Long-Run and Intergenerational Impacts of Disease and Medical Technology:
Evidence from the Sulfa Drug Innovation

*Sonia Bhalotra**

*Atheendar Venkataramani***

*University of Bristol UK, **University of Washington St Louis

A Preliminary First Draft

Abstract

This paper investigates the long run and intergenerational impact of the arrival of sulfa drugs in the US in 1937 on a diverse set of developmental outcomes including cognition, health and survival. Identification exploits the discontinuous change in the rate of decline of mortality in 1937 that was unique to diseases treatable with sulfa drugs and the force of which varied systematically with regional pre-intervention levels of mortality. The analysis focuses upon infant mortality which exhibits a trend break in 1937 that, we find, is more pronounced in regions with lower pre-intervention mortality. Mortality rates from diarrhea for children under the age of two and from tuberculosis test as suitable controls. We merge historical disease-specific mortality rates at the state or region level by mother's birth year with outcomes for mothers and their births in the CPP and the census. Using the CPP, we find weak and mixed evidence of the impact of infant mortality rates at birth on the later life outcomes of women. Women born in regions and years of lower infant mortality in general and after 1937 in particular are significantly advantaged in terms of stature but appear to be disadvantaged in terms of schooling, although there is some evidence of an advantage in the likelihood of progressing to college and in normalized family income. Controlling for indicators of the mother's health and socioeconomic status, we find more pervasive and robust impacts of disease risk in her birth year on outcomes for her births. Births of mothers who were born after the arrival of sulfa drugs performed significantly better in terms of birth weight, motor skills, death risk and IQ. Since our indicator of medical innovation is the (change in the) infant mortality rate these may be regarded as third generation effects. The effects are consistently stronger for boys than for girls, consistent with the well documented excess vulnerability of boys to the early life environment. Preliminary results from census data are broadly corroborative. Overall, the results suggest that the benefits of medical innovation are transmitted across generations, evident up to three decades later, and include gains in both health and cognition.

What we do

Investigate the long range impact of medical innovation on human capital outcomes.

- Sulfa drugs arrived exogenously in the US in 1937 and were widely used to treat a range of high-prevalence infectious diseases.

- We investigate whether the later life outcomes of individuals born in and after 1937 are systematically better than for individuals born before.

- We also investigate whether children of mothers (and fathers) born in and after 1937 do better.

Motivation

Contributes to evidence of:

(a) the (typically neglected longer run) return to investments in medical innovation.

(b) the lasting impact of the early life health and socioeconomic environment (Almond and Currie 2010), and the importance of investments at critical ages (Cunha and Heckman 2007).

- Possibly the first direct analysis of medical innovation, though closely related to previous analyses of the early life disease environment (Almond 2006, Deaton 2007).

- Few previous studies analyse intergenerational effects (Almond & Chay 2006, Almond et al. 2007, Fung 2010, Bhalotra 2010)

- We analyse a wider range of outcomes than previous studies.

Sulfa drugs

Sulfa drugs is a shorthand for antimicrobial sulfonamides.

- Precursor to modern day antibiotics, which did not arrive till the mid-40s. Still used but sparingly because of common side effects
- Innovation in German lab in 1932
- First clinical trial in NY in 1935, NY Times article Dec 1936 launched it. Widespread and affordable by 1937.
- A sharp, large, exogenous change

Evidence of their short range impact

Clinical trials in the US and Europe document sizeable effects for a number of infectious diseases.

Using nation and age averages of disease-specific mortality rates, Jayachandran et al 2010 (*JLS*) demonstrate a trend break in 1937 for mortality from diseases treatable with sulfa drugs, namely, maternal mortality (mmr), pneumonia & influenza (flu) and scarlet fever.

They show that diseases like tuberculosis (*TB*) that were not treatable with sulfa drugs showed a more continuous decline.

Size of effect

Using TB as a control, they estimate that sulfa explained 56, 39 and 76 % of the pre to post 1937 decline in (all-age) mortality in mmr, flu & scarlet respectively.

These estimates ignore the contribution of sulfa to

(a)   mortality from other treated diseases for which data are unavailable and morbidity

(b)   their longer run benefits

Our project

We take this research agenda forward-

- First, we investigate the impact of sulfa drugs on infant mortality in 1928-1943. This extends the evidence on short run impact.

  - *IMR* captures the total effect of *disease-specific* improvements on newborns.

  - It responds predictably to changes in the unobserved distribution of cohort health.

