SREE 2011: Methods symposium submission

Contact (session chair) email
Mark W. Lipsey (mark.lipsey@vanderbilt.edu)
Peabody Research Institute, Vanderbilt University

Title
The age cutoff regression-discontinuity design for preschool effects: Methodological quirks and implications for application

First and second choice of conference section
Research Methods
Early Childhood Education

Symposium justification – 500 word limit
Many preschool programs have a strict age cutoff for admission that allows children born before the cutoff date to be enrolled and those born after that date to wait until the next year to enter the program. An increasingly popular research design for investigating the effects of preschool is regression-discontinuity with age as the selection variable and the cutoff date for a given school year differentiating a group of children who participate in preschool that year (treatment) and a group who participate the next school year (control). A comparison of the outcomes measured for each group at the beginning of the subsequent school year (K for treatment, pre-K for control) with each regressed on the age selection variable then yields an estimate of the effect of preschool.

This design has many attractive features. When its assumptions are met, the regression-discontinuity design in general is capable of producing unbiased treatment effect estimates and, in that regard, is comparable to a randomized control design. Moreover, in this particular application it has a natural fit to the circumstances, requiring little alteration in the normal admission practices of the respective preschool programs and allowing all children who apply for preschool to actually attend. A series of studies using this design has been produced in recent years, all showing positive preschool effects (e.g., Gormley et al, 2005; Wong et al., 2008)*.

As implemented in these and similar studies, however, this design has some methodological quirks that have been overlooked in the related discussions, the most serious of which have the potential to bias the treatment effect estimates. The presentations proposed for this symposium include, first, an initial paper that will identify these issues, explain their implications for this application of the regression-discontinuity design, and consider possible solutions. Two additional presentations will describe ongoing studies that are using this design with an emphasis on the methodological issues identified and the attempts made to eliminate the respective problems from the design, minimize their adverse effects, or explore the nature of those effects. Finally, Tom Cook, who has written extensively about the regression-discontinuity design and been involved in preschool applications of it, will offer comments as a discussant.

---

Presentation #1

Title: Methodological issues, implications, and remedies for the preschool age cutoff regression discontinuity design

Author(s): Christina Weiland
Harvard Graduate School of Education
Christina_weiland@mail.harvard.edu

Mark W. Lipsey
Peabody Research Institute, Vanderbilt University
mark.lipsey@vanderbilt.edu

Hiro Yoshikawa
Harvard Graduate School of Education
yoshikhi@gse.harvard.edu

Background / Context:
State-funded prekindergarten is an increasingly popular preschool model in the U.S. The number of states offering such programs increased from 10 in 1980 to 38 in 2009 (Gormley, Gayer, Phillips, & Dawson, 2005; National Institute for Early Education Research, 2009). Several recent studies have used regression discontinuity (RD) design to estimate the impacts of these programs on children’s outcomes (Gormley, Gayer, Phillips, & Dawson, 2005; Hustedt, Barnett, Jung & Goetze, 2009; Hustedt, Barnett, Jung & Thomas, 2007; Wong, Cook, Barnett, & Jung, 2007). These studies capitalize on birthdate requirements to enter prekindergarten programs in a given year—children born before the cutoff constitute the treatment group, while children born after the cutoff, who have to wait a year to enter the program, constitute the control group. Encouragingly, these studies have found small to moderate positive impacts on children’s language, literacy and numeracy skills.

More generally, capitalizing on pre-existing birthdate cutoffs to estimate the effects of prekindergarten has several attractive features. First, provided its assumptions are met, the RD design produces unbiased treatment effect estimates that are comparable to a randomized control design. Second, the design does not disrupt pre-existing prekindergarten admissions practices and systems; unlike a randomized controlled trial, no children have to be denied seats in a prekindergarten program to accommodate the design. Third, although inferences from RD studies apply directly only to those at the cutoff, prekindergarten RD studies may have improved external validity compared to randomized controlled trials. Districts able to participate in RD studies may be more representative than districts in situations such as excess demand that would be most amenable to participation in a randomized controlled trial.

As implemented in prekindergarten and other studies that employ age cutoffs, however, RD design has some methodological quirks that have been overlooked in the related discussions. The most serious of these quirks have the potential to bias the treatment effect estimates.

