistance programs, but their effects on families are understudied. In fact, we were able to find only one paper that examined the effects of housing programs on the outcomes of children less than five, and only a handful that examined effects on children at all. Hence, we have no summary table regarding the effects of public housing programs on young children, but offer the following description of what has been done in this area.

Since by design, families receiving housing assistance are among the poorest of the poor, it is clearly important to address the endogeneity of program receipt. Currie and Yelowitz (2000) look at the effects of living in a public housing project in families with two children. They combine information from the Census and from the Survey of Income and Program Participation in a two-sample IV framework where the instrument for receipt of housing assistance is the sex composition of the siblings (families with a boy and a girl are entitled to larger apartments, and so are more likely to take up housing benefits). They find that families living in projects are less likely to be subject to overcrowding and that the children are much less likely to have been held back in school. The latter effect is three times bigger for boys (who are more likely to be held back in any case) than for girls. Since most "holding back" occurs at younger ages (Kindergarten and grade 1), this suggests that this type of assistance is in fact beneficial for young children. Goux and Maurin (2005) focus on the effect of overcrowding in France using a similar instrumental variables strategy: They argue that children in families in which the two eldest children are the same sex are more likely to live in crowded conditions in childhood. They also propose an alternative strategy in which crowding is instrumented with whether or not the parent was born in an urban area – parents who are from urban areas are more likely to live in crowded conditions. They find evidence consistent with Currie and Yelowitz in that crowding has a large and significant effect on the probability that a child falls behind in school and eventually drops out. Fertig and Reingold (2007) examine the effect of receipt of public housing assistance using data from the Fragile Families Study and three instruments: The gender composition of children in the household, the supply of public housing in each location, and the length of waiting lists in each location. They find improvements in maternal health and also in maternal reports of child health at age 3. Newman and Harkness (2002) use data from the PSID to examine the effect of and find that living in public housing as a child on future earnings and employment. Living in public housing is instrumented using the residual from a regression of local housing supply on the demographic characteristics of the area. They find that public housing is associated with increases in the probability of any employment (from 88 percent to about 95%) and increases in annual earnings (by $1,861 from a mean of $11,210). While all of these IV strategies are subject to caveats (is gender composition really uncorrelated with sibling's outcomes? Are characteristics of local housing markets associated with unobserved factors such as the quality of schools that might also affect child outcomes?) they certainly all point in a similar direction.

An important question is whether public housing assistance benefits children more than the equivalent cash transfer. It is difficult to answer this question given the available data. However, it is possible to eliminate some possible channels through which public housing programs might have different effects. One is that public housing programs may constrain the recipients choice of neighborhoods, with either positive or negative effects. Jacobs (2004) studies students displaced by demolitions of the most notorious Chicago high-rise projects. The U.S. Congress passed a law in 1996 that required local housing authorities to destroy units if the cost of renovating and maintaining them was greater than the cost of providing a voucher for 20 years. Jacobs argues that the order in which doomed buildings were destroyed was approximately random. For example, in January 1999, the pipes froze in some buildings in the Robert Taylor Homes, which meant that those buildings were demolished before others in the same complex. By comparing children who stayed in buildings scheduled to be demolished to others who had already been displaced by demolitions, he obtains a measure of the effect of living in high-rise public housing. Despite the fact that the high rises in Jacob's study were among the most notorious public housing projects in the country, he finds very little effect of relocation on children's educational outcomes. However, this may be because for the most part, children stayed in the same neighborhoods and in the same schools.

The most exhaustive examination of the effects of giving vouchers to project residents is an ongoing experiment called "Moving to Opportunity" (MTO). MTO was inspired by the Gautreaux program in Chicago, which resulted from a consent decree designed to desegregate Chicago's public housing by relocating some black inner-city residents to white suburbs. MTO is a large-scale social experiment that is being conducted in Chicago, New York, Los Angeles, Boston and Baltimore (see Orr et al. 2003). Between 1994 and 1998, volunteers from public housing projects were assigned by lottery to one of three groups. The first group received a voucher that could only be used to rent housing in a low-poverty area (a Census tract with a poverty rate less than 10 percent). This group also received help locating a suitable apartment (referred to here as the "MTO group"). The second group received a voucher which they could use to rent an apartment in any neighborhood. The third group was the control and received no vouchers or assistance although they were eligible to remain in their project apartment. Families in the first group did move to lower poverty neighborhoods and the new neighborhoods of the MTO group were also considerably safer. However, contrary to expectations, the move to new neighborhoods had positive effects on the mental health and schooling attainment of girls, and negative effects on the probability that they were ever arrested. But MTO either had no effect, or negative effects, on boys. Boys in the experimental group were 13 percent more likely than controls to have ever been arrested. This increase was due largely to increases in property crimes. These boys also report more risky behaviors such as drug and alcohol use. And boys in the MTO and voucher groups were more likely to suffer injuries. These differences between boys and girls are apparent even within families (Orr et al., 2003).

It remains to be seen how the long-term outcomes of the MTO children will differ from controls. Oreopoulos (2003) uses data from Canadian income tax records to examine the earnings of adults who lived in public housing projects in Toronto as children. There are large differences between projects in Toronto, both in terms of the density of the projects, and in terms of the poverty of the neighborhoods. Oreopoulos argues that the type of project a family lives in is approximately randomly assigned because the family is offered whatever happens to be available when they get to the top of the waiting list. Oreopoulos finds that once the characteristics of the family are controlled, the neighborhood has no effect on future earnings or on the likelihood that someone works.

We summarize the findings on near cash program effects as follows. There is credible evidence that the FSP may improve birth weight. More work remains to be done to determine whether it has positive effects on the nutrition of children after birth, whether similar programs in other countries have positive effects, and whether this particular type of in kind program has effects that are different than cash subsidies to poor households. The evidence regarding housing programs also suggests that they can be beneficial to families, but offers little guidance about the important question of whether housing programs matter primarily because they subsidize family incomes or operate through some other mechanism. It seems doubtful, given the available evidence, that housing programs benefit child outcomes primarily by improving their neighborhoods (especially since many housing projects are located in less desirable neighborhoods). It is conceivable that at least in some cases, housing assistance causes parents to allocate a greater share of their budgets to housing than they would otherwise, and that this has beneficial effects on children, but this is merely a conjecture.

5.3 Early Intervention Programs

Many programs specifically seek to intervene in the lives of poor children in order to improve their outcomes. Three interventions that have been shown to be effective are nurse home visiting programs, nutritional supplementation for pregnant women, and quality early childhood education programs. Table 9 summarizes some recent evidence about home visiting programs.

5.3.1 Home Visiting

Unlike many social programs, home visiting has been subject to numerous evaluations using randomized control trials. A recent survey appears in Howard and Brooks-Gunn (2009). David Olds and collaborators have developed a particular model for home visiting and conducted randomized controlled trials in a number of settings (Elmira, New York; Memphis, Tennessee; and Denver, Colorado) to evaluate it. Olds' programs focus on families that are at risk because the mother is young, poor, uneducatedand/or unmarried, and involve home visits by trained public health nurses from the prenatal period up to two years post-partum. The evaluations have shown many positive effects on maternal behavior, and on child outcomes. As of two years of age, children in Elmira were much less likely to have been seen in a hospital emergency room for unintentional injuries or ingestion of poisonous substances, although this finding was not replicated at other study sites. As of age 15, children of visited mothers were less likely to have been arrested or to have run away from home, had fewer sexual partners, and smoked and drank less. The children were also less likely to have been involved in verified incidents of child maltreatment. This finding is important given the high incidence of maltreatment among U.S. children (and especially among poor children), and the negative outcomes of maltreated children discussed above. There was little evidence of effects on cognition at four years of age (except among children of initially heavy smokers), though one might expect the documented reduction in delinquent behavior among the teens to be associated with improvements in eventual schooling attainment. Olds' model views using nurses as home visitors is key to the acceptability of the visitors (families want medical services, but may be suspicious of social workers or community workers). A randomized trial of nurses versus trained paraprofessionals (Olds, Robinson, and O'Brien, 2002) suggests that the effects that can be obtained by paraprofessionals are smaller. Also, the Olds programs are strongly targeted at families considered to be at risk and so they do not shed light on the cost-effectiveness of universal home visiting programs for pregnant women and/or newborns that exist in many countries.