- We then study the impact of birth year exposure to the arrival of sulfa drugs on later life and next generation outcomes, using timing to effectively instrument infant mortality with the sulfa innovation.

Identification

- Short-run impact:
  - Assess trend break and regional convergence in IMR

- Long-run impact:
  - Exploit birth region*year variation in IMR, net of birth region*year variation in other childhood and infectious diseases and conditional on birth region and birth year fixed effects

  - Assess whether long-run outcomes follow similar patterns as short run

Data

The analysis uses three sources of data-

- Longitudinal microdata on births, mothers (and a subset of fathers) from the National Collaborative Perinatal Project (*CPP*), 1959-1974.

- Microdata from the 1970 and 1980 census. Information on 1st and 2nd generation.

- State level disease-specific mortality rates, 1925-1945

    - merged by census region with the CPP
    - merged by state with the census.

The CPP data

- Large. >55,000 pregnant women enrolled in 1959-1966
  - Useful as long run effects are often small.

- Longitudinal.
  - Multiple survey years help separate effects of age and cohort of mother.
  - Tracking of her births allows us to study persistence of effects and to control for child-cohort and year effects.

- Timing. Women born post-sulfa are 22-29 in the CPP. This is the peak childbearing age. We restrict mother's age at birth to the interval 19-35.

- Very detailed data, range of health and cognitive outcomes.

Analysis

Short range impact

(i)Trend break in early life disease

(ii)Region variation in post-trend with pre-sulfa level of mortality

Trend break in infant mortality series

m_{it} is log infant mortality rate in state i in year t

$post = 1(year>=1937)$, so sample includes 9 years pre and 7 years post. Will check robustness to excluding 1936.

$$m_{it} = b_0 + b_1 post + b_2 year + u_i + e_{it}$$
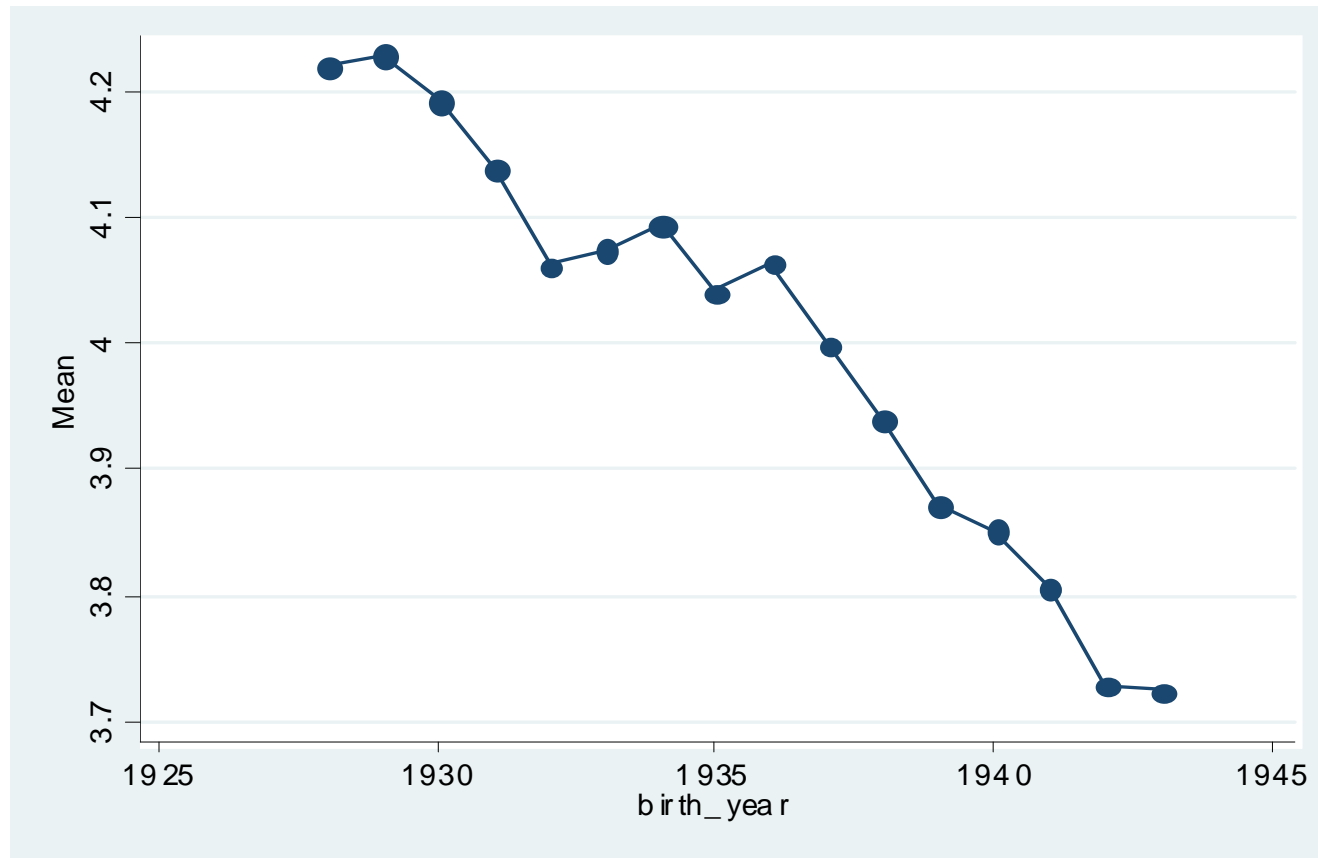
$b_1<0$ indicates an intercept shift associated with arrival of sulfa

