Human Capital Development Before Age 5

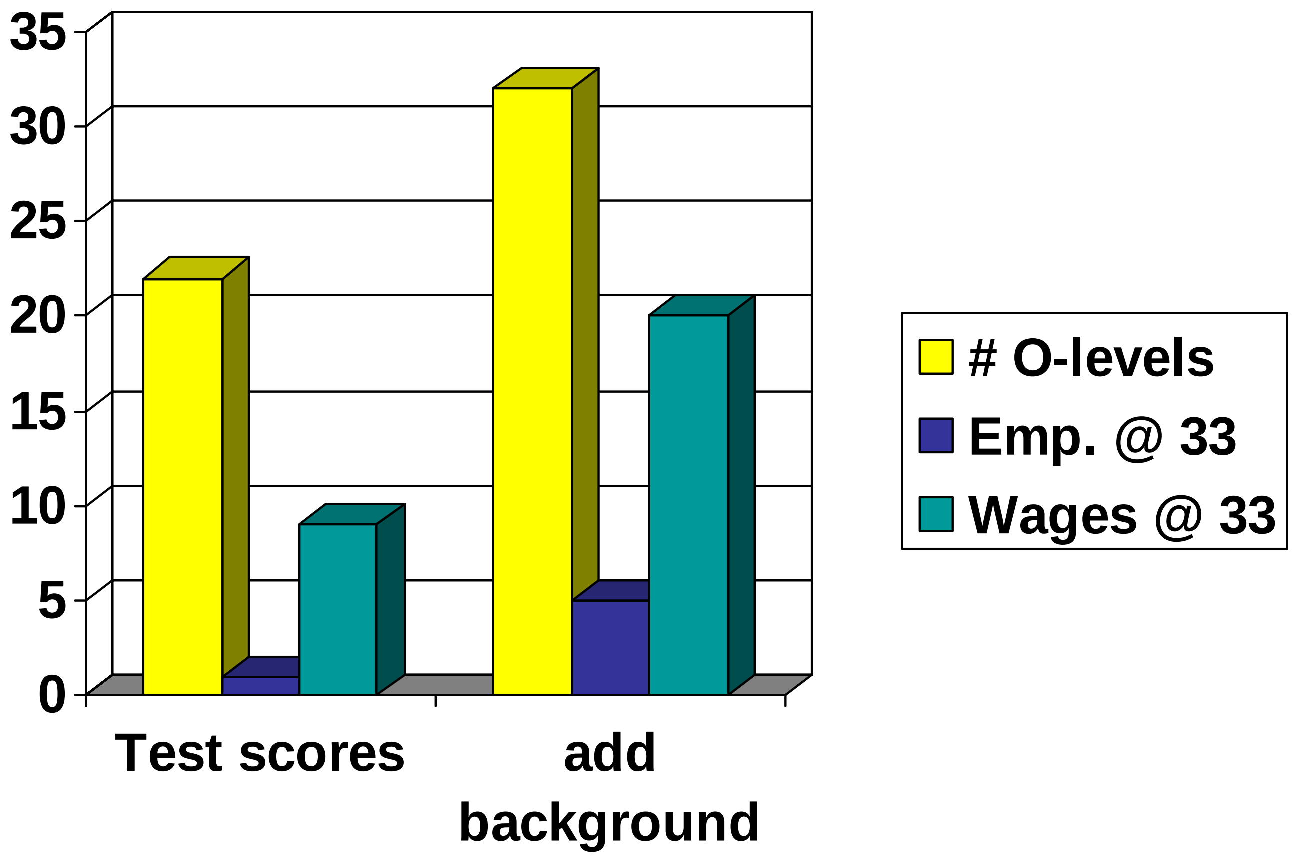
Notes based on Douglas Almond and Janet Currie (forthcoming, Handbook of Labor Economics)

This is a growing area of research

- In 2000 there were no papers on this topic in the AER, QJE, JPE. Since 2005, there have been 5 or 6 per year.
- Why? Why do Economists care about human capital development in the early years?
- Things measured in early childhood predict much of the variation in future outcomes.

