Regression Discontinuity Design

Motivating example

- Many districts have summer school to help kids improve outcomes between grades
  - Enrichment, or
  - Assist those lagging
- Research question: does summer school improve outcomes
- Variables:
  - x=1 is summer school after grade g
  - y = test score in grade g+1

- Equation of interest
- $y_i = \beta_0 + x_i\beta_1 + \varepsilon_i$

- Problem: what do you anticipate is $\text{cov}(x_i, \varepsilon_i)$?

LUSDINE

- To be promoted to the next grade, students need to demonstrate proficiency in math and reading
  - Determined by test scores
- If the test scores are too low – mandatory summer school
- After summer school, re-take tests at the end of summer, if pass, then promoted

Situation

- Let Z be test score – Z is scaled such that
  - Z≥0 not enrolled in summer school
  - Z<0 enrolled in summer school
- Consider two kids
  - #1: Z=ε
  - #2: Z=-ε
  - Where ε is small

5

Intuitive understanding

- Participants in SS are very different
- However, at the margin, those just at Z=0 are virtually identical
- One with z=-ε is assigned to summer school, but z= ε is not
- Therefore, we should see two things

6

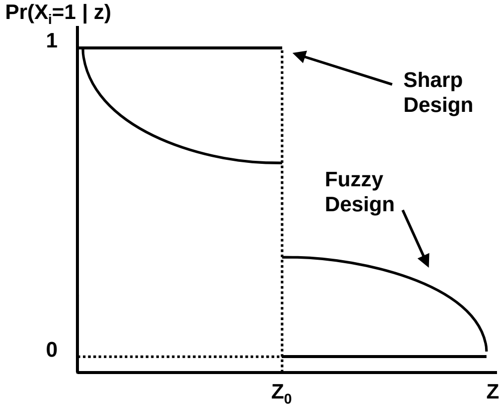
- There should be a noticeable jump in SS enrollment at z<0.
- If SS has an impact on test scores, we should see a jump in test scores at z<0 as well.

7

Variable Definitions

- y_i = outcome of interest
- x_i =1 if NOT in summer school, =1 if in
- $D_i = I(z_i \geq 0)$ -- I is indicator function that equals 1 when true, =0 otherwise
- z_i = running variable that determines eligibility for summer school. z is re-scaled so that z_i=0 for the lowest value where D_i=1
- w_i are other covariates

8



9

Initial equation

$$x_i = \theta_0 + D_i\theta_1 + h_f(z_i) + w_i\theta_2 + u_i$$

$$h_f(z_i) = polynomial\ in\ z$$

$$h_f(z_i) = 0\ at\ z = 0$$

10

$\hat{x}$ *just at* $z_i = 0$ *with summer school option*

$$\hat{x}_i^1 = \hat{\theta}_0 + \hat{\theta}_1 + w_i\hat{\theta}_2$$

$\hat{x}$ *just at* $z_i = 0$ *without summer school*

$$\hat{x}_i^0 = \hat{\theta}_0 + w_i\hat{\theta}_2$$

therefore

$$\hat{\theta}_i^1 - \hat{\theta}_i^0 = \hat{\Delta}_1$$

If $\hat{\Delta}_1 = 1$ *Sharp design*

If $\hat{\Delta}_1 < 1$ *fuzzy design*

11

RDD System

Structural equation:

$$y_i = \beta_0 + x_1\beta_1 + h(z_i) + w_i\beta_2 + \varepsilon_i$$

First stage:

$$x_i = \theta_0 + D_i\theta_1 + h_f(z_i) + w_i\theta_2 + u_i$$

reduced − form

$$y_i = \pi_0 + D_i\pi_1 + h_r(z_i) + w_i\pi_2 + v_i$$

Note that

$$\beta_1 = \pi_1 / \theta_1$$

12

RDD Equation

$\hat{y}$ *just at* $z_i = 0$ *with treatment*

$$\hat{y}_i^1 = \hat{\pi}_0 + \hat{\pi}_1 + w_i \hat{\pi}_2$$

$\hat{y}$ *just at* $z_i = 0$ *without treatment*

$$\hat{y}_i^0 = \hat{\pi}_0 + w_i \hat{\pi}_2$$

therefore

$$\hat{y}_i^1 - \hat{y}_i^0 = \hat{\pi}_1$$

13

Order of polynomial

$h(z_i) = polynomial\ in\ z$

$First\ order: h(z_i) = D_i z_i \gamma_1 + (1 - D_i) z_i \alpha_1$

$$Third\ order: h(z_i) = D_i z_i \gamma_1 + D_i z_i^2 \gamma_2 + D_i z_i^3 \gamma_3 \\ + (1 - D_i) z_i \alpha_1 + (1 - D_i) z_i^2 \alpha_2 + (1 - D_i) z_i^3 \alpha_3$$

14

Key assumption of RDD models

- People right above and below Z_0 are functionally identical
- Random variation puts someone above Z_0 and someone below
- However, this small different generates big differences in treatment (x)
- Therefore any difference in Y right at Z_0 is due to x

15

Limitation

- Treatment is identified for people at the z_i=0
- Therefore, model identifies the effect for people at that point
- Does not say whether outcomes change when the critical value is moved

16

Table 1

	Grade 3		
		Attended SS	
	Total	Yes	No
Outcomes			
2002 math score	641.8	620.4	648.5
	(.142)	(.241)	(.16)
	[36.57]		
2002 reading score	649.7	621.6	658.6
	(.176)	(.241)	(.204)
	[46.40]		
Summer school attendance			
Attended summer school 2001	.24	1	0
	(.002)	(0)	(0)
Days attended	4.373	18.208	0

17

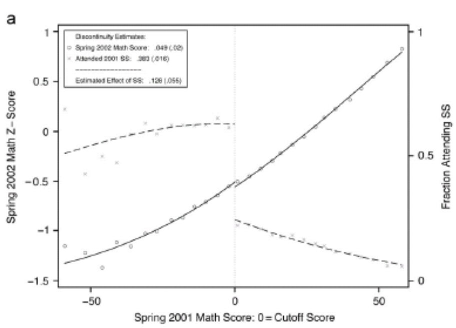
Table 2

	Effect of being mandated		Effect of SS attendance	
	Attendance (1st Stage)	Math (Reduced form)	Math (TSLS)	Reading (TSLS)
Strong 1st stage discontinuity				
Grade 3	.383	.049	.128	.087
	(.016)	(.02)	(.055)	(.065)
Grade 5	.385	.093	.241	.083
	(.006)	(.015)	(.039)	(.055)
Grade 6	.320	.061	.19	n.a.
	(.011)	(.014)	(.047)	(–)