Purpose / Objective / Research Question / Focus of Study:
We identify several methodological issues that are largely unexplored in existing prekindergarten studies that use the age cutoff RD design. We suggest strategies for addressing
these issues and we use empirical data from a large-scale public prekindergarten program to illustrate both the potential problems we identify and the results of the potential solutions we suggest. We then make practical recommendations for future research using this design.

**Setting:**

The data from the study used to illustrate some of the points in this presentation were collected in public prekindergarten programs located in elementary schools in a large urban school district in the Northeast.

**Population / Participants / Subjects:**

The sample in the study used to illustrate key points includes 2018 racially, linguistically, and socio-economically diverse children (see Table 1) who attended a citywide 4-year-old publicly funded pre-k program in either 2008-2009 (treatment) or 2009-2010 (control).

**Intervention / Program / Practice:**

The illustrative pre-k program is universal in design, as any child within the city who turns four by September 1 in a given year may apply for the program. Currently, approximately 35% of the city’s 4-year-old population attends the prekindergarten program. Importantly, the district strictly enforces the cutoff; in recent years, no child has been admitted into the program whose birthdate did not qualify. Approximately 94% of children offered a spot in the program in 2008-2009 enrolled for at least one day.

**Data Collection and Analysis:**

In the illustrative pre-k study, children were tested by study-trained child assessors using measures of pre-literacy, numeracy and executive function skills. All assessors were college educated and approximately one third held masters’ degrees. We randomly varied the order of the tests to limit the possibility of systematically biasing results due to child fatigue. The assessors visited classrooms in Fall 2009, as close to the start of the school year as possible.

**Significance / Novelty of study:**

There are several unexplored issues in studies of prekindergarten programs that use the age cutoff regression-discontinuity design that plausibly could effect estimates and/or threaten internal validity. First, the treatment group children have more time both to join and to drop out from the program than control children due to the timing of study assessments. All children are assessed at the same time: treatment children in the fall of their K year and control children in the fall of their pre-k year. This timing is determined by a practical constraint—control children cannot be identified until they actually enter the pre-k program—but it has implications for defining the treatment and control groups. Children can be counted as part of the treatment group if they enter pre-k in any part of the school year but, to be in the control group, children must have entered pre-k during the fall. Similarly, treatment children could drop from the program at any time during their pre-k year while attrition by control children is not yet known at the time their outcomes are assessed. Of concern for estimating impacts of pre-k programs, these time frame differences could plausibly create non-equivalence between the treatment and control groups that was not controlled by the intrinsic features of the RD design.

Second, many early childhood assessment instruments have different start/stop rules for 4-year-old vs. 5-year-old children so that outcome measurement is not strictly parallel for the
treatment and control groups. Estimates of the effects on such measures thus could be influenced by these protocol differences. On the Peabody Picture Vocabulary Test-III, for example, which has been used in several pre-k RD studies, 5-year-olds start the test one set (12 items) ahead of 4-year-olds children. The final score is determined by the highest item reached before the stop rule applies minus the total number of errors. When used as instructed by the test developers, this scoring assumes that a 5-year old child has a perfect score on the base set of items given to 4-year-olds. Two children of equal abilities born one day apart could conceivably be scored differently due to this test design feature.

Third, the birthdate cutoff policy might influence the counterfactual. For example, children born just after the cutoff (control) might have peers born just before the cutoff (treatment). Consequently, parents may seek out more academically oriented programs for those children than they would have in absence of the policy during the year control group children are denied access to the prekindergarten program. This would artificially attenuate the impacts of the pre-k program.

Fourth, districts may require either active or passive parental consent for children to participate in an RD study. Consent rates and the characteristics of the children whose parents consent may differ for children just beginning pre-k and those just beginning K with implications for internal validity. There may also be issues of external validity-- children in active consent districts whose parents consent likely differ from children whose parents do not return consent forms. Whether districts required active or passive consent is important in interpreting and comparing results from evaluations of prekindergarten programs, as well as in understanding to whom results can be generalized.

Fifth, ‘redshirting’ is an increasingly common practice with parents holding children out of school to allow them additional time to mature and to have a competitive edge in school. If children are permitted to enter pre-k at older ages than would be allowed under strict age cutoffs, this raises a host of problems for studies that use RD based on those cutoffs. Using Instrumental Variables (IV) to correct for this problem is one possible solution. However, a recent paper (Aliprantis, 2010) points out that redshirted children violate the monotonicity assumption of IV; a sizable percentage of redshirted children will result in biased estimates if one uses IV to calculate a local average treatment effect (LATE).