Olds' positive results do not imply that all home visiting programs are likely to be equally effective. In fact, Table 9 suggests that the average home visiting program has relatively small effects. They often improve parenting in subtle ways and may result in some improvements in specific health outcomes. However, these may not be sufficient to justify the cost of a large scale program (Aos (2004) offers a cost benefit analysis of several programs). Home visiting programs can be viewed as a type of parenting program—presumably the reason why Olds home visitors improved outcomes is because they taught mothers to be better parents. Since parents are so important to children, programs that seek to improve parenting practices are perennially popular. Yet studies of these programs suggest that it is remarkably difficult to change parent's behavior and that many attempted interventions are unsuccessful. The most successful parenting programs are those that combine parent education with some other intervention that parents want, such as visits by nurses (as in Olds case) or child care (Brooks-Gunn and Markham, 2005).

5.3.2 U.S. Supplemental Feeding Program for Women, Infants, and Children (WIC)

A second type of early intervention program that has been extensively studied is the U.S. Supplemental Feeding Program for Women, Infants, and Children (WIC). As its name implies, WIC is a program targeted at pregnant and lactating women, infants, and children up to age 5. Participants receive vouchers that can be redeemed for particular types of food at participating retailers. Participants must generally go to the WIC office to receive the vouchers, and generally receive nutrition education services at that time. Many WIC offices are run out of clinics and may also facilitate access to medical care. Dozens of studies, (many of them reviewed in Currie (2003)) have shown that participation in WIC during pregnancy is associated with longer gestations, higher birth weights, and generally healthier infants, and that the effects tend to be largest for children born to the most disadvantaged mothers. Economists have critiqued these studies, on the grounds that there may be unobservable variables that are correlated with WIC participation among eligibles and also with better birth outcomes. Moreover, it may be implausible to expect WIC to have an effect on preterm birth. A recent Institute of Medicine report on the subject reviewed the evidence and concluded that randomized trials of many different interventions with women at risk of preterm birth had failed to find effects (Behrman, 2006). So it might be surprising to find an effect for WIC, when more specific and intensive interventions aimed at preventing preterm birth have generally failed. A number of new studies have attempted to deal with various aspects of this critique, as shown in Table 10. Bitler and Currie (2005) look at data from the Pregnancy Risk Monitoring System, which contains very detailed data from new mothers obtained by combining data from birth records and survey data taken from women before and after pregnancy. They directly address the question of selection bias by examining the population of mothers eligible for Medicaid (all of whom are adjunctively eligible for WIC) and asking how participants differ from non-participants along observable dimensions. They find that the WIC women are more disadvantaged than the non-participants along all observables. This finding does not prove that WIC women are also negatively selected in terms of unobservable variables, but it does mean that women who were very negatively selected in terms of education, health, family relationships and so on would have to have other attributes that were systematically correlated with positive outcomes. Like previous studies, Bitler and Currie also find that WIC participation is associated with higher maternal weight gain, longer gestation, and higher birth weight, particularly among women on public assistance, high school dropouts, teen mothers, and single mothers.

Joyce, Gibson, and Colman (2004) adopt a similar strategy with regard to selection, and focus on a sample of first births to women who initiated prenatal care in the first four months of pregnancy in order to ensure that participants and non-participants were more likely to be similar in terms of unobservables. In their sample of women giving birth in New York City, they find positive effects of WIC among U.S. born black women, but not in other groups. Joyce, Yunzal-Butler and Racine (2008) use a national sample of women, compare women who enrolled in WIC pre and post delivery, and focus on whether infants are small for gestational age (SGA). If one does not believe that WIC can affect gestation, then focusing on SGA is appropriate because it is not affected by gestational age. The find the incidence of SGA is lower for the prenatal enrollees than for the post-partum enrollees. Guerguieva, Morse, and Roth (2009) use a large sample of births from Florida and try to deal with potential selection using propensity score matching. They side step the issue of whether WIC affects gestation by presenting separate analyses for pregnancies of different length, and focusing on SGA. They find that longer participation in WIC is associated with reductions in the incidence of SGA. Kowaleski-Jones and Duncan examine sibling pairs from the NLSY and find that WIC participation is associated with an increase of seven ounces in birth weight. However, the number of pairs in which one child participated and one did not is quite small, so that it would be useful to try to replicate this finding in a larger sample of siblings.

Figlio et al. (2009) present an innovative instrumental variables strategy using a large sample of births from Florida that have been merged to school records of older siblings. While the characteristics of WIC programs vary across states, they do not show a lot of variation over time, and previous analyses have demonstrated that these characteristics are weak instruments (Bitler and Currie). Figlio et al. first try to select participant and non-participant groups who are very similar. They do this by defining "marginally ineligible" families as those who participated in the National School Lunch Program (NLSP) in the year before or after the birth, but did not participate in the birth year. Thus, the study focuses on families whose incomes hover around the eligibility threshold for NSLP, which is the same as the eligibility threshold for WIC. The instrument is a change in income reporting requirements for WIC in Sept. 1999 which made it more difficult for eligible families to receive benefits. Figlio et al. find that WIC participation reduces the probability of low birth weight, but find no significant effect on gestational age or prematurity.

There has been much less study of the effects of WIC on other outcomes, or other groups of participants. A couple of studies that make some attempt to deal with the selection issue are summarized in Table 10. One problem with WIC is that it subsidizes baby formula, which is likely to discourage breast-feeding. Chatterji et al. use the NLSY Mother-Child file and estimate both sibling fixed effects models and instrumental variables models using characteristics of state programs as instruments. They find that WIC reduces breast feeding initiation and the length of breastfeeding. However, these results are subject to the caveats above (i.e. small samples and possibly weak instruments). Turning to the effects of WIC on older children, Black et al. (2004) compare WIC eligible participants and those who did not participate due to "access problems." These problems were assessed based on the families own reports about why they were not participating. They found that infants who received WIC were less likely to be underweight, short, or perceived by their parents to be in fair or poor health. Lee and Mackay-Bilaver (2007) use a large data base from Illinois that integrates administrative data from several sources. Using sibling fixed effects models, they find that siblings who received WIC were less likely to be anemic, to have exhibited failure-to-thrive, or other nutritional deficiencies, and that the infants were less likely to be abused or neglected. As discussed above, one issue in the interpretation of these findings is why one infant would receive WIC while the other did not?

In summary, the latest group of studies of WIC during pregnancy largely support the findings of earlier studies which consistently found beneficial effects on infant health. The finding is remarkable because WIC benefits are relatively modest (often amounting to about $40 per month) and Americans are generally well fed (if not overfed at least in terms of total calories). Research that attempted to peer into the "black box" and shed light on why the program is effective would be extremely interesting. Another question that cries out for future research is whether WIC benefits infants and children (i.e. children who participate after birth)? While a few studies suggest that it does, the effects of WIC in this population has been subject to much less scrutiny than the effects on newborns.

5.3.3 Child Care

There have been many evaluations of early intervention programs delivered through the provision of child care. One reason for focusing on early intervention through the provision of quality child care is that the majority of young children are likely to be placed in some form of care. In 2006, 64% of women with children under 6 worked for pay (U.S. Bureau of Labor Statistics, 2008). While the U.S. may be an outlier in this respect, labor force participation among women with children is high and rising in many other economies. Currie (2001, 2006, 2009), Karoly et al. (1998) and Barnett (1995) all provide overviews of the literature on early intervention through child care. Many studies concern experimental evaluations of model programs that serve relatively small numbers of children and involve intensive services delivered by well-trained and well-supervised staff. These studies generally find that early intervention has long-lasting effects on schooling attainment and other outcomes such as teen pregnancy and crime, even if it does not result in any lasting increase in cognitive test scores. These results point to the tremendous importance of "non-cognitive skills" (c.f. Heckman and Rubinstein, 2001) or alternatively, to the importance of mental as well as physical health in the production of good child outcomes (Currie and Stabile, 2006). A few of these model programs are summarized in Table 11. Two studies of "model" early intervention child care programs stand out because they randomly assigned children to treatment and control group, had low dropout rates, and followed children over many years. They are the Carolina Abecedarian Project and the Perry Preschool Project. Both found positive effects on schooling. A recent cost-benefit analysis of the Abcedarian data through age 21 found that each dollar spent on Abecedarian saved tax payers four dollars. And by focusing only on cost savings, this calculation does not even include the value of higher achievement to the individual children and society (Masse and Barnett, 2002). Each dollar spent on Perry Preschool has been estimated to have saved up to seven dollars in social costs (Karoly et al., 1998), although this high benefit-cost ratio is driven largely by the effect of the intervention on crime, which in turn depends on a handful of individuals.