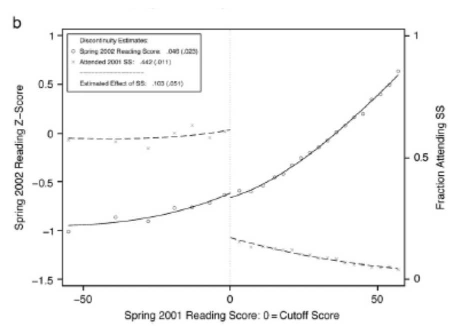
Math:
Grade 3: 0.049/0.383=0.128
Grade 5: 0.093/0.385=0.241
Grade 6: 0.061/0.320=0.190

18

J.D. Matsudaira / Journal of Econometrics 142 (2008) 829–850

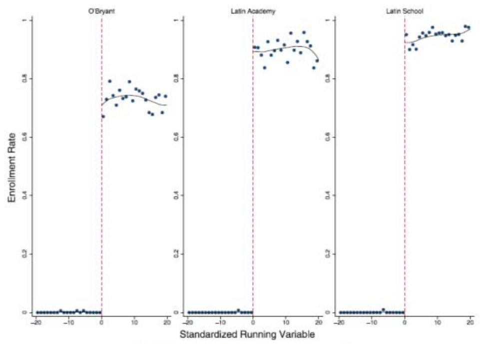


19



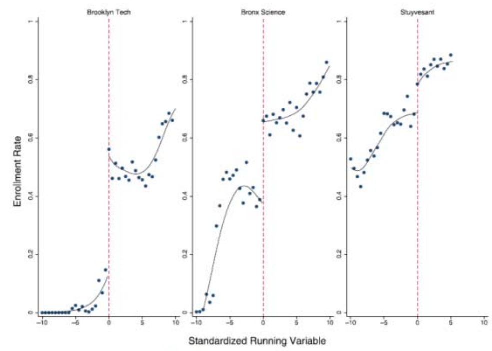
20

Example: Selective High Schools



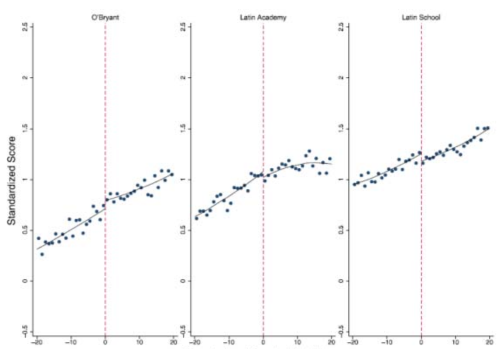
(b) Enrollment at each Boston exam school

21



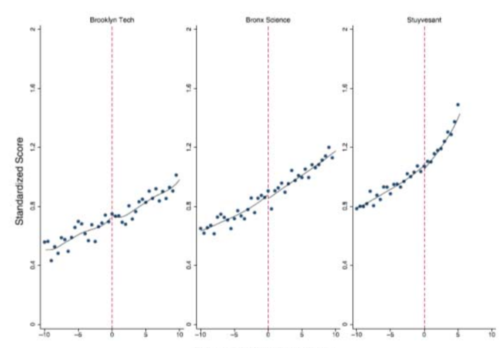
(b) Enrollment at any NYC exam school

22



(a) 10th grade math at Boston exam schools for 7th and 9th grade applicants

23



(a) Regents Advanced Math at NYC exam schools

24

Carpenter and Dobkin

Alcohol and Mortality

- Alcohol use
  - Reduces inhibition, increases aggression, compromises motor skills, blurres vision
- Use is associated with increased
  - Motor vehicle accidents, suicides, homicides, falls, burns, drowning
- Between 1975-95 Alcohol was involved in
  - 40% traffic deaths, 47% homicides, 30% suicides

Alcohol Abuse among Young Adults

- 4 million adults reported driving impaired in 2010
  - 112 million episodes
  - 81% due to men
  - Men aged 21-34 1/3 of all episodes
- Drunk driving deaths in 2012
  - 10,322 (1/3 of all traffic deaths)
  - In fatal crashes, 1/3 of drunk drivers are aged 21-24

Binge Drinking

- Definition
  - Men: 5+ drinks in a row one sitting
  - Women: 4+
- 30-day Prevalence by age
  - 18-24: 28.2%
  - All ages: 17.1%
- Frequency (among binge drinkers)
  - 18-24: 4.2 times
  - All ages: 4.4 times

- Intensity (max number of drinks among bingers)
  - 18-24: 9.3
  - All ages: 7.9

- Easy to establish
  - Pr(Drinking | MV death) >
    Pr(~ Drinking | MV Death)
- Much harder to establish
  ∂(MV Death)/∂(Alcohol use)
- What is required to identify this derivative?
  - A change in alcohol use
- Best option: variation in use generated by state policies

State alcohol control policies

- MLDA
- Price/taxes
- Retail sales restrictions
  - Date/time, Dram shop rules
- Drunk driving laws
  - BAC thresholds
  - Per se license revocation
  - Checkpoints
  - Mandatory minimum sentences

MLDA

- Used to vary across states
- In 1983, 35 states had MLDA<21
- National Minimum Drinking Age Act 1984
  - Passed July 17, 1984
  - Reduced federal highway funds for states by 10% if they had MLDA < 21
  - All states now have MLDA 21
  - US one of 4 countries with MLDA of 21

Previous research

- Difference in difference models
- 1983 law as the impetus
- MLDA <21 increases
  - Drinking, binge drinking, MV fatalities
  - MLDA 18 real problematic because it gets beer into high schools

33

This paper

- How does aging into drinking age impact use?
  - Estimated by RDD
  - sharp increase in use right at 21
- Given the change in use – is there a corresponding change in mortality outcomes

34

Nat. Health Interview Survey

- 1997-2005
- Random sample of US households
- Have date of birth and date of survey
- Measures drinking participation, heavy drinking over past week, month, year
  - Why is past-year drinking problematic for this question?
  - 71% use last month or week as reference period

35

Mortality detail files

- Annual data – authors use 1997-2005
- Contain census of deaths in the US (2.7 million/year)
- Variables: demographics, place, date, cause
- Restricted use data has date of birth
- Place people into months of age

36

Two groups of measures for alcohol use

- Participation
  - Any drinking in lifetime
  - 12 or more drinking in a year
  - Any heavy drinking past year
- Intensity
  - Proportion of days drinking
  - Proportion days heavy drinking
  - Drinks/day

37

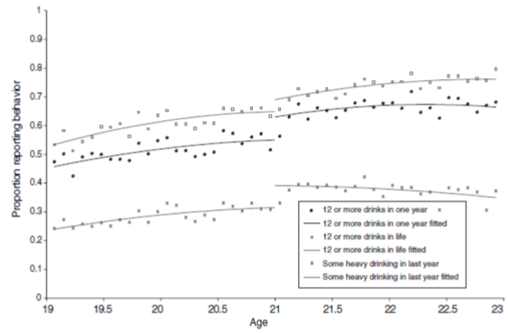


FIGURE 1. AGE PROFILE OF DRINKING PARTICIPATION

Notes: NHIS Sample Adult 1997–2005. Cells are the proportion of people in a 30-day block that report the behavior. The regression line is a second-order polynomial fitted on unweighted individual observations on either side of the age 21 cutoff.