$$m_{it} = a_0 + a_1 post + a_2 year + a_3 post*year + u_i + e_{it}$$

A level break again implies $b_1<0$ but we are now testing also for a trend break, $b_3<0$

There is a trend break in *imr* in 1937- this defines *post-37*

log (IMR) by Year, State Averaged Data, 1928-1943

Trend Breaks for Treated v Untreated Diseases

	IMR	Pneumonia	MMR	Diarrhea	TB
Post	-0.018	-0.123***	-0.254***	0.140***	0.037***
	(0.011)	(0.025)	(0.023)	(0.046)	(0.011)
Post	-0.010	-0.025	-0.123***	0.106***	0.031***
	(0.010)	(0.023)	(0.019)	(0.049)	(0.012)
Post*Year	-0.020***	-0.076***	-0.101***	0.024**	0.005*
	(0.002)	(0.005)	(0.004)	(0.011)	(0.003)
N	771	882	882	881	882
States	48	48	48	47	48

Dependent variable is logged mortality rate; sample is 1925-43
Post=(year>1937). Models include year and state fixed effects.

Trend Breaks for Treated v Untreated Diseases

	ln(M)
Post*Treated	-0.0260**
	(0.00974)
Post*Treated*Year	-0.00586***
	(0.00108)
Treated*Year	0.0108***
	(0.00178)
Post	0.0236**
	(0.00939)
Year	-0.0174***
	(0.0012)
Treated (=1)	-0.172***
	(0.0429)
N	2373
R-squared	0.699

