Long literature on effects of poor neonatal health on long-term outcomes

- There is good evidence that early life circumstances have long-term effects on human capital accumulation
  - Fetal origins hypothesis – Barker (1995)
- In utero stress & disease:
  - Almond (2006); Almond & Currie (2011); Almond, Edlund & Palme (2009); Camacho (2008); Chay and Greenstone (2003)
- Long-term effects (wages, disability, years of education) of low birth weight as a signal of poor neonatal health:
  - Behrman & Rosenzweig (2004); Black, Devereux, Salvanes (2007); Oreopoulous et al. (2008); Royer (2009)
- Most compelling causal studies make use of twin comparisons
Adult outcome literature provides no guidance regarding potential policy levers

- There is a “permanent” effect of poor neonatal health, but we don’t know the potential pathways through which these effects come into being
  - Does the effect of poor neonatal health on cognitive development vary at different ages?
  - Can school quality help mitigate these effects
- All existing research come from ethnically and economically homogeneous societies
  - Are results different for different racial, ethnic, or socio-economic groups?
  - Can we learn whether early health and parental inputs are substitutes or complements?
- One more thing: Adult outcome studies necessarily make use of older data before modern neonatology advances
  - Mean birth weights in earlier studies (births 1930s-1970s) ranged from 2517-2598 grams, versus 2409 grams in our study (births 1992-2002)
Reasons for holes in existing literature

- Existing registry databases (e.g., Denmark, Germany, Norway, Sweden) don’t include much, if any, information on test scores or other measures of cognitive development.

- Those countries (Chile, China, Taiwan) where there have been some links between birth and school records are ethnically and economically relatively homogeneous → very little work done to study heterogeneity.
A new data resource: Florida registry data

- We make use of the first, to our knowledge, large-scale dataset that
  - Links birth records to school records in a western industrialized context
  - Includes annual assessment data for children to track children’s trajectories over time – important for observing whether birth weight effects open or close over time so that we might be able to pinpoint resources
  - To date, children born from 1992-2002 matched to school records

- Florida is a location with many desirable characteristics for study:
  - **Large**: Florida’s population of ~17M and ~200K births/year compares to Norway, Denmark, and Sweden combined
  - **Heterogeneous**: 45% of moms racial/ethnic minorities; 25% of moms foreign born
  - The median voter in the United States lives there
  - Excellent institutional conditions for matching birth and school data
What we do

- Twin-pair comparisons (>14k twin pairs old enough to have third-grade test scores so far)
- Estimate twin fixed effects models to measure the effects of birth weight on
  - Test scores in grades 3-8
  - Kindergarten readiness measures (pre-literacy skills at age 5)

Our contributions

- First comprehensive exploration of effects of birth weight over the schooling career in a western industrialized context
- First opportunity anywhere to study a wide range of potential heterogeneous effects
- First study anywhere to measure the role of school quality in remediating birth weight disadvantage
The Florida data

- Only observe school history in Florida if a child
  - Remains in Florida until school age
  - Attends a Florida public school
  - Is successfully matched between birth and school records

- How good is the match?
  - Match based on name (with some fuzziness), date of birth, and social security number
  - American Community Survey: 80.9% of children born in Florida live in Florida at age 5 and attend public school – this is an overstatement
  - Our match: 79.6% of all births (79.5% for twins)
  - Therefore, nearly all potentially matchable children are matched
Attributes of all Florida births and Florida-born twins attending Florida public schools