Anderson (2008) conducts a re-analysis of the Perry Preschool and Abcedarian data (and a third intervention called the Earkly Training Project) and finds that like MTO public housing experiment, the significant effects of the intervention was largely concentrated among girls. In addition to analysing the data by gender, Anderson pays careful attention to the idea that there may be a reporting bias in the published studies of early intervention experiments; that is, researchers who found largely null effects of the experiment might still be able to publish results focusing on one or two positive outcomes out of many outcomes investigated. Conversely, if all effects tended in the same direction, but there was insufficient power to detect significant effects on each outcome, it might be possible to detect a significant effect on an index of the outcomes. Anderson finds positive effects (for girls) on a summary index of effects, and the effects are quite large at about a half a standard deviation. This study highlights an interesting question which is whether it is generally easier to intervene with girls than with boys, and why that might be the case?

The fact that special interventions like Perry Preschool or Abcedarian had an effect on at least some target children, does not prove that the types of programs typically available to poor inner-city children will do so. Head Start is a preschool program for disadvantaged 3, 4, and 5 year olds which currently serves about 800,000 children each year. It is funded as a federal-local matching grant program and over time, federal funding has increased from $96 million when the program began in 1965 about $7 billion in 2009 (plus additional "stimulus" funds). Head Start is not of the same quality as the model interventions, and the quality varies from center to center. But

Head Start centers have historically been of higher average quality than other preschool programs available to low income people. This is because, in contrast to the private child care market, there are few very low-quality Head Start programs (see Blau and Currie, 2004 for an overview of preschool quality issues).

An experimental evaluation of Head Start is currently being conducted (U.S. Dept. of Health and Human Services, 2005). The evaluation compares Head Start children to peers who may or may not be in some other form of preschool (including state-funded preschools modelled in Head Start). In fact, the majority of children who did not attend Head Start did end up attending some other preschool program. Even relative to this baseline, initial results show that Head Start children make some gains, particularly in terms of language ability. But the first followup followed children only into the first grade, and so did not address the important issue of whether Head Start has longer term effects. A second follow up will assess children at the end of third grade. This example illustrates one of the limitations of experiments for the study of longer-term effects, which is that one may have to wait a long time for evidence to accumulate. There has also been a federal evaluation of Early Head Start (EHS), a version of the program geared to infants and toddlers under three years old. As Table 11 shows, EHS has small positive effects on cognitive test scores and some measures of behavior though Aos (2004) concludes that it does not pass a cost-benefit test.

Table 12 summarizes notable non-experimental evaluations of Head Start and other public preschool programs. In a series of studies with Duncan Thomas and other colleagues, I have used national publicly-available survey data to try to measure the effect of Head Start. In most of these studies, we compare the outcomes of children who attended Head Start to their own siblings who did not attend. The idea is that siblings share many common background characteristics. By choosing the child's own sibling as a control for the Head Start child, we effectively eliminate the effect of shared fixed family background characteristics on child outcomes. As discussed above, sibling fixed effects are not a panacea. However, careful examination of differences between participant and non-participant children within families suggested that the Head Start sibling typically attends when the family is relatively disadvantaged. For example, a young single mother might have her first child attend Head Start. If she then marries, her next child will enjoy higher income and be ineligible for Head Start. We found no within-family differences in birth weight or other individual characteristics of the children. We also investigated spillover effects, which as discussed above, can bias the estimated effect of Head Start. We found some evidence (Garces et al. 2002) that having an older sibling attend Head Start had positive effects on younger siblings. In all, it seems likely that sibling fixed effects models understate the true effect of Head Start.

Nevertheless, we have found significant positive effects of Head Start on educational attainments among white youths, and reductions in the probabilities of being booked or charged with a crime among black youths (Garces et al., 2002). Test score gains for blacks and whites were initially the same, but these gains tended to fade out more quickly for black than white students, perhaps because black former Head Start students typically attend worse schools than other students (Currie and Thomas, 1995). Effects were especially large for Hispanic children (Currie and Thomas, 1999). More recently, Deming (2008) replicates the results of Currie and Thomas (1995) using the same cohort of NLSY children observed at older ages. Like Anderson, he focuses on an index of outcomes (although he also reports results for separate outcomes) and finds that Head Start results in an increase of .23 standard deviations, which is equivalent to about 1/3 of the gap between Head Start and other children. He notes that projected gains in earnings are enough to offset the cost of the program, so that there is a positive cost/benefit ratio. Carniero and Ginja (2008) use the same data but a different identification strategy: they focus on families around the cutoff for income eligibility for the program and compare families who are just below (and therefore eligible) to those who are just above (and therefore ineligible). A potential problem with this strategy is that implicitly assumes that families cannot game the system by reducing their incomes in order to become eligible for the program. Consistent with other studies, they find positive effects of Head Start attendance on adolescents including reductions in behavior problems, grade repetition, depression, and obesity.

It should also be noted that since its inception, Head Start has aimed to improve a broad range of child outcomes (not just test scores). When the program was launched in 1965, the Office of Economic Opportunity assisted the 300 poorest counties in applying for Head Start funds, and these counties were significantly more likely than other counties to receive funds. Using a regression discontinuity design, Ludwig and Miller (2007) show that mortality from causes likely to be affected by Head Start fell among children 5 to 9 in the assisted counties relative to the others. Mortality did not fall in slightly older cohorts who would not have been affected by the introduction of the program.

Frisvold (2006) and Frisvold and Lumeng (2009) also focus on health effects by examining the effect of Head Start on obesity. The former instruments Head Start attendance using the number of Head Start places available in the community, while the later takes advantage of a cut in a Michigan Head Start program which resulted in the conversion of a number of full-day Head Start places to half day places. Both studies find large and significant effects on Head Start on the incidence of obesity. In defense of their estimates, which some might find implausibly large, Frisvold and Lumeng point out that a reduction of only 75 calories per day (i.e. less than a slice of bread or an apple) would be sufficient to yield their results. In small children, even small changes in diet may have large cumulative effects. Anderson, Foster, and Frisvold (2009) follow Garces et al. and use sibling fixed effects and data from the PSID to estimate the effect of Head Start on smoking as an adult. Again, they find large effects.

Head Start has served as a model for state preschools targeted to low-income children in states such as California, and also for new (non-compulsory) universal preschool programs in Georgia, and Oklahoma. The best available evaluations of universal preschool programs highlight the importance of providing a high quality program that is utilized by the neediest children. Baker, Gruber and Milligan (2005) examine the introduction of a universal, \$5 per day (later \$7), preschool program in the Canadian province of Quebec. The authors find a strong response to the subsidy in terms of maternal labor supply and likelihood of using care, but they find negative effects of children on a range of outcomes. Lefebvr, Merrigan, and Verstraete (2008) focus on the same natural experiment and examine the effects on children's vocabulary scores, which have been shown to be a good predictor of schooling attainment in early grades. They find strong evidence of negative effects. In interpreting this study, it is important to consider who was affected by the program. Because poor children were already eligible for child care subsidies, the marginal child affected by this program was a child who probably would have stayed home with his or her middle-class, married, mother, and instead was put into child care. Moreover, the marginal child care slot made available by the program was of low quality—the sudden influx of children into care caused the province to place more emphasis on making slots available than on regulating their quality. Hence, the study should be viewed as the consequence of moving middle class children from home care to relatively poor quality care. It is not possible to draw any conclusion from this study about the effect of drawing poor children into care of good quality, which is what the studies of model preschool programs and Head Start focus on.