38

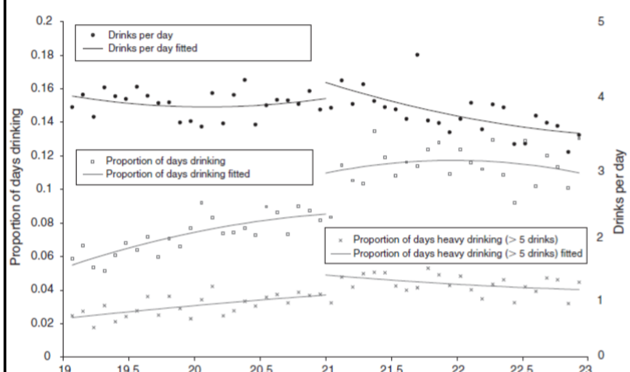


FIGURE 2. AGE PROFILE OF DRINKING INTENSITY

39

Table 1: Participation

	(1)
12 or more drinks in lifetime	
Over 21	0.0418
	(0.0242
Observations	16,107
R^2	0.02
Prob > Chi-Squared	
12 or more drinks in one year	
Over 21	0.0796
	(0.0254
Observations	16,107
R^2	0.02
Prob > Chi-Squared	
Any heavy drinking in last year	
Over 21	0.0761
	(0.0248
Observations	16,107
R^2	0.01
Prob > Chi-Squared	
Covariates	N
Weights	N
Quadratic terms	Y
Cubic terms	N
LLR	N

Table 2: Intensity

	(1)
Proportion of days drinking	
Over 21	0.0245
	(0.0086)
Observations	16,107
R^2	0.02
Prob > Chi-Squared	
Proportion of days heavy drinking	
Over 21	0.0120
	(0.0061)
Observations	15,825
R^2	0.00
Prob > Chi-Squared	
Drinks per day on days drinking	
Over 21	0.2387
	(0.2810)
Observations	9,906
R^2	0.00
Prob > Chi-Squared	
Covariates	N
Weights	N
Quadratic terms	Y
Cubic terms	N
LLR	N

40

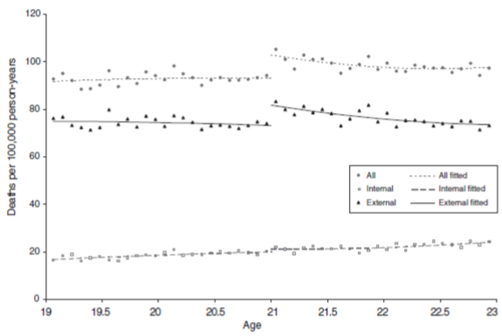


Figure 3. Age Profile for Death Rates

Notes: Deaths from the National Vital Statistics Records. Includes all deaths that occurred in the United States between 1997–2003. The population denominators are derived from the census. See online Appendix C for a list of causes of death.

Table 4—Discontinuity in Log Deaths at Age 21

	(1)	(2)	(3)	(4)
Deaths due to all causes				
Over 21	0.096	0.087	0.091	0.074
	(0.018)	(0.017)	(0.023)	(0.016)
Observations	1,460	1,460	1,460	1,458
R^2	0.04	0.05	0.05	
Prob > Chi-Squared		0.000	0.735	
Deaths due to external causes				
Over 21	0.110	0.100	0.096	0.082
	(0.022)	(0.021)	(0.028)	(0.021)
Observations	1,460	1,460	1,460	1,458
R^2	0.06	0.08	0.08	
Prob > Chi-Squared		0.000	0.788	
Deaths due to internal causes				
Over 21	0.063	0.054	0.094	0.066
	(0.040)	(0.040)	(0.053)	(0.031)
Observations	1,460	1,460	1,460	1,458
R^2	0.10	0.10	0.10	
Prob > Chi-Squared		0.000	0.525	
Covariates	N	Y	Y	N
Quadratic terms	Y	Y	Y	N
Cubic terms	N	N	Y	N
LLR	N	N	N	Y

Notes: See Notes from Table 1. The dependent variable is the log of the number of deaths that occurred x days from the person's twenty-first birthday. External deaths include all deaths with mention of an injury, alcohol use, or drug use. The Internal Death category includes all deaths Nt coded as external. Please see Web Appendix C for the ICD codes for each of the categories above. The first three columns give the estimates from polyNmial regressions on age interacted with a dummy for being over 21.

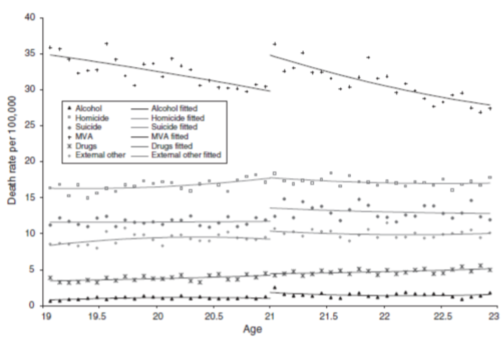


Figure 4. Age Profiles for Death Rates by External Cause

Notes: See notes to Figure 3. The categories are mutally exclusive. The order of precedence is homicide, suicide, MVA, deaths with a mention of alcohol, and deaths with a mention of drugs. The ICD-9 and ICD-10 Codes are in Appendix C.