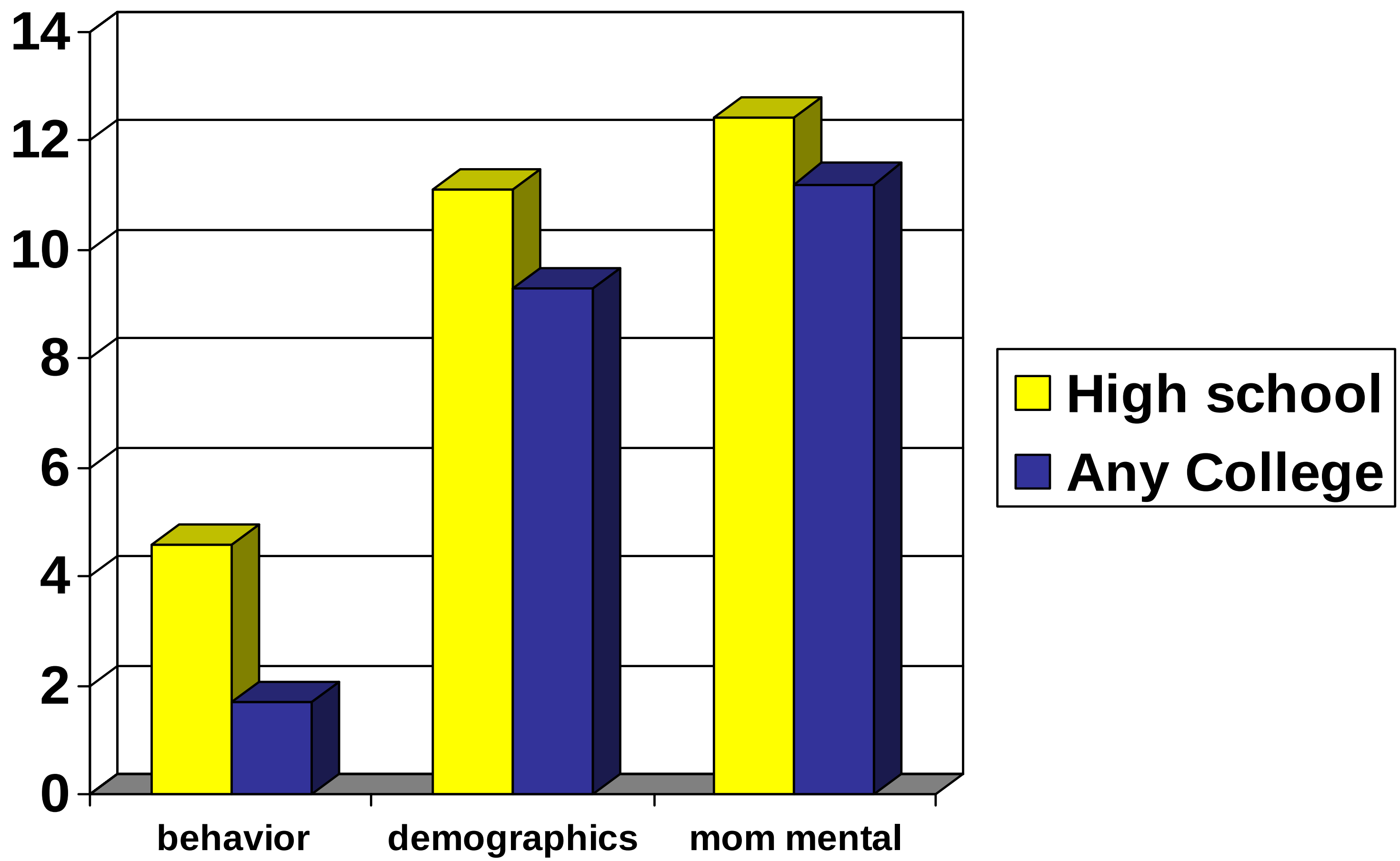
Consider the 1958 British Birth Cohort

Graph shows fraction of variance explained by math and reading scores at age 7 alone, and with background variables measured at age 7. Currie and Thomas (1999)



Consider the U.S. NLSY

Graph shows percent of variation accounted for by child behavior problems @6-8, demographics, maternal mental health, McLeod and Kaiser (2004).



These studies raise many questions:

- The studies are descriptive. Are the relationships causal? How stable are they?
- What are the mechanisms? (e.g. nature vs. nurture)
  - Biological?
  - Social?
  - Interactions biology and society?
- Can they be changed?

Overview

- Theory
- Methods (a few points)
- Evidence re: causal effects of shocks in utero and in early childhood
- Evidence re: effectiveness of remediation

Original model of Human Capital
(Becker (1964) Grossman (1972))

Consumers are assumed to maximize an intertemporal utility function:

$$\sum_{t=1}^{T} \mathrm{E}_{\mathrm{t}}(1/1+\delta)^{t} U_{t} + B(A_{T+1}), \qquad (1)$$

where δ is the discount rate, $B(\bullet)$ is a bequest function, A denotes assets.

U_t is given by:

$$U_t = U(Q_t, C_t, L_t;\ \mathbf{X}_t, u_1, \varepsilon_{1t}), \tag{2}$$

where Q is the stock of health, C is consumption of other goods, L is leisure, $\mathbf{X}$ is a vector of exogenous taste shifters, u is a vector of permanent individual specific taste shifters, and ε denotes a shock to preferences.

Utility is maximized subject to the following set of constraints:

$$Q_t = Q(Q_{t-1}, G_t, V_t; \mathbf{Z}_t, u_2, \varepsilon_{2t}), \tag{3}$$

$$C_t = Y_t + P_t G_t - (A_{t+1} - A_t), \tag{4}$$

$$Y_t = I_t + w_t H_t + r A_t, \tag{5}$$

$$L_t + V_t + H_t + S_t = 1, \tag{6}$$

$$S_t = S(Q_t, u_3, \varepsilon_{3t}), \tag{7}$$

where G and V are material and time inputs into health production, **Z** is a vector of exogenous productivity shifters, u_2 are permanent individual specific productivity shifters, ε_{2t} is a productivity shock, Y is total income, P represents prices, I is unearned income, w is the wage, r is the interest rate, S is sick time, u_3 are permanent individual specific determinants of illness and ε_{3t} are shocks that cause illness.

Endowments of health and assets, Q_0 and A_0, are assumed to be given.

- Grossman model considers health as a stock variable that varies with an individual's age
  - Responds to investments and depreciation
  - Health stocks at earlier ages matter to adult health...
    ... but their effect (and health investments they embody) dwindles over time
  - No special role for early childhood ages
- We want to leave open whether there is indeed "fade out" of health investments/experiences at early ages

Defining h as health or human capital at the completion of childhood

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

- $I_1 \cong$ investments during childhood through age 5
- $I_2 \cong$ investments during childhood after age 5.
- γ can be greater than .5
- Perfect substitutability in equation (1) not uncommon assumption in economics, but problematic for early origins

Heckman suggests more flexible "developmental" technology:

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

- Constant elasticity of substitution (CES) production function
  - Elasticity of substitution $1/(1-\phi)$
- Perfect substitutability of investments still allowed (when $\phi = 1$)

Shocks to early childhood:

- Holding other determinants of investments fixed, consider investment shock:

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

- Long-term damage from a negative μ_g is:

$$\frac{\delta h}{\delta \mu_g}$$

  - Magnitude of damage can depend on levels of $\bar{I}_1$ and $\bar{I}_2$
  - Relevant for empirical findings of heterogenous "early origins" damage
  - "Biological" effect assumes no investment response

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.
- Other things equal, those with higher baseline levels of investment will experience more muted effects on h than those where baseline investment is low.
- 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 because they are on a steeper portion of the production function.

Remediation:

Consider a change to second period investments after shock μ_g in early childhood:

- Effectiveness of remediation depends on ϕ, γ and $\bar{I}_1$
- Knowing $\frac{\delta h}{\delta \mu_g}$ to be big doesn't say much about effectiveness of remediation
- Optimal investment response also depends on utility function, e.g.:

$$U_p = (1-\alpha)\log C_p + \alpha \log h$$

- Where C_p is the consumption of parents

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

Optimized Investment Response

$\Delta I_2^* \cdot \mu_g < 0$: **Compensation** helps offset damage

$\Delta I_2^* \cdot \mu_g > 0$: **Reinforcement** accentuates damage

- For $\phi > 0$, compensation optimal
- For $\phi < 0$, reinforcement optimal

To the extent there is a response, then missing the "biological" effect.