Gormley and Gayer (2006) examine the effects of Oklahoma's universal pre-K program which is run through the public schools and is thought to be of high quality. They take advantage of strict age cutoffs for the program and compare children who had just attended for a year to similar children who were ineligible to attend because they were slightly younger. They find a 52 percent gain in pre-reading skills, a 27 percent gain in pre-writing skills, and a 21 percent gain in pre-math skills. These results suggest that a high quality universal pre-K program might well have positive effects, though one would have to track children longer to determine whether these initial gains translate into longer term gains in schooling attainment. Several other recent studies use a similar regression discontinuity design including Hustedt et al (2008) and Wong et al. (2008) who examine state pre-K programs in five states. These studies find uniformly positive effects. It has been argued in fact, that the effects of quality state preschool programs are larger than those of Head Start. However, it is difficult to control for pre-existing differences between the Head Start children and children who attend other preschools. For example in Magnuson, Ruhm, and Waldfogel (2007), the preschool children had systematically higher incomes than those who attended Head Start.

A handful of studies examine the long-term effects of public pre-school or Kindergarten programs. Cascio (2009) uses data from four decennial censuses to analyze the impact of introducing Kindergarten into public schools in the U.S., where Kindergarten was phased in on a state-by-state basis. Using a cohort-based design, she finds that white children born in adopting states after the reform were less likely to dropout of highschool and less likely to be institutionalized as adults. However, she finds no significant effect for blacks, which may be due to significant crowd out of blacks from other programs, such as Head Start. Like Anderson, she finds that the effects were larger for girls. Havnes and Mogstad (2009) study a 1975 policy change in Norway which increased the availability of regulated child care in some areas but not in others. They find that children ”exposed” to more child care received more education and were more likely to have earnings as adults. Once again, much of the benefit was concentrated among females, and children of less educated mothers were particularly likely to benefit. In terms of mechanisms, they find that the increase in formal care largely displaced informal care, without much net effect on the mother's labor force participation.

Finally, it is worth mentioning the "Sure Start" program in England and Wales. This program aimed to provide early intervention services in disadvantaged neighborhoods but allowed a wide variety of program models, which obviously complicates an assessment of the program. An evaluation was conducted by comparing communities that were early adopters to those that adopted later. A second evaluation compared Sure Start children to children from similar neighborhoods who were drawn from the Millenium Cohort study. This second study used propensity scores to balance the samples. The first evaluation found that the most disadvantaged households were actually doing more poorly in intervention areas than in other areas (NESS, 2005), while the second found some evidence of positive effects (NESS, 2008). Following the first evaluation, there has been a move to standardize the intervention and most communities are now offering Sure Start Children's Centers. This latest incarnation of the program remains to be evaluated.

This discussion shows the value of using a framework for the production of child quality as a lens for the interpretation of the program evaluation literature. As discussed above, child human capital is produced using inputs that may come from either the family or from other sources. A program that augments the resources available to the child is likely to have positive effects (subject of course to diminishing returns), while a program that reduces the resources available to the child is likely to have negative effects. Hence, a program that causes poor quality group time to be substituted for relatively high quality maternal time can have a negative effects, while a program that replaces high quality group time for relatively low quality maternal time may have positive ones.[16] An important point is that the literature does show

[16]The large literature about effects of maternal employment and of maternity leave policies is germane here. Two recent studies (Dustmann and Schonberg, 2009; Baker and Milligan, 2008) find that more generous leave policies decreases female labor supply but that they appear to have little impact on children. These results might be interpreted as evidence that replacing the average mother's time with child care of average quality has little impact.

that it is possible to intervene effectively and to improve the trajectories of young children. If parents are in fact reinforcing differences between children at home, then it may be even more important for the state to intervene.

5.4 Health Insurance

Health insurance is not an intervention program in the sense of the programs described above. Yet, there is a good deal of evidence that access to health insurance improves children's health at birth and afterwards. Much of the evidence comes from studies of the introduction, or expansion, of health insurance benefits. Some of this literature is summarized in Table 13. For example, Hanratty (1996) examined the introduction of public health insurance in Canada, which was phased in on a province-by-province basis. Using county-level panel data, she finds that the introduction of health insurance was associated with a decline of four percent in the infant mortality rate, and that the incidence of low birth weight also decreased by 1.3% for all parents and for 8.9% for single parents. Currie and Gruber (1996) conduct a similar exercise for the U.S., focusing on an expansion of public health insurance to pregnant women and infants. They find that the effects vary depending on whether the expansion covered the poorest women, or women somewhat higher in the income distribution. Narrowly targeted expansions that increased the fraction of the poorest women eligible by 30% reduced low birth weight by 7.8% and reduced infant mortality by 11.5%. Broader expansions of a similar magnitude had very small effects on the incidence of low birth weight, but reduced infant mortality. This result suggests that among women of somewhat higher income levels, the expansions did not improve health at birth, but may have increased access to life-saving technologies after birth. Currie and Grogger (2002) focus on bureaucratic obstacles to obtaining health insurance by looking at contractions of welfare (women cut from the rolls lost automatic eligibility for Medicaid) as well as outreach measures undertaken by different states. They find that changes that reduced barriers to enrollment increased use of prenatal care and had positive effects on infant health outcomes.

Baldwin et al. (1998) use individual-level data and compare expansions in Washington, which included enhanced prenatal care services, to expansions in Colorado which did not in a difference-in-differences design. They find reductions in low birth weight among medically high risk infants in Washington. Dubay et al. (2001) conduct a difference-in-difference investigation comparing the outcomes of high and low socioeconomic status women in the 1980-1986 period and in the 1986 to 1993 period. They find overall improvements in the use of prenatal care for low SES women, but find improvements in birth weight only for some groups of white women. However, this design does not really focus on health insurance *per se*, since the estimates will be affected by any other changes in health care markets between the two periods that had differential effects by SES.

Studies of the effects of health insurance expansions on children often examine preventable hospitalizations (also called ambulatory care sensitive hospitalizations). The idea is that certain conditions, such as childhood asthma, should not result in hospitalizations if they are properly treated on an outpatient basis. Hence, hospitalizations for these conditions are inefficient and indicate that children are receiving preventive inadequate care. Kaestner, Joyce, and Racine (2001) use a difference-in-differences design comparing low income and other children before and after Medicaid expansions. They find reductions in preventable hospitalizations of 10 to 20 percent.

Aizer (2003) examines a California outreach program that increased child enrollments into Medicaid. She finds that an increase in enrollments of 1,000 reduces hospitalizations by 3.26. Dafny and Gruber (2005) use a design similar to Currie and Gruber in which actual individual eligibility is instrumented using a "simulated eligibility measure" which is an index of the generosity of the Medicaid program in the state. The reason for adopting instrumental variables estimation is that eligibility for Medicaid is determined by endogenous variables such as parental labor supply. They find that Medicaid eligibility increased hospitalizations overall. However, there was no statistically significant increase in avoidable hospitalizations, suggesting that the increase was mainly due to children with unavoidable conditions gaining greater access to care. They also found increases in the probability of receiving a procedure and reductions in length of stay, suggesting that children who were hospitalized were receiving more aggressive care and that it may have improved their outcomes.

One difficulty with studying child health is that in the developmental framework of Section 2, health may be greatly influenced by investments including health insurance at younger ages, and their interaction with net investments in the current period. Currie, Decker, and Lin (2008) therefore compare the health effects of contemporanous eligibility for health insurance among older children to the effect of having been eligible since birth. They find that contemporaneous health insurance coverage has little effect but that eligibility from birth is protective. Levine and Schanzenbach (2009) link health insurance eligibility at birth to 4th and 8th grade scores on the National Assessment of Educational Progress. They find that a 50 percentage point increase in eligibility at birth is associated with a small but significant gain on reading scores at both grades, though there is no effect on math scores. A difficulty with both studies is that neither income at birth nor state of birth are directly observed in the cross sectional data sets that they use, so they must be imputed using current income and state of birth.

Another area of research focuses on the quality of care provided by public health insurance programs. Analysis of this issue is complicated by the possibility that expansions of public insurance cause people to lose private health insurance coverage, a phenomena dubbed "crowdout" (Cutler and Gruber, 1996; Dubay and Kenney, 1997; Card and Shore-Shepard, 2004; Gruber and Simon, 2008). If the private insurance that is lost (or dropped) in response to expansions of public insurance is of superior quality to the private insurance, than people's health may suffer. Koch (2009) concludes that recent expansions of public health insurance to children at higher income levels have reduced access to doctor's office visits and increased reliance on emergency rooms. He also shows some evidence consistent with the idea that this is because children are being crowded out of superior (but obviously more expensive) coverage from private heath insurance plans. In fact, it is quite possible that crowding out has increased over time as the public has become familiar with public health insurance plans for children and private health insurance costs have continued to escalate.