<table>
<thead>
<tr>
<th>Maternal attribute</th>
<th>Full population of births</th>
<th>Population of kids matched to Florida school records</th>
<th>Population of kids with a third-grade test score</th>
<th>Population of twins with a third grade test score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>22.6</td>
<td>24.8</td>
<td>25.7</td>
<td>25.9</td>
</tr>
<tr>
<td>Hispanic</td>
<td>23.0</td>
<td>23.3</td>
<td>23.8</td>
<td>18.0</td>
</tr>
<tr>
<td>Foreign-born</td>
<td>23.5</td>
<td>22.9</td>
<td>23.1</td>
<td>18.0</td>
</tr>
<tr>
<td>Married at time of birth</td>
<td>64.8</td>
<td>62.2</td>
<td>61.0</td>
<td>68.3</td>
</tr>
<tr>
<td>High school dropout</td>
<td>20.9</td>
<td>22.5</td>
<td>23.4</td>
<td>15.5</td>
</tr>
<tr>
<td>College graduate</td>
<td>20.5</td>
<td>17.5</td>
<td>15.8</td>
<td>23.0</td>
</tr>
<tr>
<td>Age 21 or below</td>
<td>22.0</td>
<td>23.6</td>
<td>24.2</td>
<td>14.4</td>
</tr>
<tr>
<td>Age 36 or above</td>
<td>9.8</td>
<td>9.4</td>
<td>9.1</td>
<td>13.6</td>
</tr>
</tbody>
</table>
Distribution of birth weight among twins
Distribution of twin birth weight discordance
Some checks of internal validity in twin FE models

<table>
<thead>
<tr>
<th></th>
<th>Appears in public school</th>
<th>In public school since first grade</th>
<th>Test scores observed</th>
<th>Test scores observed every possible year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of dependent variable</td>
<td>0.795</td>
<td>0.770</td>
<td>0.691</td>
<td>0.655</td>
</tr>
<tr>
<td>Coefficient on log birth weight</td>
<td>-0.012 (0.008)</td>
<td>-0.006 (0.008)</td>
<td>-0.003 (0.009)</td>
<td>0.006 (0.004)</td>
</tr>
</tbody>
</table>

No evidence that heavier or lighter twins are more likely to be observed in regressions.
Student outcomes in Florida

- Since 1998, Florida tests students on the criterion-referenced Florida Comprehensive Assessment Test in reading and mathematics.
- Initially tested in grades 4, 8, 10 in reading; 5, 8, 10 in math; from 2001 grades 3-10 in reading and math.
- Nearly universal testing
  - Students with some disabilities are not tested.
  - Though there are several makeup dates, it is possible to miss the test if a student is absent for a long period of time.
  - Therefore, it is important to see whether there are differential rates of missing the test.
- For ease of interpretation, we standardize scores at the state-by-grade level.
  - Average performance in matched twins sample is a little higher than state average, due to higher SES of families with twins, and the fact that those remaining in Florida from birth through school are more stable.
Heavier twins consistently score better (Dep var: state-standardized math score)
Differences are constant over the course of school

![Graph showing differences in means of test scores in mathematics over imputed grades.](image)
Same is true for reading
Combined math+reading test score gap

Difference in means of combined test scores

Difference

Imputed grade

Difference

95% confidence interval

3 4 5 6 7 8

0

0.025

0.05

0.075

0.1
Results are identical if restrict to a balanced panel
Differences are not due to differential selection
Differences are not due to differential selection
Sensitivity to selection: Impute missing test scores with 5\textsuperscript{th} or 95\textsuperscript{th} percentile score
Differences are large, but not as large as associated with mom’s education.
Empirical specification

- Following ACL, BDS:
  \[ y_{ijk} = \alpha + \beta bw_{ijk} + x_{jk} \gamma + f_{jk} + \epsilon_{ijk} \]

- Differencing within twin-pairs gives you:
  \[ y_{1jk} - y_{2jk} = \beta \left( bw_{1jk} - bw_{2jk} \right) + \left( \epsilon_{1jk} - \epsilon_{2jk} \right) \]

- In practice, we regress test scores on log(birth weight), twin-pair fixed effects, gender and birth order dummies
Empirical specification

- Twin pair fixed effects: We regress test scores on log(birth weight), twin-pair fixed effects, gender and within-pair birth order dummies
- Dependent variable: FCAT
  - Mean of math & reading, or either if one missing
- Regression sample is twin pairs where we observe both twins in a given grade
- We define “imputed grade” as the grade a student would be in if she progressed consistently after we first see her in 3rd grade
- For regressions pooled across grades, cluster standard errors at the individual level
Non-parametric relationship between birth weight and test scores

Non-parametric estimates of the effects of birth weight on cognitive development by birth weight bins

Pooled standardized test scores from grades 3-8
Mathematics and reading combined

Birth weight

- twin
- twin FE
(Note: The patterns are the same for singletons!)
### Effects of log birth weight on pooled average test score (twin FE models)