43

Table 5—Discontinuity in Log Deaths at Age 21 Due to External Causes

	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Alcohol				Homicide			
Over 21	0.388	0.346	0.406	0.441	0.009	0.002	−0.003	−0.014
	(0.119)	(0.116)	(0.156)	(0.117)	(0.045)	(0.045)	(0.061)	(0.041)
Observations	1,460	1,460	1,460	1,458	1,460	1,460	1,460	1,458
R^2	0.03	0.04	0.04			0.01	0.01	0.01
Prob > Chi-Squared		0.000	0.228			0.000	0.495	
	Suicide				Motor vehicle accidents			
Over 21	0.160	0.154	0.135	0.105	0.158	0.143	0.145	0.139
	(0.059)	(0.059)	(0.086)	(0.045)	(0.033)	(0.032)	(0.044)	(0.032)
Observations	1,460	1,460	1,460	1,458	1,460	1,460	1,460	1,458
R^2	0.02	0.02	0.02	0.15	0.16	0.16		
Prob > Chi-Squared		0.000	0.892	0.000	0.666			
	Drugs				Other external causes			
Over 21	0.070	0.067	0.004	−0.016	0.087	0.098	0.098	0.074
	(0.081)	(0.082)	(0.107)	(0.078)	(0.060)	(0.059)	(0.075)	(0.043)
Observations	1,460	1,460	1,460	1,458	1,460	1,460	1,460	1,458
R^2	0.04	0.04	0.04	0.01	0.01	0.01		
Prob > Chi-Squared		0.000	0.643			0.000	0.877	
Covariates	N	Y	Y	N	N	Y	Y	N
Quadratic terms	Y	Y	Y	N	Y	Y	Y	N
Cubic terms	N	N	Y	N	N	N	Y	N
LLR	N	N	N	Y	N	N	N	Y

Notes: See Notes from Table 4. There are 276 observations where there are N deaths coded as due to alcohol; for this variable 0.5 was added to the dependent variable before taking the log. There are 15 observations where there are N deaths coded as due to drug use; for this variable 0.5 was added to the count before taking the log.

44

Summary – Table 5

Cause of death	Coefficient (std. error) on Over 21 dummy variable
Alcohol	0.388 (0.119)
Homicide	0.009 (0.045)
Suicide	0.160 (0.059)
MV accidents	0.158 (0.033)
Drugs	0.070 (0.081)
Other external causes	0.087 (0.060)

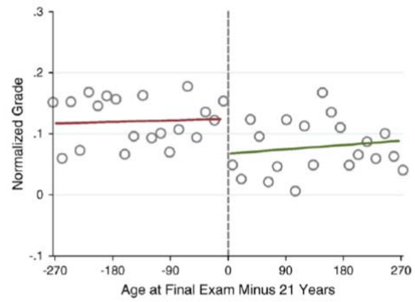


Fig. 3. Regression discontinuity estimates of the effect of drinking on achievement.

Estimating RDD models

- All states moved to MLDA 21 by 1988
- Use data on deaths among people with Social Security Numbers from 1989-2008
- Generate monthly counts of deaths by age/months – from age=19, month=0 through age=21, month=11
- 48 observations


```
Contains data from monthly_deaths.dta
  obs:            48
 vars:             4                          4 Aug 2015 09:35
 size:           240
--------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------
----
              storage   display    value
variable name   type    format     label      variable label
--------------------------------------------------------------------------------------
------------------------------------------------------------------------------
index           byte    %8.0g                 1-48, =1 for deaths at age 19, 0 months,
                                               =1 for deaths at 19, 1, ..48 for 22, 11
year            byte    %8.0g                 year of death
month           byte    %8.0g                 month of death
deaths          int     %8.0g                 annual deaths for monthly cohort
--------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------
Sorted by:
```

```
* generate ln death counts
gen deathsl=ln(deaths)


* rescale the running variable so that
* index = 0 in the month someone turns 21
gen rv=index-25

* treatment dummy
gen treatment=index>=25

* generate separate running variables before and
* after the discontinuity

gen rv_after1=treat*rv
gen rv_after2=rv_after1*rv_after1
gen rv_after3=rv_after2*rv_after1

gen rv_before1=(1-treat)*rv
gen rv_before2=rv_before1*rv_before1
gen rv_before3=rv_before2*rv_before1
```

49

```
. * basic RD -- linear running variables
. reg deathsl rv_after1 rv_before1 treat

      Source |       SS       df       MS              Number of obs =      48
-------------+------------------------------           F(  3,    44) =  461.28
       Model |  .631817937     3  .210605979           Prob > F      =  0.0000
    Residual |  .020088866    44  .000456565           R-squared     =  0.9692
-------------+------------------------------           Adj R-squared =  0.9671
       Total |  .651906803    47  .013870358           Root MSE      =  .02137

------------------------------------------------------------------------------
     deathsl |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   rv_after1 |   .0029281   .0006301     4.65   0.000     .0016582    .0041979
  rv_before1 |   .0072171   .0006301    11.45   0.000     .0059472    .0084869
   treatment |   .0925353   .0123525     7.49   0.000     .0676403    .1174302
       _cons |   7.861978   .0090032   873.25   0.000     7.843833    7.880122
```

50

```
.
. * basic RD -- cubic running variables
. reg deathsl rv_after* rv_before* treat

      Source |       SS       df       MS              Number of obs =      48
-------------+------------------------------           F(  7,    40) =  191.03
       Model |  .632972863     7  .090424695           Prob > F      =  0.0000
    Residual |   .01893394    40  .000473348           R-squared     =  0.9710
-------------+------------------------------           Adj R-squared =  0.9659
       Total |  .651906803    47  .013870358           Root MSE      =  .02176

------------------------------------------------------------------------------
     deathsl |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   rv_after1 |   .0055154   .0058914     0.94   0.355    -.0063915    .0174223
   rv_after2 |  -.0004198   .0006027    -0.70   0.490     -.001638    .0007983
   rv_after3 |   .0000146   .0000172     0.85   0.400    -.0000201    .0000494
  rv_before1 |   .0057035   .0071083     0.80   0.427     -.008663      .02007
  rv_before2 |   3.35e-06   .0006536     0.01   0.996    -.0013177    .0013244
  rv_before3 |   2.88e-06   .0000172     0.17   0.868    -.0000319    .0000377
   treatment |   .1007563   .0259471     3.88   0.000     .0483153    .1531974
       _cons |   7.853171    .020946   374.92   0.000     7.810837    7.895504
------------------------------------------------------------------------------
```

51

Medicare

- Introduced in 1963
- Federal health insurance programs for
  - the elderly
  - Disabled
- Among elderly – become eligible at age 65
- Two things happen at age 65
  - More become insured
  - Insurance is more generous

52

Medicare

- 2007
- 44.1 million recipients
- $432 bill. exp.
- 3.2% of GDP
- 16% of fed. budget

- 2040
- 87 million recipients
- 7.6% of GDP
- 30% of fed. budget

53

This paper

- Change in eligibility at age 65
- We should see
  - Greater levels of insurance
  - Greater use of medical services
- If health insurance improves health, we should also see a reduction in mortality

54

Sample

- CA hospital admissions 1992-2002
- Restrict sample to those admitted through emergency department
  - e.g., Chronic bronchitis, heart attack, stroke
  - Why?