Medicaid managed care has also been shown (in at least some cases) to reduce the quality of care. Conover et al. (2001) conduct a difference-in-difference analysis of Tennessee and North Carolina before and after Tennessee switched its Medicaid patients to managed care. They find that use of prenatal care and birth outcomes deteriorated in Tennessee after the switch. Aizer, Currie and Moretti (2007) examine data from California, where Medicaid managed care was adopted on a county-by-county basis. They also find that compulsory managed care had a negative impact on use or prenatal care and birth outcomes. This may be because the California Medicaid managed care program "carved out" care for sick newborns–that is, they were covered by a state fund rather than by the managed care companies.

In summary, health insurance matters for children's outcomes. But quality of care also matters. And it is important to remember that for most children threats to health and well being come from sources such as injuries, poor nutrition, and toxins rather than from lack of access to conventional medical care.

6 Discussion and Conclusions

Try and compare impact estimates in sections 5 and 6 to the *R-squared* stats in T1 – to what extent are the impact estimates smaller and why?

Note, in contrast to generally gloomy connotations of first part of the paper which shows long term lasting damage from early shocks, intervention literature suggests that there really is a great deal of remediation that can be accomplished...

- Do shocks at certain key ages matter more than others?
  - From natural experiments, tend to learn more about the "near" age comparisons than the "far" age comparisons. I.E. flu pandemic helps say something about fetal versus early post-neonatal, but less about fetal versus age 5 exposure.
- To what extent do different types of shocks have different effects?
- Are there interactions between shocks to health and shocks to cognition?
- Do effects differ by gender (and why?)?
  - In nations or sub-populations with son preference, do early childhood investments differ by gender?
  - How have advances in fetal sex determination affected prenatal investments in the context of son preference? See Lhila and Simon [2008] for recent work on the investment pattern by gender following ultrasound diffusion. Should systematic investment responses be found, a relatively open question is how human capital outcomes will be affected by technological improvements in sex determination and sex selection.
- What is the least cost way to improve outcomes?
- What can the burgeoning literature on biomarkers contribute beyond prediction? Is that a growth area for economists interested in "early origins"?

- What is the scope for randomized interventions in understanding long-term effects? For example, is it feasible to extend previous randomizations of prenatal nutritional supplementation (e.g., Harlem study by Rush, Stein, and Susser [1980]) to evaluate cognitive outcomes in secondary school, and whether parental investments during childhood were affected by the randomization? Are certain new controlled studies ethical? (E.g. providing some pregnant women information on risks of infections)

- Changing definition of live births – what does inclusion of births that would formerly have been classified as fetal deaths (because of improved neonatal medicine, e.g. MacDorman, Martin, Mathews, Hoyert, and Ventura [2005]) imply for estimates of program effectiveness and cohort effects? For example, twinning rate goes up much faster for live births than for fetal deaths, suggesting that:

  - Increase in twinning attributable to assisted reproductive technologies (ART) is overstated (maybe grossly)

  - More low-weight babies who would formerly have been called fetal deaths now show up as live births

Do in kind programs have larger effects than cash? Until recently there was little evidence that cash had any effect, so this question was relatively easy to answer. Inefficiencies of in-kind vs. potential to reach needist kids?

Talk again about the importance of new data (or better access to old data?)

References

An overstretched hypothesis? *The Lancet*, 357(9254):405, 2001.

Anna Aizer and Laura Stroud. Education, medical knowledge and the evolution of disparities in health. manuscript, Brown University, September 2009.

Douglas Almond. Is the 1918 influenza pandemic over? long-term effects of in utero influenza exposure in the post-1940 U.S. population. *Journal of Political Economy*, 114(4):672–712, August 2006.

Douglas Almond, Kenneth Y. Chay, and David S. Lee. The costs of low birth weight. *The Quarterly Journal of Economics*, 120(3):1031–1084, August 2005.

Douglas Almond, Lena Edlund, and Mårten Palme. Chernobyl's subclinical legacy: Prenatal exposure to radioactive fallout and school outcomes in sweden. *The Quarterly Journal of Economics*, 2009. forthcoming.

Abhijit V. Banerjee, Esther Duflo, Gilles Postel-Vinay, and Timothy Watts. Long run health impacts of income shocks: Wine and phylloxera in 19th century france. *Review of Economics and Statistics*, 2009. forthcoming.

Spencer Banzhaf and Randall P. Walsh. Do people vote with their feet? an empirical test of tiebout. *American Economic Review*, 98(3):84363, June 2008.

D.J.P. Barker, editor. *Fetal and Infant Origins of Adult Disease*. British Medical Journal, London, 1992.

Jrg Baten, Dorothee Crayen, and Joachim Voth. Poor, hungry and stupid: Numeracy and the impact of high food prices in industrializing britain, 1780-1850. Economics Working Papers 1120, Department of Economics and Business, Universitat Pompeu Fabra, October 2007. URL `http://ideas.repec.org/p/upf/upfgen/1120.html`.

Jere R. Behrman, Robert A. Pollak, and Paul Taubman. Parental preferences and provision for progeny. *Journal of Political Economy*, 90(1):52–73, 1982.

Jere R. Behrman and Mark R. Rosenzweig. Returns to birthweight. *The Review of Economics and Statistics*, 86(2):586–601, May 2004.

Gerard J. Van Den Berg, Maarten Lindeboom, and France Portrait. Economic conditions early in life and individual mortality. *American Economic Review*, 96(1):290–302, March 2006.

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

Hoyt Bleakley. Disease amd development: Evidence from hookworm eradication in the south. *The Quarterly Journal of Economics*, 122(1):73–, 2007.

Carlos Bozzoli, Angus S. Deaton, and Climent Quintana-Domeque. Child mortality, income and adult height. *NBER Working Paper No. 12966*, March 2009. forthcoming, "Demography".

Anne Case, Darren Lubotsky, and Christina Paxson. Economic status and health in childhood: The origins of the gradient. *American Economic Review*, 92(5):1308–1334, December 2002.

Anne Case and Christina Paxson. Early Life Health and Cognitive Function in Old Age. *American Economic Review Papers and Proceedings*, 99(2): 104–109, May 2009.

Kenneth Y. Chay and Michael Greenstone. Air quality, infant mortality, and the clean air act of 1970. Working Paper 10053, National Bureau of Economic Research, October 2003a. URL http://www.nber.org/papers/w10053.

Kenneth Y. Chay and Michael Greenstone. The impact of air pollution on infant mortality: Evidence from the geographic variation in pollution shocks induced by a recession. *Quarterly Journal of Economics*, 118(3):1121–1167, August 2003b.

Kenneth Y. Chay, Jonathan Guryan, and Bhashkar Mazumder. Birth cohort and the black-white achievement gap: The roles of access and health soon after birth. Working Paper 15078, National Bureau of Economic Research, June 2009. URL http://www.nber.org/papers/w15078.

Alpha C. Chiang. *Fundamental Methods of Mathematical Economics*. McGraw-Hill, New York, third edition, 1984.

Dalton Conley and Neil G. Bennett. Birth weight and income: Interactions across generations. *Journal of Health and Social Behavior*, 42(4):450–465, 2001. ISSN 00221465. URL http://www.jstor.org/stable/3090189.

Dora L. Costa and Joanna N. Lahey. Predicting older age mortality trends. *Journal of the European Economic Association*, 3(2-3):487–493, April-May 2005.

Eileen M. Crimmins and Caleb E. Finch. Infection, inflammation, height, and longevity. *Proceedings of the National Academy of Sciences*, 103(2):498–503, January 2006.

Flavio Cunha and James J. Heckman. The technolgy of skill formation. *American Economic Review*, 97(2):31–47, May 2007.

Janet Currie and Rosemary Hyson. Is the impact of shocks cusioned by socioeconomic status? the case of low birth weight. *American Economic Review*, 89(2):245–250, May 1999.