- Can understate total damage from μ_g by focussing exclusively on reduced form $\frac{\delta h}{\delta \mu_g}$

Consider parents making optimal investments

- If there is perfect substitution between I_1 and I_2 then,
  - If there is a negative shock in period 1, then the MU of h becomes higher relative to parental consumption, so parents will increase investments in period 2.
  - If there is a positive shock, then the opposite is true and parents should decrease investments in period 2.
  - So investments are compensating.

- But if the production technology is Leontief, then compensation is ineffective. H depends on the minimum of I_1 and I_2 so a negative shock in the first period will cause a reduction of investment in the second.
- Hence, optimal parental response reflects not only parental preferences but production technologies/budget constraints.

Summary

- There is a good deal of evidence that events in early life affect adult outcomes.
- Measured relationships are not necessarily “biological” but reflect parental preferences and available production technologies for child outcomes (both of which may be subject to change).

Some issues with respect to methods

- Estimation of production functions
- Sibling fixed effects
- Power
- Data issues

Production Functions

- Economists think of there being a "production function" for child outcomes that turns "inputs" (like reading to a child) into "outputs" (like higher educational attainment.
- This model suggests several reasons that parent background (especially income and education) matters:

The Production Paradigm

– Richer families can afford more and better “inputs” (more books, more activities).
– Parents with more education may be better able to take a given set of “inputs” and produce “outputs”(e.g. a mother with a richer vocabulary will teach her child more words per time spent with them).
– Children born to parents with lower incomes and education have worse health at birth on average, so start from behind.

The Production Paradigm Has Several Implications:

- There may be more than one to produce the same outcome (e.g. it may be possible to substitute some money for some time).
- The effects of a given input will vary depending on what other inputs are chosen (e.g. effects of more money could be negative with inadequate parental supervision).
- More inputs may matter more to children who start with few.
- Although this depends in part on whether “early” and “late” investments are substitutes or complements.

So the production function paradigm provides a useful framework

- But in the absence of appropriate measures of inputs and outputs, or any knowledge of appropriate functional forms we should not take the paradigm too literally.

Sibling Fixed Effects

- A useful way to control for fixed characteristics of family background.
- However, the preceding discussion emphasized that parents might respond to shocks to an individual child.
- This is just one example of individual-specific omitted factors that might affect estimates.

What is the evidence regarding parental investments?

- Some emphasis of reinforcing behavior in developing countries (Rosenzweig and Schultz, 1982; Rosenzweig and Wolpin, 1988).
- These results might reflect budget constraints – if a poor family can only invest in one child, they could choose to invest only in the child most likely to succeed.

Evidence from Developed Countries

- Datar, Kilburn and Loughran (2010) use data from the NLSY and show that low birth weight children are less likely to be breastfed, have fewer well-baby visits, are less likely to be immunized, and are less likely to attend preschool than normal birth weight sibs.
- However, these differences might be due to poorer health among the LBW sibs.

- Hence, they also look at how the presence of a LBW child in the household affects investments in other children. Is there evidence that a LBW child causes parents to reallocate resources to the healthy one?
- No. A LBW sib has no effect on a normal birth weight child’s breastfeeding, immunizations, or preschool.
- There is an effect on visits for well-baby care, but this could be due to economies of scale in medical utilization.

- Similarly, Del Bono, Ermisch, and Francesconi (2008) estimate a model that allows endowments of other children to affect parental investment in the index child. But results are very similar to mother fixed effects.
- Royer (2009) also finds that birth weight differences are unrelated to breastfeeding, NICU admission in twin pairs in the ECLS-B.
- Kelly (2009) finds that parental investments (time reading to child) are unrelated to damage from the 1958 flu epidemic (using NCDS).

- Hsin (2009) uses PSID and suggests that maternal time use is positively related to birth weight in low income families, and negatively related in high income families. In other words, poor families reinforce (as in the LDCs) and rich families compensate.
- This is an interesting hypothesis, though the sample size is very small (65 sibling pairs with a difference in LBW).

Summary:

- This is a very new area of research
- So far, little evidence of reinforcement in rich countries
- Note that in fixed effects models, reinforcement will cause effects to be over-estimated, while compensation will cause effects to be understated. So compensation will not bias towards finding effects which are not there.

- In sibling FE it is important to consider possible sources of individual differences and test for them to the extent possible.

Power calculations

- Many data sets that have detailed information about child “inputs” and “outputs” are relatively small.
- Power calculations can be used to determine whether it is likely that any effect can be detected.

Example

- Black et al. (2007) use registry data from Norway to determine that a 1% increase in birth weight increases high school completion by .09 pp.
- How large a sample size is necessary to detect an effect of this magnitude? i.e. the null is no effect, and we want to be able to reject the null with 95% confidence?

True model: $p(HS)=.7+.1*\ln(birthweight) + e$

- Birth weight (in a sample of twins) ~N(2598,612)
- Mean HS=.73 with sd.44
- Can compute e~N(0,.44)
- Do Monte Carlo simulation with 1000 replications

Results: Need a sample size of ~4000

Sample size	Power
100	.097
300	.167
700	.351
1000	.446
1500	.617
2500	.825
3000	.892
3500	.928
4000	.962

Given a sample size, how large would an effect have to be to detect it 95% of the time?

- Suppose we take a sample size from the PSID of ~1,500.
- Let $y = b0 + b1*X + e$
- $Z \sim N(2598,612)$, $x=\ln(z)$
- $e \sim N(0,.44)$

Results: We can only reliably detect a b1 of >.15.