<table>
<thead>
<tr>
<th>Population</th>
<th>Percent</th>
<th>Mean score</th>
<th>Mean (sd) birth weight (g)</th>
<th>Pooled estimate OLS (SE)</th>
<th>Pooled estimate FE (SE)</th>
<th>P-value of difference in FE</th>
</tr>
</thead>
<tbody>
<tr>
<td>All twins</td>
<td>100%</td>
<td>0.074</td>
<td>2420 (565)</td>
<td>0.310 (0.019)</td>
<td>0.441 (0.029)</td>
<td>n/a</td>
</tr>
<tr>
<td>Same sex</td>
<td>68.2</td>
<td>0.073</td>
<td>2405 (568)</td>
<td>0.315 (0.022)</td>
<td>0.447 (0.032)</td>
<td>0.773</td>
</tr>
<tr>
<td>Opposite sex</td>
<td>31.8</td>
<td>0.076</td>
<td>2454 (557)</td>
<td>0.312 (0.035)</td>
<td>0.427 (0.062)</td>
<td>0.223</td>
</tr>
<tr>
<td>Mom white</td>
<td>72.0</td>
<td>0.256</td>
<td>2457 (554)</td>
<td>0.230 (0.022)</td>
<td>0.466 (0.034)</td>
<td></td>
</tr>
<tr>
<td>Mom black</td>
<td>26.1</td>
<td>-0.466</td>
<td>2318 (585)</td>
<td>0.283 (0.033)</td>
<td>0.381 (0.061)</td>
<td>0.518</td>
</tr>
<tr>
<td>Mom non-Hisp</td>
<td>82.0</td>
<td>0.098</td>
<td>2413 (565)</td>
<td>0.313 (0.021)</td>
<td>0.434 (0.033)</td>
<td></td>
</tr>
<tr>
<td>Mom Hispanic</td>
<td>18.0</td>
<td>-0.036</td>
<td>2454 (564)</td>
<td>0.308 (0.044)</td>
<td>0.478 (0.059)</td>
<td></td>
</tr>
<tr>
<td>Non-immigrant</td>
<td>82.0</td>
<td>0.072</td>
<td>2413 (564)</td>
<td>0.324 (0.021)</td>
<td>0.440 (0.033)</td>
<td>0.899</td>
</tr>
<tr>
<td>Immigrant</td>
<td>18.0</td>
<td>0.080</td>
<td>2451 (570)</td>
<td>0.232 (0.043)</td>
<td>0.449 (0.058)</td>
<td></td>
</tr>
<tr>
<td>Mom unmarried</td>
<td>31.8</td>
<td>-0.360</td>
<td>2336 (574)</td>
<td>0.310 (0.033)</td>
<td>0.362 (0.057)</td>
<td>0.064</td>
</tr>
<tr>
<td>Mom married</td>
<td>67.6</td>
<td>0.272</td>
<td>2458 (556)</td>
<td>0.271 (0.023)</td>
<td>0.485 (0.033)</td>
<td></td>
</tr>
</tbody>
</table>
Effects of log birth weight on pooled average test score (twin FE models)