55

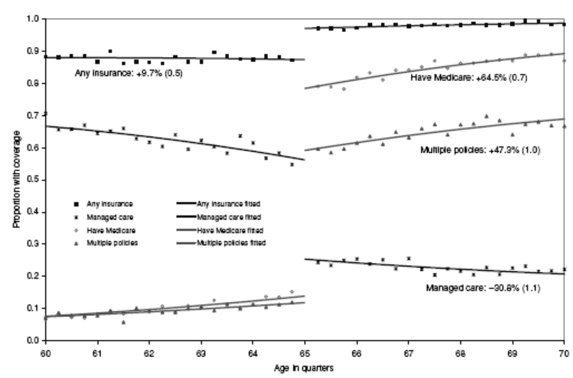


FIGURE I

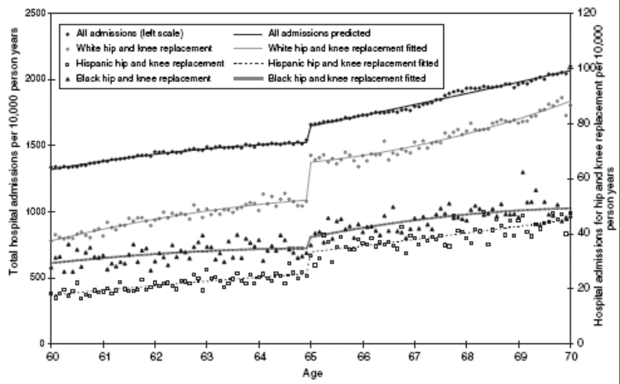


FIGURE 3. HOSPITAL ADMISSION RATES BY RACE/ETHNICITY

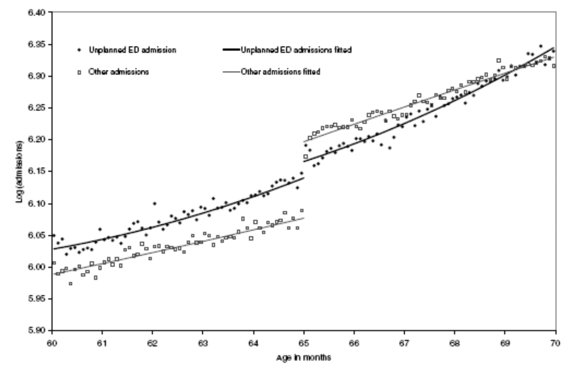


FIGURE II
Number of Admissions by Route into Hospital, California, 1992–2002

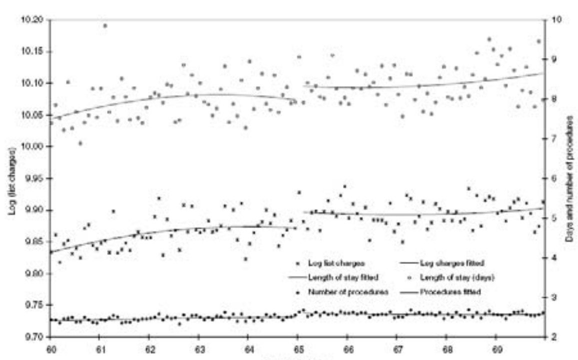


FIGURE V
Three Measures of Inpatient Treatment Intensity

TABLE II
REGRESSION DISCONTINUITY MODELS FOR LOG (NUMBER OF ADMISSIONS) TO CALIFORNIA HOSPITALS BY AGE OF PATIENT AT ADMISSION

	Non-ED or planned		ED and unplanned		Weekend t-stat > 6.62	
	(1)	(2)	(3)	(4)	(5)	(6)
Age over 65 (×100)	11.9	12.0	2.4	2.6	3.2	3.3
	(0.5)	(0.5)	(0.5)	(0.5)	(1.0)	(1.1)
Dummy for just under 65	No	Yes	No	Yes	No	Yes

TABLE III
REGRESSION DISCONTINUITY MODELS FOR PROBABILITY OF DIFFERENT FORMS OF PRIMARY INSURANCE COVERAGE

	Medicare		Private		Medicaid		Uninsured	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age over 65 (×100)	43.9	47.5	−24.8	−26.8	−10.1	−10.8	−7.4	−8.0
	(0.4)	(0.4)	(0.4)	(0.4)	(0.3)	(0.3)	(0.2)	(0.2)
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes
Mean of dependent variable for patients aged 64–65 (×100)	24.0		43.3		43.3		9.7	

TABLE IV
REGRESSION DISCONTINUITY MODELS FOR CHANGES IN TREATMENT INTENSITY

	Length of stay (days)		Number of procedures		Log list charges (×100)	
	(1)	(2)	(3)	(4)	(5)	(6)
Age over 65	0.37	0.35	0.09	0.11	2.5	2.6
	(0.24)	(0.26)	(0.03)	(0.03)	(1.1)	(1.0)
Additional controls	No	Yes	No	Yes	No	Yes
Mean of dependent variable for patients aged 63 or 64	8.12		2.50		9.87	

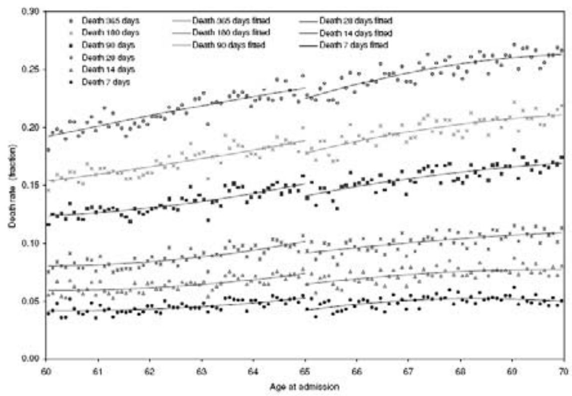


FIGURE VI
Patient Mortality Rates over Different Follow-Up Intervals

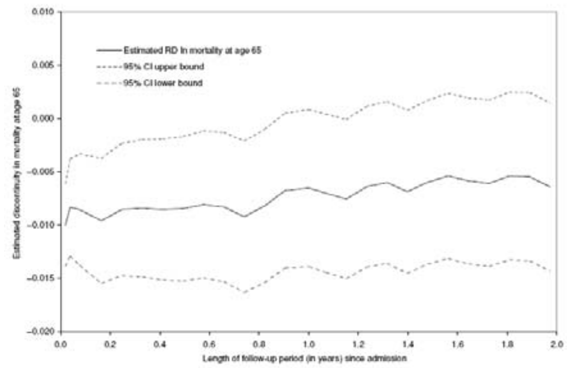


FIGURE VII
Estimates of the Discontinuity in Mortality Rates at Age 65 over Various Follow-Up Periods