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
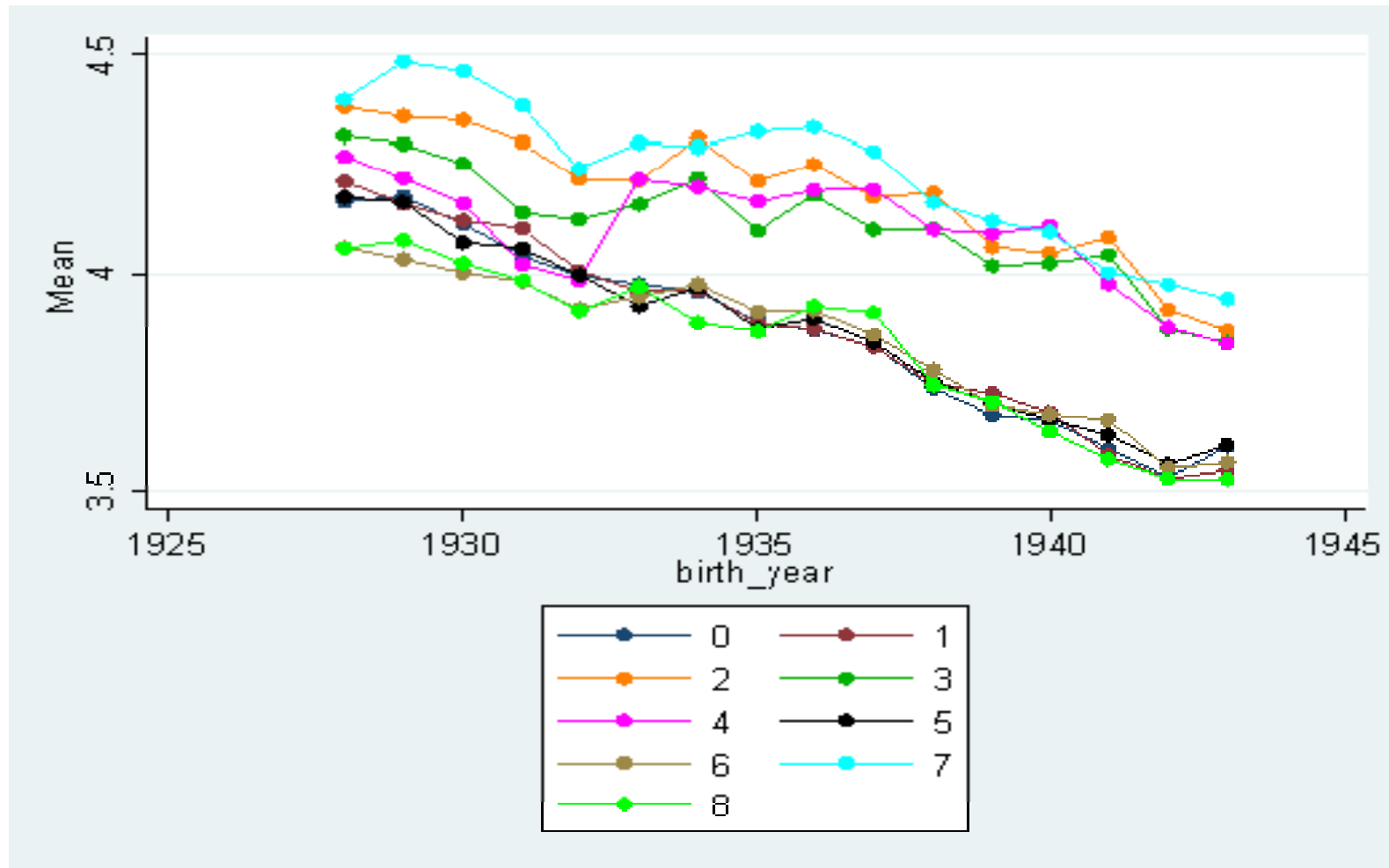
Convergence/Divergence in IMR

Variation in treatment intensity, as in Bleakley 2007

$$m_{it} = b_0 + b_1 post*base + b_2 year + u_i + e_{it}$$

$$m_{it} = a_0 + a_1 post*base*year + a_2 post*base + a_3 post*year + a_4 base*year + a_5 year + u_i + e_{it}$$

Steeper decline in regions with lower initial levels of mortality

Divergence in IMR

	IMR	Pneumonia	MMR	TB
Post*Baseline	0.074***	-0.025	0.273***	-0.032**
	(0.024)	(0.050)	(0.047)	(0.014)
Post*Baseline	0.086*	0.148	0.191**	-0.001
	(0.051)	(0.137)	(0.075)	(0.027)
Post*Baseline*Year	-0.012	-0.004	-0.0001	-0.012**
	(0.012)	(0.031)	(0.018)	(0.006)
N	748	859	859	859
States	46	46	46	46

Mortality rates are logged, baseline is level in 1925-43.
All main and two-way effects are included.

What do we make of this?

- Untreated diseases show convergence. Consistent with health being concave in overall progress.

- In contrast, the sulfa treatment appears to have caused divergence across regions in the level of treated disease mortality.

    - Medical innovations like sulfa tend to be adopted earlier/ more widely in urban, white communities (Almond et al., JLS). These regions will tend to have lower pre-intervention mortality (endogenous adoption)

    - Additional survivors from sulfa in high mortality regions succumbed to untreated diseases

First generation outcomes

The impact of infant mortality rates (and selected disease-specific mortality rates) in an individual's birth year on their later life outcomes

The long run impact of infant mortality risk in the birth year

The expectation is that the short run patterns are mirrored in longer run outcomes.

$$y_{ijc} = a_0 + a_1 imr_{jc} + a_2 diar_{jc} + a_3 tb_{jc} + a_4 X_{ijc} + \gamma_c + \delta_j + e_{it}$$

Y is an indicator of health, cognition or SES in adulthood
Subscripts are i for individual, j for region of birth, c for cohort
imr, diar, tb are log rates of infant mortality, diarrhea <age 2, TB

$a_1 < 0$ indicates scarring.

- X includes age at interview, race, religion, city of current residence. In a variation, it includes region-specific linear trends

- This specification does not directly use the timing of sulfa arrival.

- It exploits the sulfa shock only in looking at the impact of IMR which we have seen declined more quickly after 1937 relative to TB and diarrhea mortality (which were not treatable by sulfa and declined less quickly after 1937).