<table>
<thead>
<tr>
<th>Population</th>
<th>Percent</th>
<th>Mean score</th>
<th>Mean (sd) birth weight (g)</th>
<th>Pooled estimate OLS (SE)</th>
<th>Pooled estimate FE (SE)</th>
<th>P-value of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS dropout</td>
<td>15.8</td>
<td>-0.476</td>
<td>2338 (570)</td>
<td>0.224 (0.047)</td>
<td>0.359 (0.070)</td>
<td>0.163</td>
</tr>
<tr>
<td>HS grad</td>
<td>61.4</td>
<td>0.003</td>
<td>2430 (563)</td>
<td>0.328 (0.025)</td>
<td>0.434 (0.038)</td>
<td></td>
</tr>
<tr>
<td>College grad</td>
<td>22.8</td>
<td>0.663</td>
<td>2451 (562)</td>
<td>0.310 (0.038)</td>
<td>0.529 (0.059)</td>
<td></td>
</tr>
<tr>
<td>Mom &lt;=21</td>
<td>14.7</td>
<td>-0.396</td>
<td>2269 (574)</td>
<td>0.253 (0.046)</td>
<td>0.372 (0.086)</td>
<td></td>
</tr>
<tr>
<td>Mom 22-29</td>
<td>40.2</td>
<td>-0.006</td>
<td>2419 (561)</td>
<td>0.309 (0.029)</td>
<td>0.443 (0.044)</td>
<td>0.698</td>
</tr>
<tr>
<td>Mom 30-35</td>
<td>31.6</td>
<td>0.277</td>
<td>2465 (557)</td>
<td>0.319 (0.035)</td>
<td>0.483 (0.052)</td>
<td></td>
</tr>
<tr>
<td>Mom &gt;=36</td>
<td>13.5</td>
<td>0.343</td>
<td>2479 (559)</td>
<td>0.345 (0.058)</td>
<td>0.413 (0.078)</td>
<td></td>
</tr>
<tr>
<td>Low income</td>
<td>29.8</td>
<td>-0.216</td>
<td>2393 (567)</td>
<td>0.331 (0.035)</td>
<td>0.389 (0.057)</td>
<td>0.657</td>
</tr>
<tr>
<td>Middle income</td>
<td>26.8</td>
<td>0.122</td>
<td>2409 (568)</td>
<td>0.322 (0.034)</td>
<td>0.457 (0.054)</td>
<td></td>
</tr>
<tr>
<td>High income</td>
<td>24.5</td>
<td>0.437</td>
<td>2435 (561)</td>
<td>0.261 (0.037)</td>
<td>0.446 (0.059)</td>
<td></td>
</tr>
</tbody>
</table>
Patterns remain steady as children age

<table>
<thead>
<tr>
<th></th>
<th>Pooled</th>
<th>Grade 3</th>
<th>Grade 4</th>
<th>Grade 5</th>
<th>Grade 6</th>
<th>Grade 7</th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>All twins</td>
<td>0.441</td>
<td>0.442</td>
<td>0.526</td>
<td>0.430</td>
<td>0.426</td>
<td>0.386</td>
<td>0.373</td>
</tr>
<tr>
<td>Same sex</td>
<td>0.447</td>
<td>0.460</td>
<td>0.527</td>
<td>0.406</td>
<td>0.464</td>
<td>0.394</td>
<td>0.363</td>
</tr>
<tr>
<td>Opposite sex</td>
<td>0.427</td>
<td>0.398</td>
<td>0.524</td>
<td>0.486</td>
<td>0.335</td>
<td>0.365</td>
<td>0.395</td>
</tr>
<tr>
<td>White</td>
<td>0.466</td>
<td>0.504</td>
<td>0.546</td>
<td>0.440</td>
<td>0.419</td>
<td>0.417</td>
<td>0.389</td>
</tr>
<tr>
<td>Black</td>
<td>0.381</td>
<td>0.291</td>
<td>0.476</td>
<td>0.412</td>
<td>0.447</td>
<td>0.300</td>
<td>0.341</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>0.434</td>
<td>0.442</td>
<td>0.518</td>
<td>0.440</td>
<td>0.395</td>
<td>0.376</td>
<td>0.358</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.478</td>
<td>0.442</td>
<td>0.567</td>
<td>0.390</td>
<td>0.576</td>
<td>0.432</td>
<td>0.439</td>
</tr>
<tr>
<td>Non-immigrant</td>
<td>0.440</td>
<td>0.469</td>
<td>0.518</td>
<td>0.439</td>
<td>0.408</td>
<td>0.367</td>
<td>0.348</td>
</tr>
<tr>
<td>Immigrant</td>
<td>0.449</td>
<td>0.323</td>
<td>0.563</td>
<td>0.394</td>
<td>0.510</td>
<td>0.470</td>
<td>0.478</td>
</tr>
<tr>
<td>Unmarried</td>
<td>0.362</td>
<td>0.336</td>
<td>0.402</td>
<td>0.413</td>
<td>0.376</td>
<td>0.363</td>
<td>0.218</td>
</tr>
<tr>
<td>Married</td>
<td>0.485</td>
<td>0.497</td>
<td>0.588</td>
<td>0.446</td>
<td>0.454</td>
<td>0.400</td>
<td>0.458</td>
</tr>
</tbody>
</table>