Janet Currie and Enrico Moretti. Biology as destiny? short- and long-run determinants of intergenerational transmission of birth weight. *Journal of Labor Economics*, 25(2):231–264, 2007. URL http://www.journals.uchicago.edu/doi/abs/10.1086/511377.

Janet Currie and Matthew Neidell. Air pollution and infant health: What can we learn from california's recent experience? *Quarterly Journal of Economics*, 120(3):1003–1030, 2005. URL http://www.mitpressjournals.org/doi/abs/10.1162/003355305774268219.

Janet Currie, Matthew Neidell, and Johannes F. Schmieder. Air pollution and infant health: Lessons from new jersey. *Journal of Health Economics*, 28(3):688 – 703, 2009. ISSN 0167-6296. URL http://www.sciencedirect.com/science/article/B6V8K-4VPV58R-1/2/102b434aaf36b4f2b9997

Janet Currie and Mark Stabile. Socioeconomic status and child health: Why is the relationship stronger for older children? *American Economic Review*, 93(5):1813–1823, December 2003.

Janet Currie and Reed Walker. Traffic congestion and infant health: Evidence from e-zpass. Working Paper 15413, National Bureau of Economic Research, October 2009. URL http://www.nber.org/papers/w15413.

David M. Cutler, Grant Miller, and Douglas M. Norton. Evidence on early-life income and late-life health from America's Dust Bowl era. *Proceedings of the National Academy of Sciences*, 104(33):13244–13249, 2007. URL http://www.pnas.org/content/104/33/13244.abstract.

Ashlesha Datar, Rebecca Kilburn, and David Loughran. Endowments and parental investments in infancy and early childhood. *Demography*, february 2010. forthcoming.

Emilia Del Bono, John Ermisch, and Marco Francesconi. Intrafamily resource allocations: A dynamic model of birth weight. IZA Discussion Papers 3704, Institute for the Study of Labor (IZA), September 2008. URL `http://ideas.repec.org/p/iza/izadps/dp3704.html`.

Olivier Deschnes and Michael Greenstone. Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the us. *NBER Working Paper No. 13178*, June 2007.

Gabriele Doblhammer and James W. Vaupel. Lifespan depends on month of birth. *Proceedings of the National Academy of Sciences*, 98(5):2934–2939, February 2001.

Matthew C. Farrelly, Christian T. Nimsch, and Joshua James. State cigarette excise taxes: Implications for revenue and tax evasion. Report 08742.000, RTI International, Research Triangle Park, NC, May 2003.

Erica Field, Omar Robles, and Maximo Torero. Iodine deficiency and schooling attainment in tanzania. *The American Economic Journal - Applied*, 2009. forthcoming.

Peter D. Gluckman and Mark A. Hanson. Adult disease: Echoes of the past. *European Journal of Endocrinology*, 155(Supl. 1), 2006.

Zvi Griliches. Sibling models and data in economics: Beginnings of a survey. *Journal of Political Economy*, 87(s5):S37, 1979. URL `http://www.journals.uchicago.edu/doi/abs/10.1086/260822`.

Michael Grossman. On the concept of health capital and the demand for health. *Journal of Political Economy*, pages 223–255, 1972.

James J. Heckman. The economics, technology, and neuroscience of human capability formation. *PNAS*, 104(33):13250–13255, August 14 2007.

Fritz Heider. The influence of the epidemic of 1918 on deafness: A study of birth dates of pupils registered in schools for the deaf. *American Journal of Epidemiology*, 19:756–765, 1934.

Amy Hsin. Is biology destiny? birth weight and differential parental treatment. Manuscript, Population Studies Center, University of Michigan, 2009.

Elaine Kelly. The scourge of asian flu: in utero exposure to pandemic influenza and the development of a cohort of british children. *Institute for Fiscal Studies*, Working Paper 09/17, September 2009. University College London.

W.O. Kermack, A.G McKendrick, and P.L. McKinlay. Death-Rates in Great Britain and Sweden. *The Lancet*, pages 698–703, March 31 1934.

Aparna Lhila and Kosali I. Simon. Prenatal health investment decisions: Does the childs sex matter? *Demography*, 45(4):885–905, November 2008.

Diana S. Lien and William N. Evans. Estimating the Impact of Large Cigarette Tax Hikes: The Case of Maternal Smoking and Infant Birth Weight. *J. Human Resources*, XL(2):373–392, 2005. URL `http://jhr.uwpress.org/cgi/content/abstract/XL/2/373`.

MF MacDorman, JA Martin, TJ Mathews, DL Hoyert, and SJ Ventura. Explaining the 2001-2002 infant mortality increase in the united states: data from the linked birth/infant death data set. *International Journal of Health Services*, 35(3):415–442, 2005.

Peter Nilsson. The long-term effects of early childhood lead expsoure: Evidence from sharp changes in local air lead levels induced by the phase-out of leaded gasoline. manuscript, Uppsala University, September 2009.

Phil Oreopoulos, Mark Stabile, Randy Walld, and Leslie Roos. Short, medium, and long-term consequences of poor infant health: An anlysis using siblings and twins. *Journal of Human Resources*, 43(1), Winter 2008.

Kathleen Maher Rasmussen. The "fetal origins" hypothesis: Challenges and opportunities for maternal and child nutrition. *Annual Review of Nutrition*, 21:73–95, July 2001.

Jessica Wolpaw Reyes. Environmental policy as social policy? the impact of childhood lead exposure on crime. Working Paper 13097, National Bureau of Economic Research, May 2007. URL http://www.nber.org/papers/w13097.

Mark R. Rosenzweig and Kenneth I. Wolpin. Heterogeneity, intrafamily distribution, and child health. *Journal of Human Resources*, 23(4):437–461, Fall 1988.

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

David Rush, Zena Stein, and Mervyn Susser. *Diet in Pregnancy: A Randomized Controlled Trial of Nutritional Supplements*. Alan R. Liss, Inc., New York, 1980.

Sicco A. Scherjon, Hans Oosting, Bram W.Ongerboer de Visser, Ton de Wilded, Hans A. Zondervan, and Joke H. Kok. Fetal brain sparing is associated with accelerated shortening of visual evoked potential latencies during early infancy. *American Journal of Obstetrics and Gynecology*, 175 (6):1569–1575, December 1996.

Gary Solon. *Handbook of Labor Economics*, volume 3, chapter Intergenerational Mobility and the Labor Market, pages 1761–1800. Elsevier, 1999. edited by Orley Ashenfelter and David Card.

Zena Stein, Mervyn Susser, Gerhart Saenger, and Francis Marolla. *Famine and Human Development: The Dutch Hunger Winter of 1944-1945*. Oxford University Press, New York, 1975.

Emma Tominey. Maternal smoking during pregnancy and early child outcomes. CEP Discussion Papers dp0828, Centre for Economic Performance, LSE, October 2007. URL http://ideas.repec.org/p/cep/cepdps/dp0828.html.

Peter Zweifel, Friedrich H. J. Breyer, and Mathias Kifmann. *Health Economics*. Oxford University Press, second edition, 2009. forthcoming.

Appendix

Appendix A

In general, we need to observe the investments $\bar{I}_1$ and $\bar{I}_2$ to estimate parameters of the production function. However, if we lack measures of $\bar{I}_1$ and $\bar{I}_2$ but suspect them to be similar, for $\mu_g = {\mu_g}'$ (4) reduces to:

$$\frac{1-\gamma}{\gamma},$$

So comparing damage to h from the shock $\mu_g = \mu_g^{'}$ isolates γ (while remaining silent on the magnitude of ϕ).