True b1	power
.005	.046
.01	.047
.03	.092
.05	.198
.08	.461
.12	.769
.15	.926
.17	.975
.2	.990

Data Issues

- Lack of large scale longitudinal data is a huge problem for this line of research.
- New longitudinal data collection is not necessarily a panacea
  - Expensive
  - Takes a long time to be able to see long-term outcomes
  - Sample attrition can pose problems for inference

Existing Data Can Offer a Solution

- Can add questions to existing data sets
- Can merge new information to existing data sets
- Can merge several administrative data sets

Example of Adding Questions:

1. Garces, Thomas, and Currie (2002) had special questions on early childhood education added to the 1995 Panel Study of Income Dynamics. Adult respondents <=30 were asked whether they had ever been enrolled in Head Start and whether they had attended any other preschool.
2. Smith (2007) examines retrospective questions added to PSID on health when <16.

Pros:

- The PSID is a long panel with rich contextual information.
- The PSID offers a large sample, and also samples of siblings.
- Given this data we were able to assess the impact of Head Start on Young Adults, Smith examines the impact of early health problems.

Cons:

- Retrospective data may be reported with error.
- Only outcomes already in the data can be assessed.

Respondent Error in Retrospective Data Can be Assessed

- Compare reported participation rates to known rates
- Compare distributions of participant characteristics to known distributions from other data.
- Investigate whether there is more “forgetting” further in the past
- These tests are not conclusive but offer some reassurance of data quality.

Merging new data to existing data sets

- Generally done by geographic area and date, so requires geocodes.
- Many examples:
  - Ludwig and Miller (2007) exploit the fact that the Office of Economic Opportunity initially offered the 300 poorest counties assistance in applying for Head Start. Using the NELS, they show that children in counties with assistance were much more likely to have attended Head Start than children in otherwise similar counties.

Ludwig and Miller go on to show:

- Using Vital Statistics data they show that county-level mortality rates were lower in counties whose Head Start enrollment rates 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, the Census does not collect county of birth ($\rightarrow$ measurement error, possibility of selective migration).

Merging Administrative Data Bases

- Requires personal identifiers.
- Example: Currie, Stabile, Manivong, and Roos (2008) merge data from Canadian public health insurance records to welfare rolls and data from the educational system.
- Using this data, it is possible to ask whether health problems in early childhood are related to future welfare use or lower educational attainment.

Pros and Cons:

Pro: Large sample

Objective indicators of health and of outcomes available for full sample

Sibling comparisons possible

Long follow up period

Con: Limited background information

Health measures depend on utilization of care

Outcomes limited to those in data bases that can be merged in.

Findings:

- Major physical health conditions at age 0-3 are predictive of poorer educational attainment and welfare use.
- This is mainly because poorer physical health conditions at 0-3 predict poorer future health.
- Mental health conditions at early ages are independently predictive of poorer outcomes.

Other examples:

- Black, Devereux, Salvanes (QJE 2005,2007) use Norwegian data to look at long-term effects of birth weight, birth order, and family size on educational outcomes. Requires linking birth and education registries.
- Doyle (2008) links child protection and criminal justice data bases in Illinois. Shows that on the margin, children who remain at home rather than being placed in foster care are less likely to be incarcerated later.

The Major Challenge:

- Privacy concerns are making it harder to obtain data just as it is actually becoming more feasible to link it. E.g. public use vital statistics natality data no longer report county of birth.

Potential Solutions

- Creators of large scale data sets need to be sensitive to the fact that their data may well be useful for questions they have not yet envisaged.
- In order for data to be used in this way, it is essential to retain information that can be used for linkage.
  - If it is undesireable to keep personal identifiers, geographic identifiers at a fine unit of disaggregation (e.g. Census tract or zip code) should be kept.

Explore Methods for Making Sensitive Data Available

- Suppress small cells, and/or data on rare outcomes in public use files. E.g. NCHS data sets such as NHANES and NHIS ought to report state at least for large states.
- Add a small amount of noise to public use files (or data swapping to prevent identification of outliers).
- Data use agreements (e.g. NLSY, NELS).
- Creation of de-identified merged files.
- Secure data facilities.

Summary

- Many secrets are currently “locked up” in existing data that researchers do not have access to.
- Exploring ways to make these data available might in many cases be a more cost-effective way to answer questions than carrying out new data collections.