Note: point estimates are almost always statistically significant at conventional levels.
Patterns remain steady as children age

<table>
<thead>
<tr>
<th></th>
<th>Pooled</th>
<th>Grade 3</th>
<th>Grade 4</th>
<th>Grade 5</th>
<th>Grade 6</th>
<th>Grade 7</th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>All twins</td>
<td>0.441</td>
<td>0.442</td>
<td>0.526</td>
<td>0.430</td>
<td>0.426</td>
<td>0.386</td>
<td>0.373</td>
</tr>
<tr>
<td>HS dropout</td>
<td>0.359</td>
<td>0.249</td>
<td>0.484</td>
<td>0.431</td>
<td>0.257</td>
<td>0.369</td>
<td>0.342</td>
</tr>
<tr>
<td>HS grad</td>
<td>0.434</td>
<td>0.466</td>
<td>0.493</td>
<td>0.410</td>
<td>0.443</td>
<td>0.365</td>
<td>0.365</td>
</tr>
<tr>
<td>College grad</td>
<td>0.529</td>
<td>0.517</td>
<td>0.656</td>
<td>0.493</td>
<td>0.503</td>
<td>0.477</td>
<td>0.433</td>
</tr>
<tr>
<td>Age &lt;=21</td>
<td>0.372</td>
<td>0.371</td>
<td>0.411</td>
<td>0.495</td>
<td>0.237</td>
<td>0.399</td>
<td>0.233</td>
</tr>
<tr>
<td>Age 22-29</td>
<td>0.443</td>
<td>0.417</td>
<td>0.509</td>
<td>0.374</td>
<td>0.534</td>
<td>0.417</td>
<td>0.385</td>
</tr>
<tr>
<td>Age 30-35</td>
<td>0.483</td>
<td>0.466</td>
<td>0.585</td>
<td>0.496</td>
<td>0.465</td>
<td>0.388</td>
<td>0.426</td>
</tr>
<tr>
<td>Age &gt;=36</td>
<td>0.413</td>
<td>0.529</td>
<td>0.570</td>
<td>0.393</td>
<td>0.182</td>
<td>0.270</td>
<td>0.354</td>
</tr>
<tr>
<td>Low income</td>
<td>0.389</td>
<td>0.428</td>
<td>0.445</td>
<td>0.310</td>
<td>0.328</td>
<td>0.399</td>
<td>0.396</td>
</tr>
<tr>
<td>Middle income</td>
<td>0.457</td>
<td>0.409</td>
<td>0.534</td>
<td>0.491</td>
<td>0.504</td>
<td>0.387</td>
<td>0.338</td>
</tr>
<tr>
<td>High income</td>
<td>0.446</td>
<td>0.507</td>
<td>0.550</td>
<td>0.383</td>
<td>0.376</td>
<td>0.320</td>
<td>0.413</td>
</tr>
</tbody>
</table>

Note: point estimates are almost always statistically significant at conventional levels.
Test performance and estimated birth weight effects across groups

Relationship between group average test scores and estimated birth weight effects
Do the effects differ by discordance in birth weight? Non-parametric estimation
Does school quality affect the birth weight gap?

- Since 1999, Florida has graded schools on an A (best) to F (worst) basis
  - Initially based mainly on average proficiency rates on the criterion-referenced Florida Comprehensive Assessment Test
  - From 2002 based on a combination of average proficiency rates and average student-level test score gains from year to year

- We measure “school quality” in three ways:
  - State-awarded school grade
  - Average FCAT performance level
  - Average FCAT gain score
### Does school quality affect the birth weight gap?