62

TABLE V
REGRESSION DISCONTINUITY ESTIMATES OF CHANGES IN MORTALITY RATES

	Death rate in					
	7 days	14 days	28 days	90 days	180 days	365 days
	Estimated discontinuity at age 65 (×100)					
Fully interacted quadratic with no additional controls	−1.1 (0.2)	−1.0 (0.2)	−1.1 (0.3)	−1.1 (0.3)	−1.2 (0.4)	−1.0 (0.4)
Fully interacted quadratic plus additional controls	−1.0 (0.2)	−0.8 (0.2)	−0.9 (0.3)	−0.9 (0.3)	−0.8 (0.3)	−0.7 (0.4)
Fully interacted cubic plus additional controls	−0.7 (0.3)	−0.7 (0.2)	−0.6 (0.4)	−0.9 (0.4)	−0.9 (0.5)	−0.4 (0.5)
Local linear regression procedure fit separately to left and right with rule-of-thumb bandwidths	−0.8 (0.2)	−0.8 (0.2)	−0.8 (0.2)	−0.9 (0.2)	−1.1 (0.3)	−0.8 (0.3)
Mean of dependent variable (%)	5.1	7.1	9.8	14.7	18.4	23.0

63

Elder and Lubotsky

64

The downside of being the youngest in your class

- Suggestive evidence that children "young" for their class perform worse in school
  - Lower test scores/more repeated grades/more disciplinary problems/more ADHD diagnoses
- This has lead to two trends
  - Academic "red shirting"
  - States have moved the "age of entry" earlier
    - 1980, 10% of 5 years olds not in k-garten
    - 2002, this number was 21%

65

- Suppose all schools start September 1
- Consider the youngest possible kid in the class
- Three state laws – to start k-garten, a kid must turn 5 by: December 1, September 1 or June 1
- In these three states, at school start, the ages of the youngest kids in class are
  - 4 years, 9 months at start (12/1)
  - 5 years (9/1)
  - 5 years, 3 months (6/1)

66

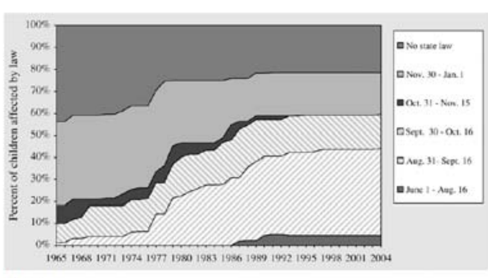


Figure 1
State Entrance Age Cutoffs, 1965-2005
Source: Authors' calculations from state kindergarten statutes. States are weighted by the number of children four to six years old in the state in 1990.

67

State	Cutoff 2005	Law changes since 1984
AL	1-Sep	1984–1989: 10/1 1990+: 9/1
AK	1-Sep	1984–1987: 11/2 1988–2003: 8/15
AZ	31-Aug*	
AR	15-Sep	1984–1997: 10/1 1998: 9/1 1999+: 9/15
CA	2-Dec	1984–1986: 12/1 1987+: 12/2
CO	LEA	
CT	1-Jan	
DE	31-Aug	1984–1992: 12/31 1993: 11/30 1994: 10/31 1995: 9/30 1996+: 8/31
DC	31-Dec	
FL	1-Sep	
GA	1-Sep	Established 1985
HI	31-Dec	
ID	1-Sep	1984–1989: 10/16 1990: 9/16 1991–1992: 8/16 1993+: 9/1
IL	1-Sep	1984–1985: 12/1 1986: 11/1 1987: 10/1 1988+: 9/1
IN	1-Jul	1984–1988: LEA 1989: 9/1 1990: 8/1 1991: 7/1 1992–2000: 6/1 2001–2005: 7/1
IA	15-Sep	
KS	31-Aug	1984–1994: 9/1 1995+: 8/31
KY	1-Oct	
LA	30-Sep	1984–1995: 12/31 1996+: 9/30
ME	15-Oct	

State	Cutoff 2005	Law changes since 1984
MD	30-Sep	1984–2002: 12/31 2003: 11/30 2004: 10/31 2005: 9/30 2006+: 9/1
MA	LEA	
MI	1-Dec	
MN	1-Sep	
MS	1-Sep	
MO	31-Jul*	1984–1986: 8/31* 1987: 7/31* 1988–1996: 6/30* 1997+: 7/31*
MT	10-Sep	
NE	15-Oct	
NV	30-Sep	
NH	LEA	
NJ	LEA	
NM	31-Aug*	
NY	LEA	
NC	16-Oct	
ND	31-Aug*	
OH	30-Sep	
OK	1-Sep	
OR	1-Sep	1984–1985: 11/15
PA	LEA	
RI	1-Sep	1984–2003: 12/31
SC	1-Sep	1984–1992: 11/1
SD	1-Sep	
TN	30-Sep	1984: 10/31
TX	1-Sep	1984–1994: ssy 1995+: 9/1
UT	1-Sep*	1984–1987: ssy 1988+: 9/1*
VT	LEA	1984–1990: 1/1 1991+: LEA
VA	30-Sep	
WA	31-Aug	
WV	31-Aug*	
WI	1-Sep	
WY	15-Sep	

68

Evidence to date

- Most of the evidence on the problems of being the youngest in your class is regression-based
- Outcome is regressed on age of child
- Control for other covariates

69

- Consider a regression
- y_i = some measure of outcomes (test score)
- EA_i =entrance age (age you enter k-garten)
- $y_i=\beta_0 + EA_i\beta_1 + w_i\beta_2 + \varepsilon_i$
- Is the estimate for β_1 unbiased?
  - Can be biased up for down

70

Research strategy

- Suppose a state has a September 1 cutoff
- Consider two kids
  - One born August 31
  - One born September 2nd
- One average – do we expect these kids to differ systematically?
- Yet – they will differ when they start school
  - August 31st birth will start at age 5
  - September 2nd birth will start at age 6

71

- Look on either side of cutoff date
- Should see a large change in age at school entry
- If this impacts outcomes, should see change in test scores at the cutoff as well
- Is the assumption that kids born 3 days apart a good assumption?