Additional Issues

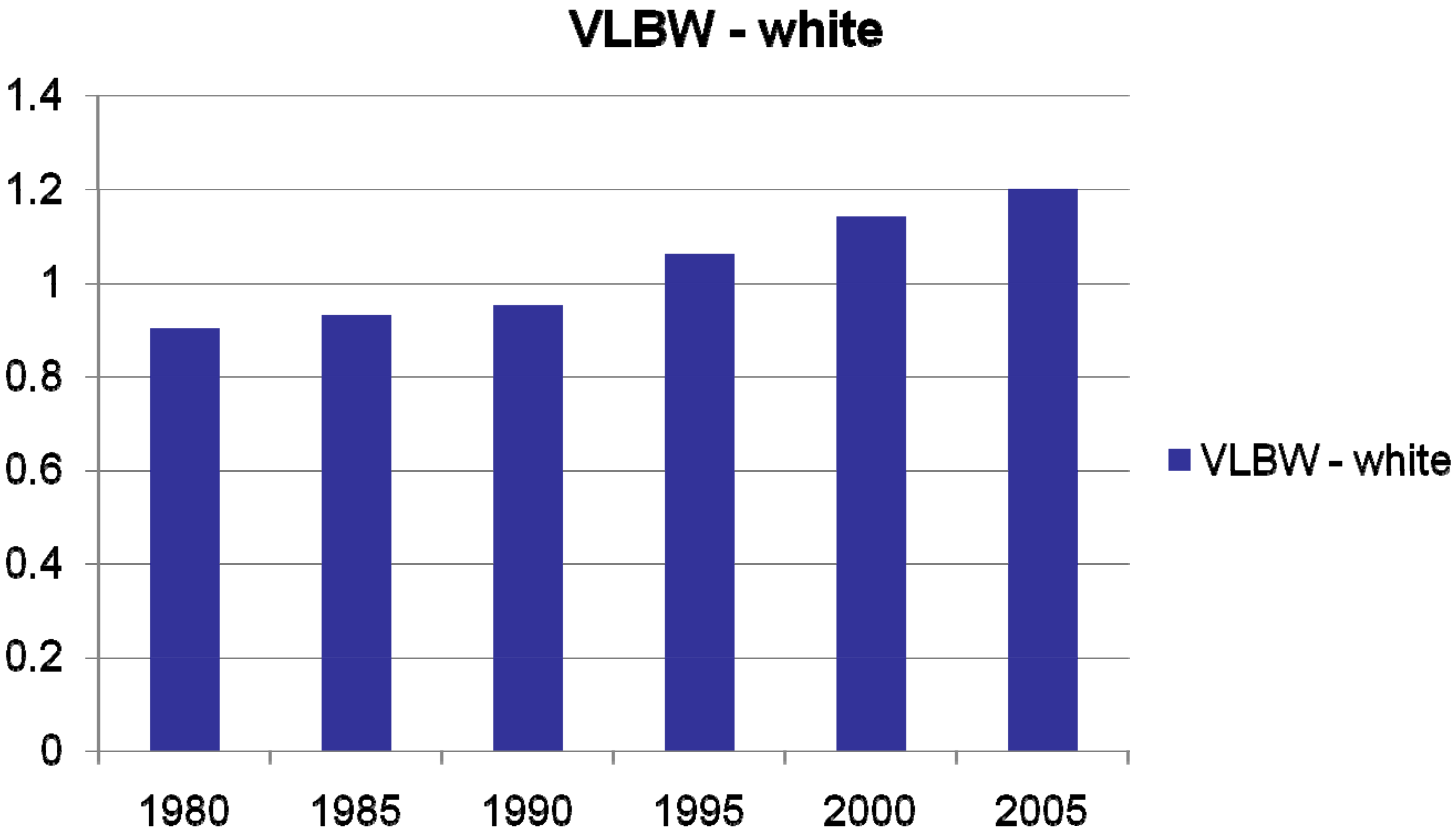
Biomarkers

More data sets now include measures such as cortisol (stress) or genetic markers.

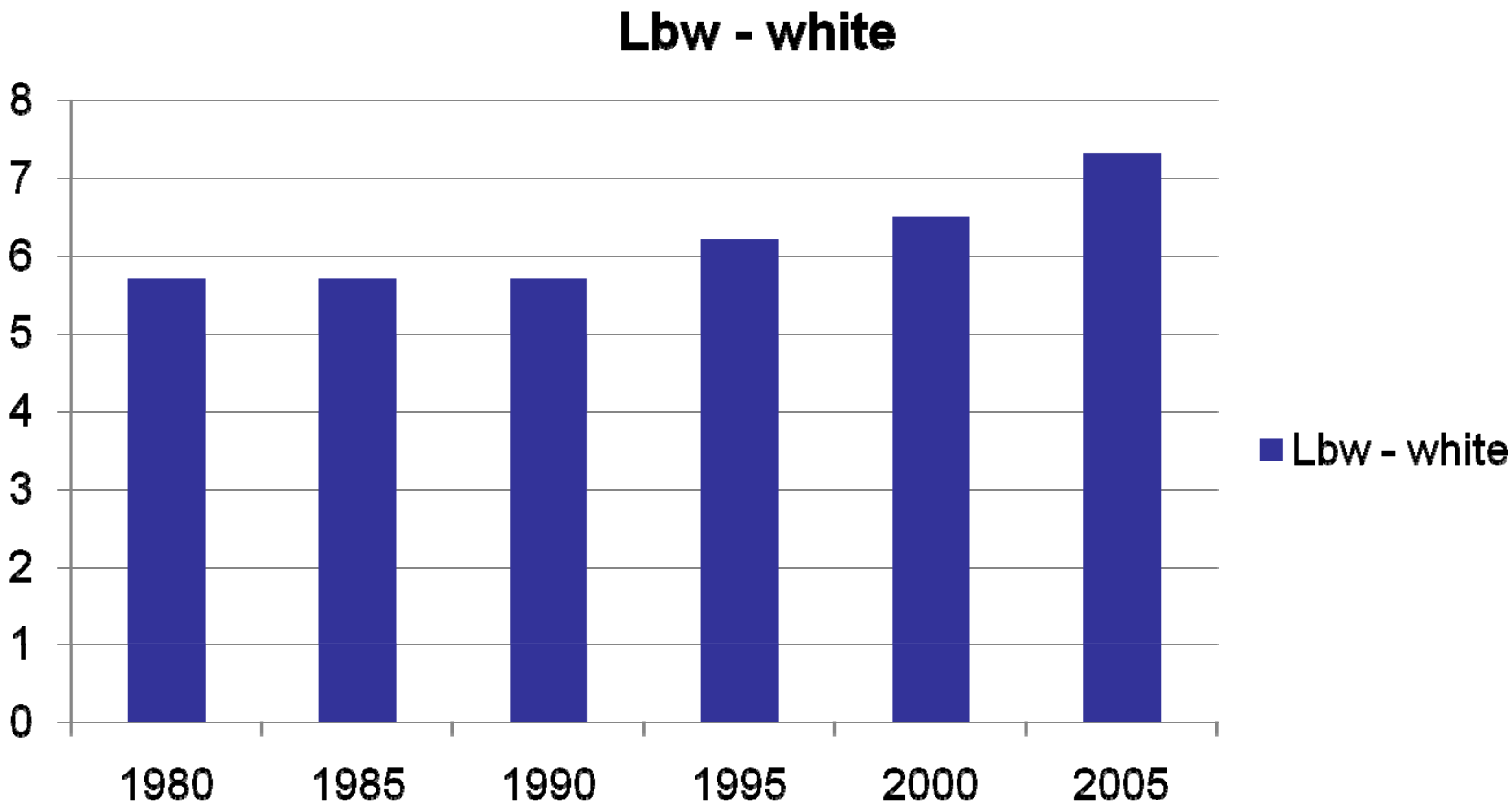
It is tempting to think of genes in particular as predetermined and therefore candidate "instruments".

But do we know enough at this point to say that a particular gene affects say schooling only through its effects on IQ and not through effects on personality (or vice versa)?