To estimate ϕ, we can use exposure to a shock in both the first *and* second childhood periods. In the OLG framework, this would require a shock lasting two childhood periods (or longer), and "half-exposed" cohorts on either end of the shock. Damage to the fully exposed cohort relative to the cohort exposed in period 1 alone is:

$$1 + \frac{1-\gamma}{\gamma}(\bar{I}_2 + \mu_g^{'})^{\phi-1}. \tag{9}$$

Between (4) and (9) we now have two equations in the unknowns ϕ and γ. Thus:

$$\phi = 1 + \frac{\mathtt{log}(\frac{c_4-1}{c_3})}{\mathtt{log}(\bar{I}_1 + \mu_g)} \tag{10}$$

Appendix B

Human capital of a child is produced with a CES technology:

$$h = A\left[\gamma(\bar{I}_1 + \mu_g)^{\phi} + (1-\gamma)I_2^{\phi}\right]^{1/\phi}, \tag{11}$$

where μ_g is an exogenous shock to (predetermined) period 1 investments. Parents value their consumption and the human capital of their child:

$$U_p = U(C, h) = B\left[\theta(C)^{\varphi} + (1-\theta)h^{\varphi}\right]^{1/\varphi}, \tag{12}$$

and have the budget constraint:

$$\bar{I}_1 + I_2 + C = \bar{y}.$$

Absent discounting, the marginal utility from consuming equals the marginal utility from investing:

$$\frac{\delta U}{\delta C^*} = \frac{\delta U}{\delta h}\frac{\delta h}{\delta I_2^*}.$$

$$\theta C^{\varphi-1} = (1-\theta)h^{\varphi-1}A[...]^{\frac{1}{\phi}-1}(1-\gamma)I_2^{*\phi-1} \tag{13}$$

$$\theta(\bar{y} - \bar{I}_1 - I_2^*)^{\varphi-1} = (1-\theta)A^{\varphi-1}[...]^{\frac{\varphi-1}{\phi}}A[...]^{\frac{1}{\phi}-1}(1-\gamma)I_2^{*\phi-1} \tag{14}$$

$$\theta(\bar{y} - \bar{I}_1 - I_2^*)^{\varphi-1} = (1-\theta)(1-\gamma)A^{\varphi}\,[...]^{\frac{\varphi-\phi}{\phi}}\,I_2^{*\phi-1} \tag{15}$$

$$G(u_g, I_2^*) \equiv \theta(\bar{y} - \bar{I}_1 - I_2^*)^{\varphi-1} - (1-\theta)(1-\gamma)A^{\varphi}\,[...]^{\frac{\varphi-\phi}{\phi}}\,I_2^{*\phi-1} = 0. \tag{16}$$

$$\frac{\delta I_2^*}{\delta \mu_g} = -\frac{\frac{\delta G}{\delta \mu_g}}{\frac{\delta G}{\delta I_2^*}}$$

$$= \frac{a(I_2^*)^{\phi-1}[...]^{\frac{\varphi-2\phi}{\phi}}\left(\frac{\varphi-\phi}{\phi}\right)\gamma\phi(\bar{I}_1+\mu_g)^{\phi-1}}{-(\varphi-1)\theta(\bar{y}-\bar{I}_1-I_2^*)^{\varphi-2} - a\left[[...]^{\frac{\varphi-\phi}{\phi}}(\phi-1)I_2^{*\phi-2} + \frac{\varphi-\phi}{\phi}[...]^{\frac{\varphi-2\phi}{\phi}}\phi(1-\gamma)I_2^{*\phi-1}I_2^{*\phi-1}\right]}, \tag{17}$$

using the implicit function theorem and defining a to be $(1-\theta)(1-\gamma)A^{\varphi} \geq 0$.

$$= \frac{(\varphi-\phi)a(I_2^*)^{\phi-1}[...]^{\frac{\varphi-2\phi}{\phi}}\gamma(\bar{I}_1+\mu_g)^{\phi-1}}{(1-\varphi)\theta(\bar{y}-\bar{I}_1-I_2^*)^{\varphi-2} \;+\; a[...]^{\frac{\varphi-\phi}{\phi}}I_2^{*\phi-2}\left[(1-\phi)+(\varphi-\phi)(1-\gamma)I_2^{*\phi}/[...]\right]} \tag{18}$$

For $\varphi > \phi$, (18) is positive, so negative shocks in the first period should be reinforced. Accommodation through preferences (i.e., more consumption and less investment, which lowers h in addition to that caused by μ_g) is optimal.

Appendix C

TBA.

(*basic idea is that gamily j has 2 kids and their h_j^1 and h_j^2 are the two arguments in utility function.*)

Table 1: How much of the differences in later outcomes (test scores, educational attainment, earnings) can be explained by early childhood factors?

Study	Dataset	Independent Vars	Dependent Vars	R-squared
STUDIES USING NLSY-CHILD SAMPLE DATA				
Childhood Emotional and Behavioral Problems and Educational Attainment (McLeod and Kaiser (2004))	NLSY-Child Sample Data on children who were 6-8 years old in 1986 for five waves through 2000 (when they were 20-22 years old). n = 424	Emotional and behavioral problems at age 6-8 measured by BPI, mother emotional problems and delinquency, poverty status, mother's AFQT score, mother's education, mother's marital status and age, child's age, sex, and race, dummy for LBW	Dummy for graduating high school by 2000, dummy for enrolling in college by 2000	**For predicting HS graduation:** Only child emotional and behavioral problems: 0.046 Adding in child and mother demographics: 0.111 Adding in mother's emotional problems and delinquency: 0.124 **For predicting college enrollment:** Only child emotional and behavioral problems: 0.017 Adding in child and mother demographics: 0.093 Adding in mother's emotional problems and delinquency: 0.112
STUDIES USING NCDS DATA				
Ability, family, education, and earnings in Britain (Dearden (1998))	NCDS: 1958 cohort Focus on individuals who participated in waves 4 and 5 of the survey in 1981 and 1991, who were employees in 1991. n = 2597 males, 2362 females	Math and verbal ability (age 7), type of school, family characteristics -- teacher's assessment of interest shown by parents in child's education at 7; type of school attended at 16 family's financial status at 11 and 16; region dummies; father's SES; parents' education levels	Years of full-time education; Earnings at age 33	**Education as outcome:** Including all explanatory variables: R-squared = 0.33-0.34 **Earnings as outcome:** Baseline earnings equation includes only years of eduation R-squared = 0.15 for males, R-squared = 0.25 for females. Adding in reading and math ability at age 7, as well as school type and regional dummies raises R-squared to 0.26 for males, 0.31 for females. Including all explanatory variables raises R-squared to 0.29 for males, 0.41 for females.
Early test scores, socioeconomic status, and future outcomes (Currie and Thomas (1999))	NCDS: 1958 cohort Full sample size (based on responses at 7): n = 14,022 Individuals surveyed are all born in the week of March 3, 1958, and followed through	Reading and math test scores at age 7, mother and father's SES and education, birth weight, other child background variables at age 7	Number of O-level passes of exams by age 16; employed at age 23, 33; log wage at age 23, 33	**For predicting age 16 exam passes:** Reading and math test scores only: 0.21-0.22 With other background vars: 0.31-0.32 **For predicting employment at 33:**

	age 33. Sample sizes for each outcome adjusted based on number of responses.			Reading and math test scores only: 0.01 With other background vars: 0.04-0.05 **For predicting log wage at 33:** Reading and math test scores only: 0.08-0.09 With other background vars: 0.18-0.20
The lasting impact of childhood health and circumstance (Case, Fertig, and Paxson (2005))	NCDS: 1958 cohort n = 14,325 (7016 men, 7039 women) Individuals surveyed are all born in the week of March 3, 1958, and followed through age 42. Sample sizes for each outcome adjusted based on number of responses.	Mother's and father's education and SES, LBW, indicators for moderate, heavy, and varied maternal smoking during pregnancy, number of chronic conditions at age 7 and 16.	Number of O-level passes of exams by age 16; adult health status at age 42; part-time or full employment at age 42.	**For predicting age 16 exam passes:** Mother's education and SES: contributes 0.062 Father's education and SES: contributes 0.241 LBW and maternal smoking: contribute 0.024 **For predicting adult health:** Mother's education and SES: contributes 0.082 Father's education and SES: contributes: 0.189 LBW and maternal smoking: contribute 0.086 **For predicting employment:** Mother's education and SES: contributes: 0.076 Father's education and SES: contributes: 0.173 LBW and maternal smoking: contribute 0.052
Explaining intergenerational income persistence: non-cognitiv skills, ability, and education (Blanden, Gregg, and Macmillan (2006))	**A.** NCDS: 1958 cohort Individuals surveyed are all born in the week of March 3, 1958, and followed through age 42. Outcomes comparable to BCS study through age 33. Focus on males (sons). n = 2163 males **B.** British Cohort Study: 1970 cohort Individuals surveyed born between April 4th and 11th, 1970. Followed through 30.	**A.** Family income at age 16, reading and math test scores at age 11, scores for "behavioral syndromes" at age 11, O-level exam scores at age 16 **B.** Years of preschool education, birth weight, height at 5 and 10, emotional/behavioral scores at ages 5, 10, & 16, family income at ages 10 and 16, reading and math test scores at age 10, IQ at age 10, dummy for HS degree,	**A.** Earnings at age 33 **B.** Earnings at age 30	**A.** Birth weight, childhood health, and age 11 test scores only: 0.116 Including "behavioral syndromes" at age 11: 0.151 All variables: 0.263 **B.** Birth weight, childhood health, and age 10 test scores only: 0. 075 Including emotional/behavioral characteristics at age 10: 0.087 All variables: 0.222

	Focus on males (sons). n = 3340 males	exam scores at age 16.			