- Untreated diseases control for the (non-linear) effects of omitted variables at the region-cohort level may otherwise load on to a_1.
- Approach resembles Case-Paxson 2009

1st generation: the impact of IMR at birth on outcomes for women in their 20s

	(1)	(2)	(3)	N
Mother Schooling (Yrs)	1.469***	3.174***	-0.316	29226
	(0.246)	(0.387)	(0.521)	
Mother Some College (=1)	-0.177***	-0.017	-0.124	29226
	(0.039)	(0.061)	(0.082)	
ln(Per Capita Income)	-0.179**	-0.327***	-0.368**	28048
	(0.075)	(0.118)	(0.159)	
Mother Height	-0.298	-0.350	-0.194	27634
	(0.319)	(0.505)	(0.682)	
Duncan SES	2.020	12.817***	-10.544**	29032
	(1.996)	(3.135)	(4.231)	
Logged Diarrheal/TB Mortality	No	Yes	Yes	
Maternal Birth Region Spec. Tre	No	No	Yes	

Each cell is the coeff on ln(IMR) in the mother's birth year from a separate regression

- The most general specification suggests birth-year disease conditions have adverse consequences for all outcomes, with (almost) significant effects for (the probability of college attendance), family income and Duncan's index.

- Effect sizes- A 10% drop in IMR is associated with a 3.4% rise in p.c. income (big) and an increase in the prob of college attendance of 0.012.

- The sign of the impact on years of schooling (not college) and Duncan's index is sensitive to controls for region-time varying variables (untreated diseases and linear region trends).

    - Differential trends in treated and untreated diseases
    - Other region-time developments (e.g Cutler and Miller 2007)
    - Selection e.g. compare years of school and prob(college).

Convergence/Divergence

$$y_{ijc} = a_0 + a_1 post_c * base_j + a_3 TB_{jc} + a_4 diar_{jc} + \lambda_t + u_i + e_{it}$$

This is the reduced form of a model in which IMR is instrumented with post*base. Untreated diseases still control for unobserved trends (jc)

Sulfa induced gains in p.c income and p(college) were decreasing in the pre-intervention infant mortality rate, $a_1 < 0$.

Years of schooling shows the opposite pattern

Second generation outcomes

The impact of infant mortality in the mother's birth year on cognitive and health outcomes of her births some 20-30 years after the arrival of sulfa

specification

$$y_{ijct} = a_0 + a_1 imr_{jc} + a_2 diar_{jc} + a_3 tb_{jc} + a_4 X_{ijc} + \gamma_c + \delta_j + \theta_t + e_{it}$$

Now y is the outcome for child i born in year t to a mother of cohort c born in region j. Child outcomes are observed at t, t+k (CPP is longitudinal, birth to age 7).

Controls now include fixed effects not only for mother's but also for child's cohort, age at interview, gender and gender*t.

We investigate models that do and do not include 1st generation (mother) outcomes in the controls, X. Estimates not sensitive to this.

estimates

	(1)	(2)	(3)	N
ln(Birth Weight)	-0.051	-0.143***	-0.150	32407
	(0.039)	(0.061)	(0.092)	
Low Birth Weight (=1)	0.090**	0.194***	0.207**	32407
	(0.037)	(0.059)	(0.089)	
Child Head Circumference	-0.370	-0.816	0.021	26918
	(0.417)	(0.659)	(1.000)	
Child Death in First Year (=1)	0.293	0.059*	0.021	32530
	(0.018)	(0.029)	(0.044)	
Bayley Mental Score (8 months)	0.140	-0.094	-2.07	26127
	(0.769)	(1.217)	(1.85)	
Bayley Motor Score (8 months)	-1.442**	-3.607***	-0.212	26130
	(0.602)	(0.952)	(1.446)	
Child Length Birth	-0.264	-1.035**	-0.521	30816
	(0.329)	(0.519)	(0.786)	
Child Height (Age 7)	0.962	2.106*	2.192	24992
	(0.748)	(1.188)	(1.798)	
Child IQ (Age 4)	-9.607***	-12.999***	-4.310	23091
	(1.837)	(2.932)	(4.435)	
Controls				
Log Diarrheal/TB Mortality _jc	No	Yes	Yes	
Linear Trends _jc	No	No	Yes	

findings

- Using only untreated diseases as specific but non-linear controls for relevant unobserved regional shocks, we identify adverse effects of mother's birth year IMR on a range of outcomes for her births-

- birth weight (above & below the lbw threshold)
- height at birth
- motor development at 8 months
- infant survival
- IQ at age 4.