Changes in birth cohorts



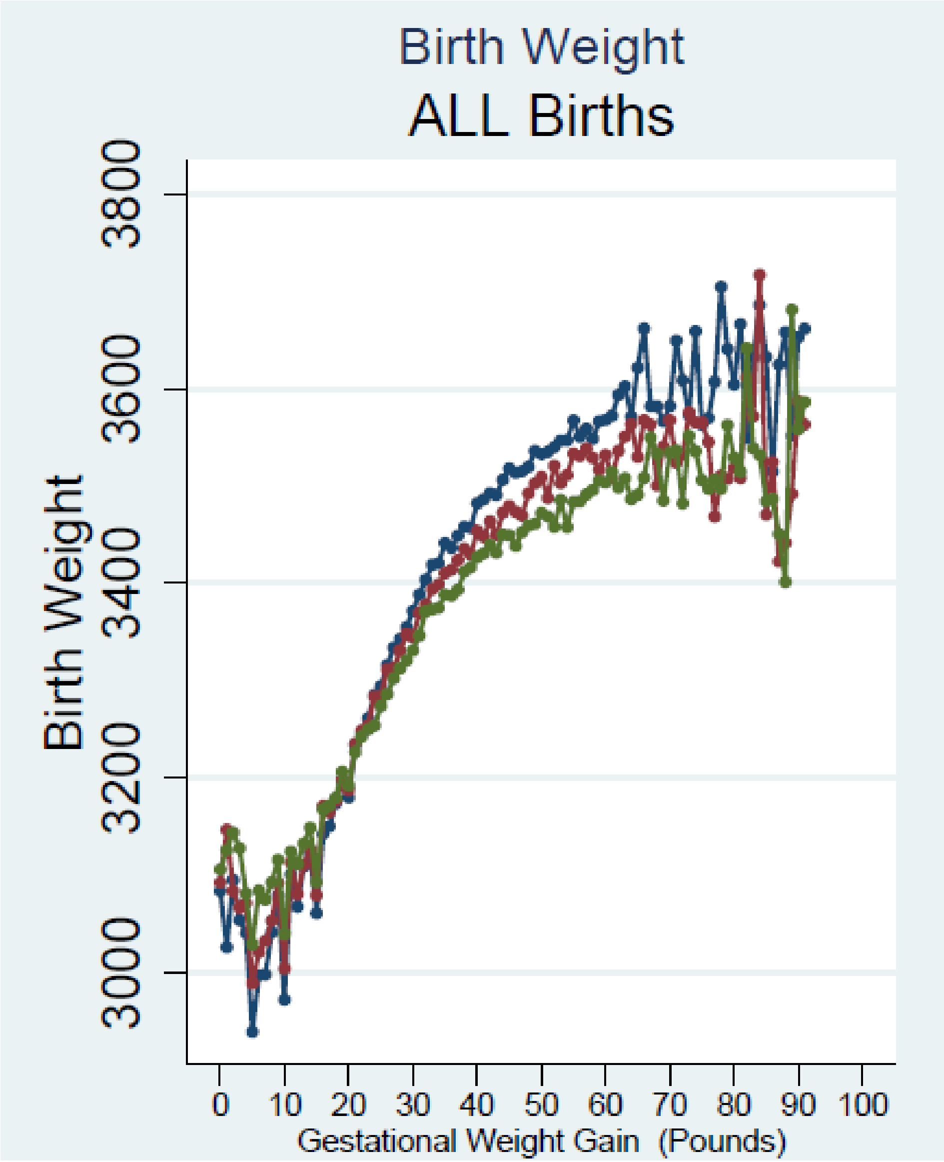
Changes in birth cohorts

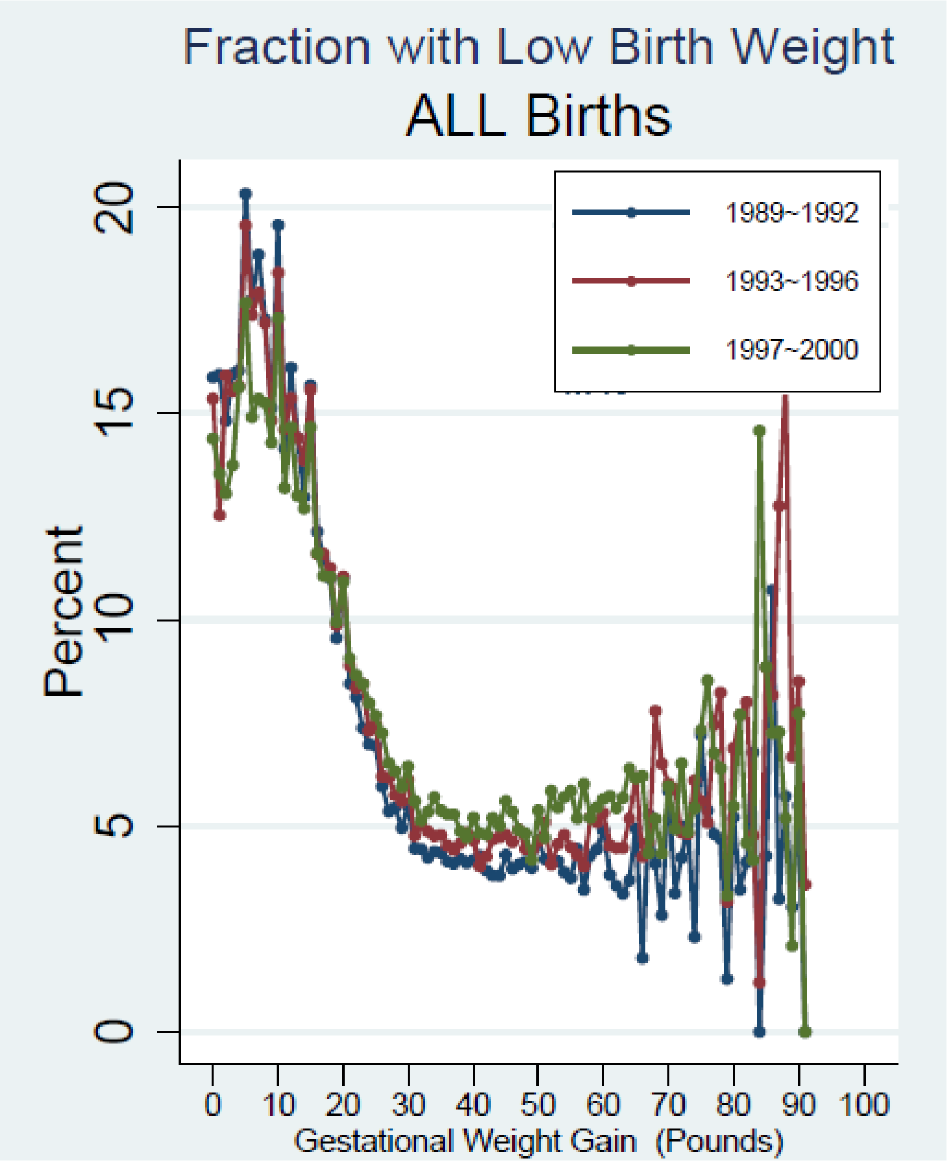


There are several possible explanations

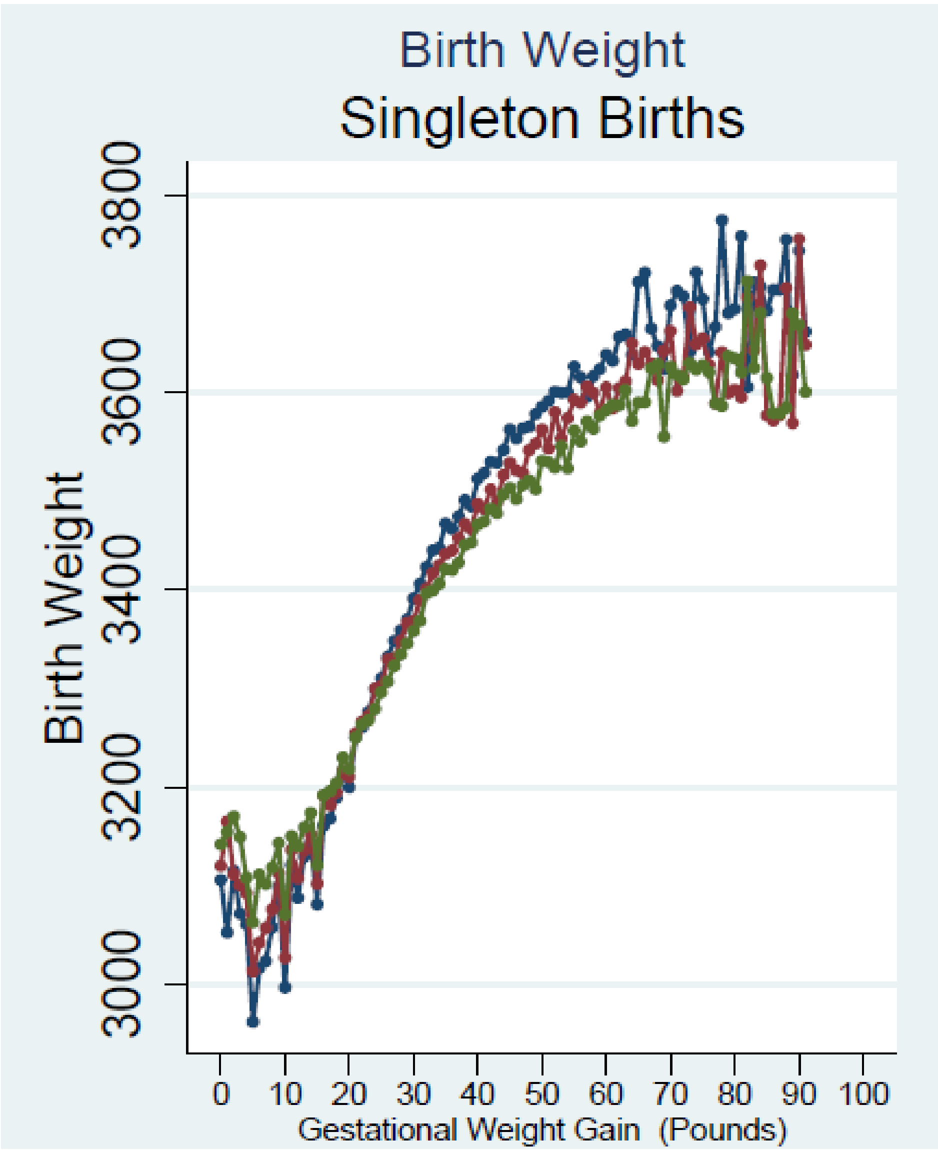
- Assisted reproductive technologies → increase in multiple births
- Older maternal/paternal age
- Better neonatal medicine
- Increasing income inequality
- Possible changes in reporting
- The implications of these changes in the distribution of birth weight for future outcomes are so far unexplored.