Table 2: SUMMARY OF RESULTS FROM ECLS DATA

Effects of Continuous Birth Weight in Kilograms on Outcomes: Mother Fixed Effects
Standard errors clustered on the mother

Note: Twin pairs where one or both twins have a congenital anomaly reported on their birth certificate are omitted.

9 month survey				
Outcome:	**ALL1**	**ALL2**	**SAMESEX**	**IDENT**
1 if child was ever breastfed	0.0187	0.0183	0.0187	0.0031
	[0.0237]	[0.0238]	[0.0277]	[0.0355]
	1531	1531	986	327
1 if child is now being breastfed	0.0031	0.0038	-0.0039	-0.0007
	[0.0126]	[0.0126]	[0.0152]	[0.001]
	1533	1533	988	327
How long child was breastfed in months	-0.0818	-0.0753	-0.2165	-0.343
given breastfed	[0.1722]	[0.1752]	[0.204]	[0.3182]
	805	805	514	171
Age solid food was introduced in months,	-0.1983	-0.1802	-0.2478	-0.6660*
given introduced	[0.152]	[0.1523]	[0.1906]	[0.2914]
	1530	1530	985	326
Number of well-baby visits	0.291	0.283	0.3803	0.5797
	[0.1855]	[0.1883]	[0.2414]	[0.5253]
	1529	1529	985	326
Number of well-baby visits	0.1976	0.1956	0.2329	0.2668
only children in excellent of very good health	[0.1587]	[0.1624]	[0.1944]	[0.3799]
	1478	1478	950	314
1 if caregiver praises child	-0.0075	-0.0015	-0.051	0.096
	[0.0915]	[0.0941]	[0.1189]	[0.2089]
	1229	1229	778	257
1 if caregiver avoids negative comments	-0.005	-0.0051	-0.0077	0
	[0.0054]	[0.0055]	[0.0084]	[.]
	1236	1236	782	259
1 if somewhat difficult or difficult to raise	-0.0097	-0.0181	-0.0772	-0.0946
(caregiver report)	[0.0577]	[0.0583]	[0.0712]	[0.1395]
	1531	1531	986	327
1 if not at all difficult or not very difficult to raise	0.09	0.1065	0.153	0.2237
(caregiver report)	[0.0707]	[0.0707]	[0.0812]	[0.1195]
	1531	1531	986	327
2-year survey				
1 if Caress/kiss/hug child	0.0232	0.0228	0.0055	0.0021
	[0.0279]	[0.0266]	[0.0254]	[0.0049]
	1350	1350	860	286
1 if Spank/slap child	-0.0152	-0.0195	-0.0095	-0.0048
	[0.0245]	[0.0249]	[0.0192]	[0.0316]
	1350	1350	860	286
1 if tme spent calming child >1 hr usually	0.0429	0.0317	-0.024	0.0719
	[0.0645]	[0.0646]	[0.0759]	[0.093]
	1439	1439	930	313
1 if somewhat difficult or difficult to raise	-0.0344	-0.0432	-0.0901	-0.1412
(caregiver report)	[0.0553]	[0.0555]	[0.0621]	[0.086]

	1439	1439	930	313
1 if not at all difficult or not very difficult to raise (caregiver report)	-0.0134 [0.0758] 1439	-0.0031 [0.0757] 1439	0.068 [0.0869] 930	0.0527 [0.1258] 313
Age when stopped feeding formula in months	-0.1901 [0.248] 1158	-0.1903 [0.255] 1158	-0.4504 [0.3204] 749	-0.5903 [0.7844] 251
Age when stopped breastfeeding in months	-0.1612 [0.6412] 113	-0.1492 [0.5981] 113	-0.0267 [0.044] 70	-0.0422 [0.069] 31
Preschool Survey				
1 if parent expects child to enter kindergarten early	-0.0052 [0.0115] 1283	-0.0082 [0.012] 1283	-0.0071 [0.0102] 822	0 0 267
1 if parents concerned about child's kindergarten readiness	-0.1214* [0.0562] 1297	-0.1435** [0.0554] 1297	-0.1299* [0.0636] 830	-0.1099 [0.1253] 273
1 ifexpect child to get >= 4 yrs of college	-0.0103 [0.0278] 1329	-0.0073 [0.0272] 1329	0.0069 [0.0327] 854	0.0228 [0.0264] 281
Number of servings of milk in the past 7 days	-0.0535 [0.2042] 1335	-0.0598 [0.2074] 1335	-0.0577 [0.2278] 860	0.0819 [0.2489] 282
Number of servings of vegetables past 7 days	0.0893 [0.2555] 1336	0.0632 [0.2634] 1336	0.2131 [0.3027] 861	0.0871 [0.4091] 283

Notes: Standard errors in brackets with sample sizes below.
ALL1 = mother FE regression with all twins, not controlling for child's sex
ALL2 = mother FE regression with all twins, controlling for child's sex
SAMESEX = mother FE regression with same sex twins only
IDENT = mother FE regression with identical twins only

Table 3: Sample Power Calculations

Given a true population effect size, what is the power of a size alpha = .05 test against the null hypothesis that there is no effect for different sample sizes?

Basis Study	Assumptions	Sample Size	Power
Black, Devereux, and	True model:	100	0.097
Salvanes (2007)	Prob(HSGRAD) = 0.7 + 0.1*ln(birthweight) + error	300	0.167
		500	0.263
Key result: a 1% increase	Calculation of error variance and SD:	600	0.298
in birth weight increases	Let y = Prob(HSGRAD), x = ln(birthweight), e = error	700	0.351
the probabililty of high	Var(y) = 0.44^2 = 0.19	800	0.376
school completion by	Var(x) = 0.26^2 = 0.07 (where SD(x)=0.26, according	900	0.409
0.09 percentage points.	to the distribution of ln(birthweight))	1000	0.446
	If y = 0.7 + 0.1x + e, and x and e are independent,	1250	0.531
Birth weight sample	Var(e) = Var(y) - (0.1^2)*Var(x)	1500	0.617
summary stats (twins):	= 0.19-(0.1^2)*(0.07) = 0.19	1620	0.660
mean = 2598g, SD = 612g		2000	0.744
	So, SD(e) = sqrt(Var(e)) = 0.44	2200	0.750
Probability of HS grad		2500	0.825
sample summary stats:	Therefore, assume: birthweight ~ N(2598, 612), and	3000	0.892
mean = 0.73, SD = 0.44	take the natural log of birthweight.	3500	0.928
	error ~ N(0, 0.44)	4000	0.962
		4500	0.975
		5000	0.982
		5500	0.993
		6000	0.994
		6500	0.996
		7000	0.999

Given a sample size, how large would the true effect size have to be in order to be able to detect it with reliable power using a test of size alpha = .05?

Basis Study	Assumptions	True B1	Power
Conley, Pfeiffer and	Model: y = B0 + B1*x + error	0.005	0.046
Velez (2006)		0.01	0.047
	Assume: z ~ N(2598, 612), x = ln(z)	0.02	0.077
Sibling sample from	error ~ N(0, 0.44)	0.03	0.092
PSID (n=1,360)	sample size = 1500	0.04	0.146
		0.05	0.198
		0.06	0.274
		0.07	0.354
		0.08	0.461
		0.09	0.525
		0.1	0.631
		0.12	0.769
		0.15	0.926
		0.17	0.975
		0.2	0.99

Notes: Power calculations are based on Monte Carlo simulations with 1000 replications.