<table>
<thead>
<tr>
<th>School quality measure</th>
<th>Percent of population</th>
<th>Mean (sd) of birth weight</th>
<th>Pooled estimate OLS (SE)</th>
<th>Pooled estimate FE (SE)</th>
<th>P-value of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A school</td>
<td>48.6</td>
<td>2437 (559)</td>
<td>0.273 (0.021)</td>
<td>0.407 (0.033)</td>
<td>0.204</td>
</tr>
<tr>
<td>B school</td>
<td>28.8</td>
<td>2409 (570)</td>
<td>0.299 (0.028)</td>
<td>0.497 (0.055)</td>
<td></td>
</tr>
<tr>
<td>C/D/F school</td>
<td>22.6</td>
<td>2375 (578)</td>
<td>0.328 (0.035)</td>
<td>0.455 (0.062)</td>
<td></td>
</tr>
<tr>
<td>&lt;=state median avg. score</td>
<td>39.7</td>
<td>2381 (580)</td>
<td>0.329 (0.028)</td>
<td>0.436 (0.048)</td>
<td>0.831</td>
</tr>
<tr>
<td>&gt;state median avg. score</td>
<td>60.3</td>
<td>2442 (555)</td>
<td>0.259 (0.021)</td>
<td>0.425 (0.033)</td>
<td></td>
</tr>
<tr>
<td>&lt;=state median avg. gain</td>
<td>49.8</td>
<td>2420 (565)</td>
<td>0.319 (0.021)</td>
<td>0.449 (0.036)</td>
<td>0.649</td>
</tr>
<tr>
<td>&gt;state median avg. gain</td>
<td>50.2</td>
<td>2421 (564)</td>
<td>0.298 (0.021)</td>
<td>0.433 (0.035)</td>
<td></td>
</tr>
</tbody>
</table>
Over-time patterns remain steady for different school quality groups as well

<table>
<thead>
<tr>
<th></th>
<th>Pooled</th>
<th>Grade 3</th>
<th>Grade 4</th>
<th>Grade 5</th>
<th>Grade 6</th>
<th>Grade 7</th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>All twins</td>
<td>0.407</td>
<td>0.381</td>
<td>0.459</td>
<td>0.379</td>
<td>0.412</td>
<td>0.339</td>
<td>0.365</td>
</tr>
<tr>
<td>A schools</td>
<td>0.497</td>
<td>0.584</td>
<td>0.523</td>
<td>0.418</td>
<td>0.333</td>
<td>0.271</td>
<td>0.624</td>
</tr>
<tr>
<td>B schools</td>
<td>0.455</td>
<td>0.396</td>
<td>0.550</td>
<td>0.362</td>
<td>0.402</td>
<td>0.512</td>
<td>0.335</td>
</tr>
<tr>
<td>C/D/F schools</td>
<td>0.436</td>
<td>0.352</td>
<td>0.466</td>
<td>0.479</td>
<td>0.520</td>
<td>0.425</td>
<td>0.444</td>
</tr>
<tr>
<td>Below median average score</td>
<td>0.425</td>
<td>0.457</td>
<td>0.499</td>
<td>0.382</td>
<td>0.427</td>
<td>0.387</td>
<td>0.342</td>
</tr>
<tr>
<td>Above median average score</td>
<td>0.449</td>
<td>0.448</td>
<td>0.530</td>
<td>0.389</td>
<td>0.370</td>
<td>0.345</td>
<td>0.435</td>
</tr>
<tr>
<td>Below median average gain</td>
<td>0.433</td>
<td>0.411</td>
<td>0.492</td>
<td>0.397</td>
<td>0.429</td>
<td>0.408</td>
<td>0.368</td>
</tr>
<tr>
<td>Above median average gain</td>
<td>0.407</td>
<td>0.381</td>
<td>0.459</td>
<td>0.379</td>
<td>0.412</td>
<td>0.339</td>
<td>0.365</td>
</tr>
</tbody>
</table>

Note: point estimates are almost always statistically significant at conventional levels.
Are the gaps present at kindergarten entry?

- Florida had two waves of universal kindergarten readiness screening included in statewide data:
  - 1998-2001: School Readiness Checklist: 17 expectations for kindergarten readiness (82.1% of twins ready)
  - 2006-2008: Dynamic Indicators of Basic Early Literacy Skills (DIBELS): Rating letter sounds and letter naming by above average, low risk, moderate risk, and high risk (83.8 % of twins above average or low risk)

- These measures are highly predictive of later test scores: pooled score differences between ready and unready kids are 0.27 standard deviations, in twin fixed effect models
- Important to see whether gaps are present at age 5, and how these gaps compare with those in the testing grades
## Are the gaps present at kindergarten entry?