72

Early Childhood Longitudinal Study

- 20 kids from each of 1,000 schools
- Kindergarten class of 1988/89
- Students re-sampled in 1st, 3rd, 5th grade
- Obtain detailed information about the kids/parents/schools/teachers

- Structural equation
  - $y_i=\beta_0 + EA_i\beta_1 + w_i\beta_2 + \varepsilon_i$
  - EA is entry age
- First stage
  - $EA_i = \theta_0 + PEA_i\theta_1 + w_i\theta_2 + \upsilon_i$
  - PEA = predicted entry age – age you would be at the start of kindergarten if you followed the state law to the letter

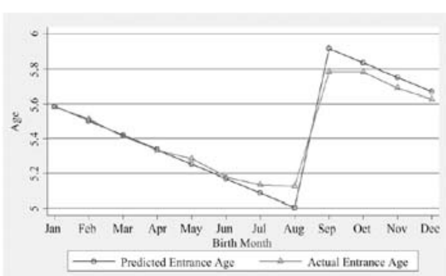


Figure 2
Average Predicted and Actual Entrance Ages by Birth Month in States with September 1 Cutoffs, ECLS-K

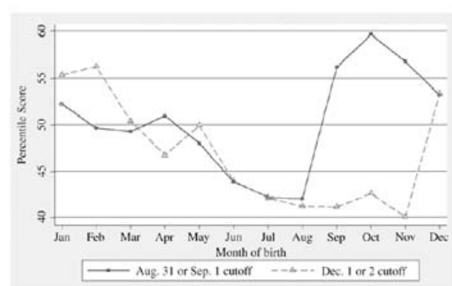
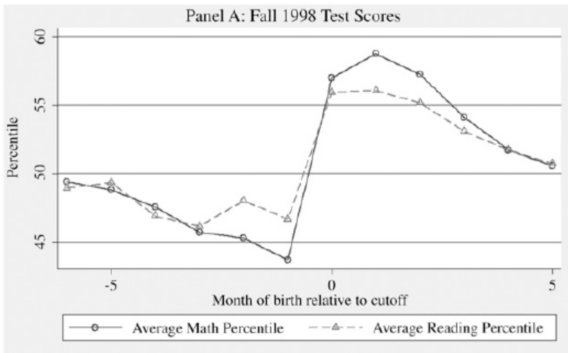
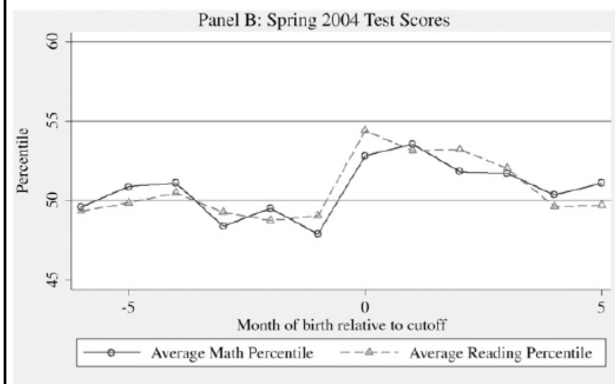


Figure 3
Fall Kindergarten Math Scores by Month of Birth, ECLS-K



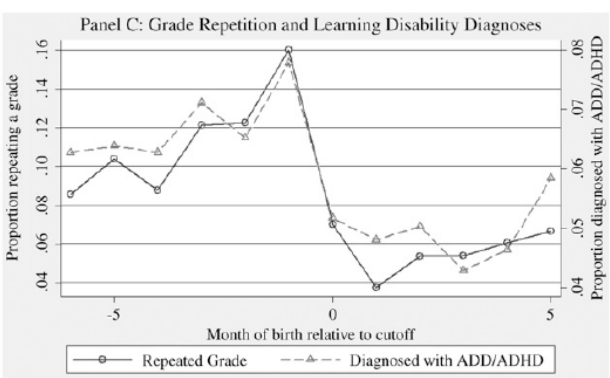
Test in fall of kindergarten

77



Test in spring of 5th grade

78



79

Table 1
Estimates of the Effect of Kindergarten Entrance Age on Reading Test Scores

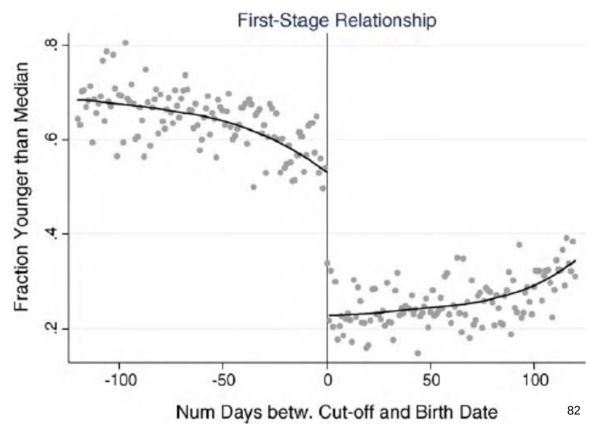
	Mean of IRT test score	Models of IRT test scores by estimation method				Test score percentile
Test date	S.D. N	OLS (1)	OLS (2)	IV (3)	IV (4)	IV (5)
ECLS-K						
Fall 1998 (Kindergarten)	27.5 10.0 11,592	3.79 (0.31) 0.018	3.69 (0.29) 0.212	4.15 (0.49) 0.018	5.28 (0.47) 0.209	16.68 (1.28) 0.248
Spring 1999 (Kindergarten)	38.9 13.4 11,975	5.07 (0.40) 0.018	5.05 (0.39) 0.192	6.20 (0.64) 0.017	8.17 (0.62) 0.187	19.33 (1.33) 0.211
Spring 2000 (First grade)	68.0 20.7 12,046	7.60 (0.59) 0.017	7.17 (0.55) 0.219	8.11 (0.95) 0.017	10.67 (0.89) 0.216	14.08 (1.22) 0.213
Spring 2002 (Third grade)	107.5 20.2 10,336	7.09 (0.72) 0.016	5.26 (0.60) 0.285	6.54 (1.03) 0.016	7.41 (0.88) 0.284	11.08 (1.27) 0.285
Spring 2004 (Fifth grade)	139.4 23.2 8,210	7.44 (0.86) 0.013	5.64 (0.73) 0.286	6.69 (1.27) 0.013	8.38 (1.09) 0.284	10.59 (1.33) 0.280

80

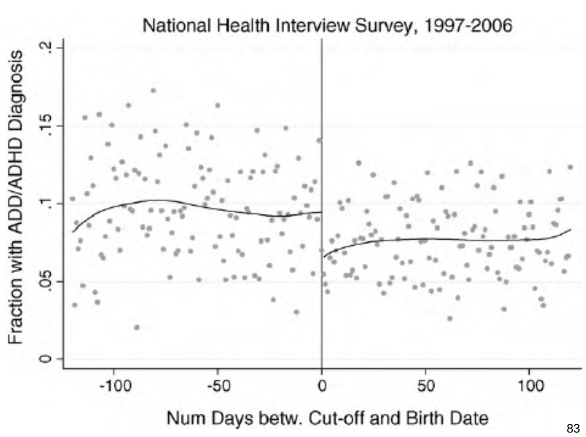
Table 5
The Effect of Kindergarten Entrance Age on Grade Retention and Learning Disabilities in the Full NELS:88 and ECLS-K Samples and by Family Background Quartile