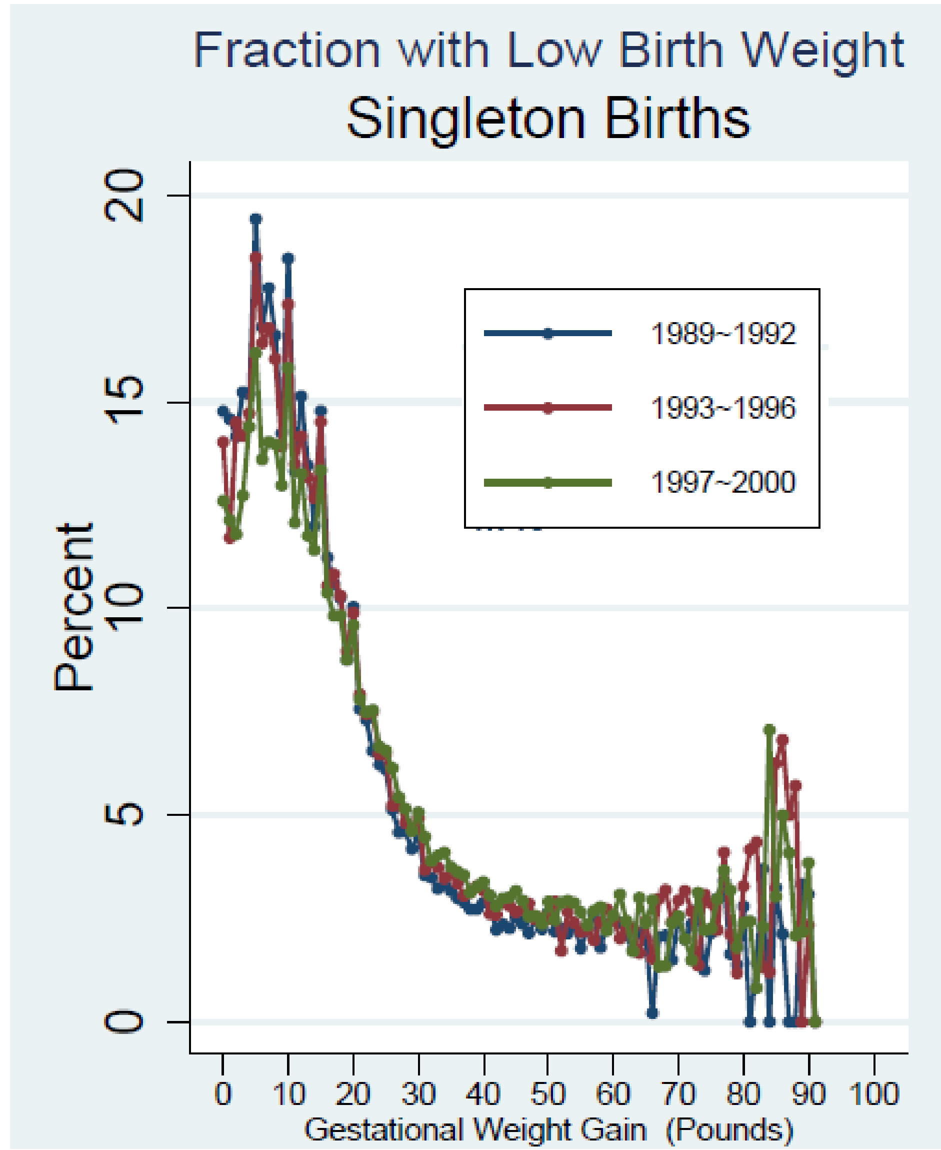
The distribution has been changing given maternal behavior



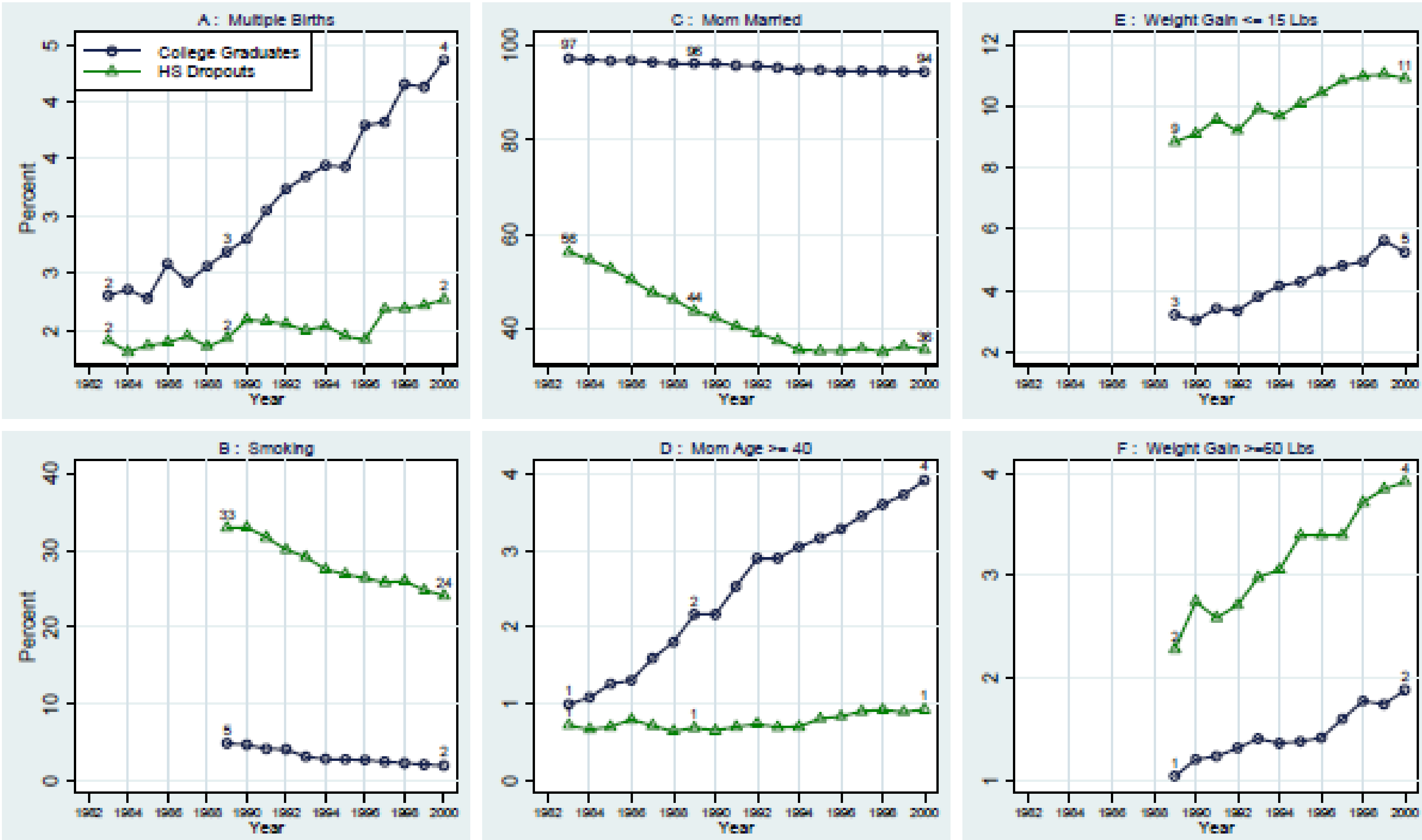


And for singleton births





But maternal behaviors have also changed over time, with unknown consequences for child outcomes.



Summary

- I have given an overview of Almond and Currie's discussion of the theory regarding early life investments, and some empirical issues including:
  - Use of fixed effects models
  - power calculations
  - Data sources
  - Biomarkers
  - Changes in birth weight distributions