<table>
<thead>
<tr>
<th>Kindergarten readiness indicator</th>
<th>Kindergarten cohorts assessed</th>
<th>Percent ready by this measure</th>
<th>Estimated effect of log birth weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>School Readiness Checklist</td>
<td>1998-2001</td>
<td>82.1%</td>
<td>0.067 (0.035)</td>
</tr>
<tr>
<td>DIBELS low risk or better</td>
<td>2006-2008</td>
<td>83.8</td>
<td>0.115 (0.043)</td>
</tr>
<tr>
<td>Pooled population of above 2 groups</td>
<td>1998-2001; 2006-2008</td>
<td>83.0</td>
<td>0.086 (0.027)</td>
</tr>
</tbody>
</table>
Comparing kindergarten readiness to test score gaps

- In order to directly compare the threshold-based kindergarten readiness indicator to the continuous grades 3-8 test scores, one must create a comparable measure.

- Therefore, we create a discrete version of the grade 3-8 test scores in which we consider a child “above threshold” on the test if he/she scores in the top 83 percent of the distribution, and “below threshold” otherwise.

- Note: both the kindergarten readiness and threshold test score measures necessarily focus on very at-risk students.
Estimated effects of log birth weight on threshold-based test scores (twin FE model)

<table>
<thead>
<tr>
<th>Panel</th>
<th>n</th>
<th>KG readiness</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt; grade threshold FCAT</th>
<th>Pooled panel FCAT</th>
<th>p-value (2)-(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KGR &amp; 3&lt;sup&gt;rd&lt;/sup&gt; grade</td>
<td>13,718</td>
<td>0.093</td>
<td>0.159</td>
<td>0.159</td>
<td>0.099</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.029)</td>
<td>(0.031)</td>
<td>(0.031)</td>
<td></td>
</tr>
<tr>
<td>KGR &amp; 3&lt;sup&gt;rd&lt;/sup&gt;-5&lt;sup&gt;th&lt;/sup&gt; grade</td>
<td>9,198</td>
<td>0.060</td>
<td>0.178</td>
<td>0.167</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.033)</td>
<td>(0.038)</td>
<td>(0.022)</td>
<td></td>
</tr>
<tr>
<td>KGR &amp; 3&lt;sup&gt;rd&lt;/sup&gt;-8&lt;sup&gt;th&lt;/sup&gt; grade</td>
<td>6,512</td>
<td>0.057</td>
<td>0.181</td>
<td>0.146</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.040)</td>
<td>(0.045)</td>
<td>(0.024)</td>
<td></td>
</tr>
</tbody>
</table>

Test score thresholds are set at the 17th percentile to match the proportion not ready for kindergarten

However…
## Unpacking School Readiness Checklist vs. DIBELS

<table>
<thead>
<tr>
<th>KG readiness exam taken</th>
<th>Coefficient on log birth weight when dependent variable is:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KG readiness exam</td>
<td>G3 Reading</td>
<td>G3 Math</td>
</tr>
<tr>
<td>School Readiness Checklist</td>
<td>0.060</td>
<td>0.098***</td>
<td>0.103***</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.041)</td>
<td>(0.045)</td>
</tr>
<tr>
<td>DIBELS</td>
<td>0.139***</td>
<td>0.101**</td>
<td>0.103***</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.049)</td>
<td>(0.051)</td>
</tr>
</tbody>
</table>

The apparent opening of the gap between kindergarten and third grade is due to measurement. When we compare apples to apples as well as possible, we see that there is no widening of the gap.
Recap: Where we are so far

- Birth weight-related test score differences within twin pairs are present throughout school
  - They are almost identical regardless of grade
  - They are similar across different family types, and across different measures of school quality
  - If anything, the gaps are larger for more advantaged groups: Are early health and parental inputs complements?

- These gaps are also present at the start of kindergarten, and apples-to-apples comparisons suggest they are similarly sized
Exploratory analysis: Evidence of parental remediation prior to school

- There exist few pieces of information regarding parental remediation in administrative data, at least in the USA
  - Hsin (2012), Rosenzweig & Zhang (2009) for China

- We do have a few examples, however, suggesting that parents remediate (or try to independently maximize children’s human capital):
  - Parents send one twin to preschool but not the other (7.6% of twin pairs)
  - Parents either send one twin to preschool but not the other OR send one twin to privately-financed preschool and the other to publicly-financed preschool (9.3% of twin pairs)
  - Parents “redshirt” one twin but not the other (i.e., twins start kindergarten in different years; 0.8% of twin pairs)
Are parents more likely to choose different pre-k paths as BW gap increases?