Dependent Variable	Mean N	OLS (1)	OLS (2)	IV (3)	IV (4)
ECLS-K					
Diagnosis of learning disability/ADD/ADHD/etc.	0.088 12,860	0.008 (0.008)	0.005 (0.009)	−0.026 (0.011)	−0.025 (0.012)
Diagnosis of ADD/ADHD	0.043 12,860	−0.004 (0.006)	−0.011 (0.006)	−0.021 (0.007)	−0.029 (0.009)
Diagnosis of non-ADD/ADHD learning disability	0.045 12,860	0.012 (0.005)	0.014 (0.005)	−0.004 (0.007)	0.001 (0.008)
In 1st or 2nd grade in Spring, 2002	0.088 10,431	−0.112 (0.010)	−0.112 (0.011)	−0.116 (0.013)	−0.131 (0.015)

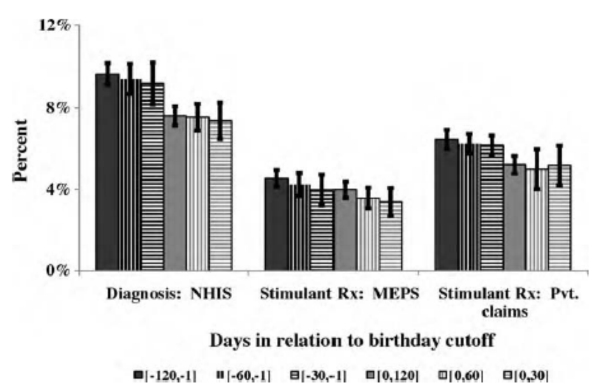
81



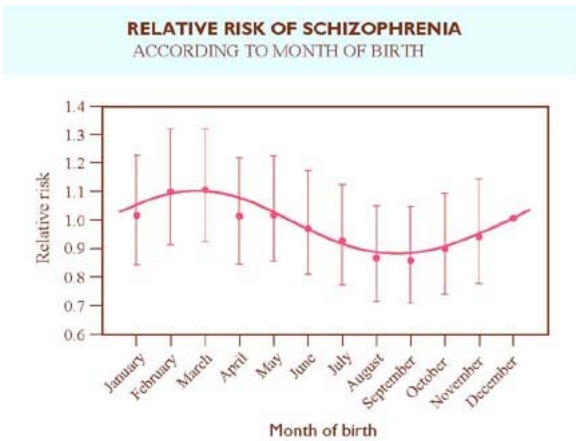
82

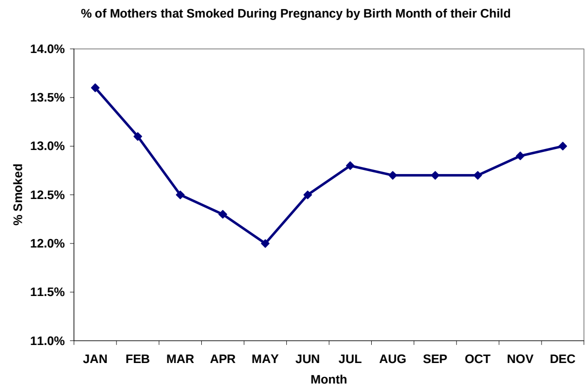


83

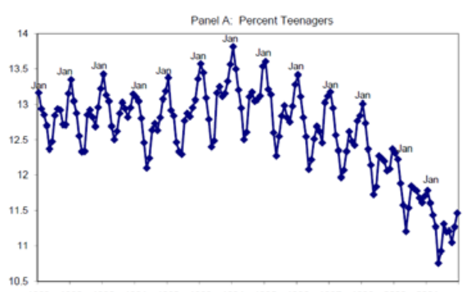


84

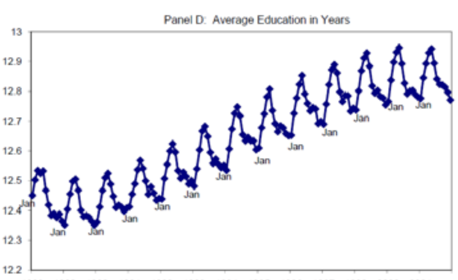

RELATIVE RISK OF SCHIZOPHRENIA
ACCORDING TO MONTH OF BIRTH
Relative risk
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
January
February
March
April
May
June
July
August
September
October
November
December
Month of birth


% of Mothers that Smoked During Pregnancy by Birth Month of their Child
% Smoked
14.0%
13.5%
13.0%
12.5%
12.0%
11.5%
11.0%
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
Month

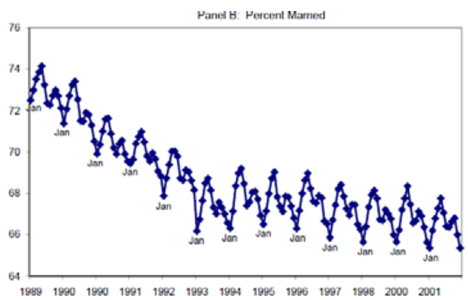
86


Panel A: Percent Teenagers
14
13.5
13
12.5
12
11.5
11
10.5
1989
1990
1990
1991
1992
1993
1994
1995
1996
1997
1998
2000
2001

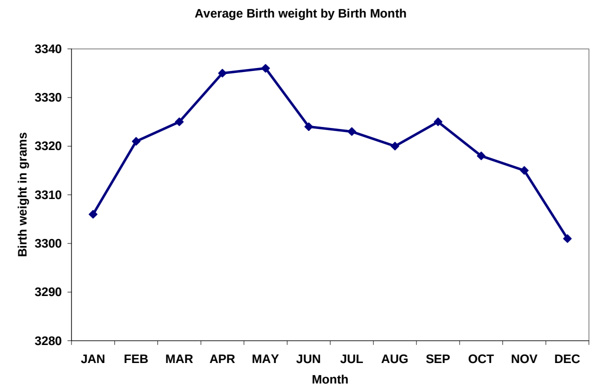
87


Panel D: Average Education in Years
13
12.9
12.8
12.7
12.6
12.5
12.4
12.3
12.2
1989
1990
1990
1991
1992
1993
1994
1995
1996
1997
1998
2000
2001

88


Panel B: Percent Married
76
74
72
70
68
66
64
1989 1990 1990 1991 1992 1993 1994 1995 1996 1997 1998 2000 2001
Jan

89


Average Birth weight by Birth Month
Birth weight in grams
3340
3330
3320
3310
3300
3290
3280
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
Month

90