We also find a "perverse" effect on child height at age 7, which we suggest is explained by selection (will investigate).

robustness to specification

- Removing controls for untreated disease mortality in general lowers the coefficients and t-statistics without changing the signs.

- Adding linear regional trends tends to raise the s.e. It lowers the coefficients for about half the outcomes.

- The only outcome for which statistical significance is preserved is low birth weight

- In urban regions, where the trend break is sharper, the estimates are relatively robust to regional trends

Differences by child gender

Boys show a response for all the outcomes indicated as significant in the model that pools boys and girls. The perverse coef on height at age 7 is insignificant for boys.

Weaker effects for girls-
- Birth weight (insig), tho similar coef on low birthweight
- Infant mortality (insig)

Stronger effects for girls-
- Motor development at 8 months
- Height at age 7 (the perversely signed coef)

Coefs on IQ, length at birth and mental development at 8 months are almost identical.

Convergence/Divergence

$$y_{ijct} = a_0 + a_1 post_c * base_j + a_2 TB_{jc} + a_3 diar_{jc} + + a_4 X_{ijct} + \gamma_c + \delta_j + \theta_t + e_{ijct}$$

As for 1st generation, this is a reduced form of model in which IMR is instrumented with post*base. Untreated diseases still control for unobserved trends (jc)

In general, the improvement in outcomes after 1937 is decreasing in pre-intervention mortality. This mirrors the short range analysis.

Coefficients (a_2) are now significant for a wider range of outcomes for boys (all previous excl length at birth but incl. head circumference) than for girls (for whom only bw and IQ are significant).

Extensions

Race differences

Sulfa diffused more rapidly amongst whites (JLS).

Short run and first generation outcomes show stronger trend breaks for whites. However second generation outcomes are stronger amongst blacks.

This is consistent with the impact of environmental conditions on the foetus being mediated by maternal health and/or access to health care at the birth of the child, both of which are lower for black women.

– e.g. Barker 1997, van den Berg et al. 2008, Bhalotra & Rawlings 2009, Kelly 2009.

Timing

We have looked for long run effects for women born in and after 1937, so we've defined post as 1937-1943.

- If it is shocks in the foetal rather than the birth year that matter then post would be 1938-43
- If shocks in early childhood matter, then post is 1936-43 (age 1),.., 1929-43 (age 4)

We investigated these alternative definitions.

The estimates point to birth year effects as most significant.

Further Analysis

- Simulation. Feed estimated short range impacts into long range analysis. Compare effects sizes for eg. lbw with previous studies. Region*year controls.
- Complement IMR analysis with analysis of cause-specific mortality
- Replace infant with child mortality

- Placebo design, alternative windows
- Interact Xs with post and with birth region
- In post*base-*IMR* models, including also base level of *outcome*.
- Clustered s.e. adjusted for small number of clusters

- Exploit panel of births to study persistence of effects, endogenous survival selection, birth spacing and fertility
- Parameterize and estimate selection
- Fathers
- Census data- similar and later ages

(skip this)

- What we refer to as 1st and 2nd generation outcomes are, in fact, 2nd and 3rd generation outcomes. This makes the findings more remarkable.

- Standard errors.

Thank you

Do email us with any further comments

DiD specifications for trend break

Same as previous specifications but now include control diseases.

$$m_{it} = b_0 + b_1 treated*post + b_2 treated*year + b_3 treated + b_4 year + b_5 state*post37 + u_i + e_{it}$$

$$m_{it} = a_0 + a_1 treated*post*year + a_2 treated*post + a_3 treated*year + a_4 treated + a_5 year + a_6 state*post37 + u_i + e_{it}$$

These models allow for different trends in treated and control disease mortality and test for a break in the level and trend of treated disease mortality.

The estimates confirm a break in trend.

First generation: post*treated

$$y_{ijc} = a_0 + a_1 post_c + a_3 TB_{jc} + a_4 diar_{jc} + a_5 year + u_i + e_{it}$$

$$y_{ijc} = b_0 + b_1 post_c + b_2 post_c*year + b_3 TB_{jc} + b_4 diar_{jc} + b_5 year + u_i + e_{it}$$

- Years of schooling and pr(college) show a significant step increase after 1937 (b_1>0) but a declining trend after 1937 (b_2<0)
- p.c. income shows a step increase, no change in trend (b_1>0, b_2=0)
- Height shows no step change and a declining trend (b=0, b_2<0)