- Yes: In models where the dependent variable is an indicator for different treatment prior to kindergarten, the coefficients on the difference in log birth weight are:
  - Pr(only one twin attends preschool): 0.053 (se=0.021)
  - Pr(different preschool settings for twins): 0.049 (se=0.023)
  - Pr(redshirt only one twin): 0.022 (se=0.010)

- If parents are attempting to remediate to this degree, it’s not surprising that the gaps would increase modestly after school starts

- But it’s important to recall that the gaps still exist at kindergarten entry, even in the face of this apparent attempted remediation – suggests that the effects of poor neonatal health are difficult to fully stamp out
Conclusions

- There exists considerable evidence that birth weight has persistent effects into adulthood, but the time from birth to 18 has been largely a black box.
- This paper represents the first systematic attempt to study the period from age 5 through school.
- We find that birth weight gaps are present for all groups studied, and persist regardless of family SES or school quality – suggesting that poor neonatal health plays a long-term role throughout schooling.
- Smaller twins from high SES families tend to do very well, but not quite as well as larger twins from the same families!
- Birth weight gaps appear to be stable throughout schooling.
Is stability driven by the specification of birth weight?

<table>
<thead>
<tr>
<th></th>
<th>Pooled</th>
<th>Grade 3</th>
<th>Grade 4</th>
<th>Grade 5</th>
<th>Grade 6</th>
<th>Grade 7</th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln (BW)</td>
<td>0.441</td>
<td>0.442</td>
<td>0.526</td>
<td>0.430</td>
<td>0.426</td>
<td>0.386</td>
<td>0.373</td>
</tr>
<tr>
<td>BW in 1000g</td>
<td>0.186</td>
<td>0.185</td>
<td>0.223</td>
<td>0.178</td>
<td>0.180</td>
<td>0.169</td>
<td>0.155</td>
</tr>
<tr>
<td>BW</td>
<td>0.198</td>
<td>0.196</td>
<td>0.234</td>
<td>0.191</td>
<td>0.193</td>
<td>0.177</td>
<td>0.171</td>
</tr>
<tr>
<td>BW*(mean twin pair BW)</td>
<td>-0.105</td>
<td>-0.117</td>
<td>-0.105</td>
<td>-0.114</td>
<td>-0.105</td>
<td>-0.058</td>
<td>-0.112</td>
</tr>
<tr>
<td>BW</td>
<td>0.450</td>
<td>0.477</td>
<td>0.487</td>
<td>0.466</td>
<td>0.446</td>
<td>0.317</td>
<td>0.441</td>
</tr>
<tr>
<td>BW²</td>
<td>-0.053</td>
<td>-0.058</td>
<td>-0.053</td>
<td>-0.057</td>
<td>-0.053</td>
<td>-0.029</td>
<td>-0.056</td>
</tr>
</tbody>
</table>
Is stability driven by the specification of birth weight?

<table>
<thead>
<tr>
<th></th>
<th>Pooled</th>
<th>Grade 3</th>
<th>Grade 4</th>
<th>Grade 5</th>
<th>Grade 6</th>
<th>Grade 7</th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Both twins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2500g</td>
<td>0.473</td>
<td>0.470</td>
<td>0.569</td>
<td>0.495</td>
<td>0.455</td>
<td>0.339</td>
<td>0.419</td>
</tr>
<tr>
<td>&gt;2500g</td>
<td>0.526</td>
<td>0.575</td>
<td>0.656</td>
<td>0.464</td>
<td>0.528</td>
<td>0.409</td>
<td>0.427</td>
</tr>
</tbody>
</table>
Is stability driven by the specification of the test score?

<table>
<thead>
<tr>
<th></th>
<th>Pooled</th>
<th>Grade 3</th>
<th>Grade 4</th>
<th>Grade 5</th>
<th>Grade 6</th>
<th>Grade 7</th>
<th>Grade 8</th>
</tr>
</thead>
</table>