Finally, calculating intent-to-treat estimates in RD studies that use age cutoffs is very challenging. We can assume that children are randomized into before- and after-cutoff groups at birth. However, children must select into pre-k programs to be included in RD studies of these programs. From an intent-to-treat perspective, the offer of the treatment is the question of interest, since attending prekindergarten is not mandatory, but age cutoff RD studies cannot produce these estimates.

**Findings / Results:**

We suggest practical design and data analytic strategies for addressing the issues we have identified in the age cutoff regression discontinuity design. Our suggestions are preliminary and we will be honing them before SREE 2011.

First, to address the problem that treatment children have more time than the controls to join and drop from pre-k, we suggest standard checks to verify that characteristics of children vary smoothly across the cutoff (Imbens & Lemieux, 2008; Lee & Lemieux, 2010). We also recommend that researchers examine the sensitivity of the impact results to different definitions of the sample. The broadest sample would include children enrolled in the prekindergarten
program at least one day, while the most narrow would be restricted to children enrolled for the full duration of the prekindergarten program. Another strategy is an adaptation of a strategy from Krueger (1999). As a lower bound on the effect of differential attrition, treatment children who drop out before their kindergarten fall could be assigned the lowest possible test score and impact models could then be refit including these children. Importantly, implementing these strategies requires that researchers gain access to student enrollment records for all prekindergarten and kindergarten children in the district, not just those who agree to participate. In addition, to identify which control group children are the counterparts of the treatment children who drop out, researchers must follow control children through their kindergarten fall.

Second, regarding the problem of assessment start/stop rules that vary by child age, we recommend that researchers apply the same rules to all children, regardless of age. Where different start/stop rules have been used, we researchers can adjust the scores for assessment rule differences as a robustness check. The exact adjustment will depend on the assessment; if treatment group children are implicitly given credit for items, as in the PPVT-III example, scores can be adjusted downward for these children by assuming they would have missed a few more items had they started the assessment in the same point as the control group children.

Third, to address the possibility that the cutoff policy may have changed the counterfactual, researchers could collect information on children’s childcare placement prior to the prekindergarten program, including type, dosage, and quality. These data can be used to see if care selection varies smoothly across the cutoff. Further, previous childcare information will aid in interpreting the impact results more generally.

Fourth, we recommend weighting data in studies that require active consent. While this will solve only the problem of selection on observables, we believe it has some merit, so long as the assumption that refusal to participate is a function of observables is reasonable. Children in the study in the Northeast we described earlier in this proposal, for example, were more likely to refuse to participate if they were Hispanic. Weighting would help make the estimates from the study we described more generalizable to Hispanics.

Fifth, regarding the increasingly common practice of redshirting, we recommend that these children not be excluded from studies a priori. Particularly if the percentage of these children is not sizable, there are methods to address observations that violate the cutoff or that are far away from the cutoff.

Finally, we have no easy solutions regarding the issue of calculating intent-to-treat estimates. Population-level indicators on young children can help to address the degree to which children who select into the program are different from those who do not, at least on observable characteristics. The Krueger (1999) strategy may be useful as well; children who do not enroll in the treatment could be assigned the lowest scores. Estimates from including this group of children could serve as a lower bound.

Usefulness / Applicability of Method:

At SREE, we will present evidence on implementing strategies to address all six issues. All results will be based on data from the large urban prekindergarten program described earlier.

Conclusions:

Closer scrutiny must be given to the age cutoff regression-discontinuity design used for assessing preschool effects. While this is a potentially strong design, there are aspects of it that may lead to biased treatment effect estimates, especially in some applications. Some of these
issues can be addressed in design and data analysis but others are less tractable and require research to determine the extent of their influence.

**Appendix A. References**

**Appendix B. Tables and Figures**

Table 1: Descriptive characteristics of sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Overall (N=2018)</th>
<th>Born before cutoff (N=969)</th>
<th>Born after cutoff (N=1049)</th>
</tr>
</thead>
<tbody>
<tr>
<td>English only home language</td>
<td>0.50</td>
<td>0.48</td>
<td>0.53</td>
</tr>
<tr>
<td>Spanish home language</td>
<td>0.27</td>
<td>0.28</td>
<td>0.27</td>
</tr>
<tr>
<td>Other home language</td>
<td>0.22</td>
<td>0.24</td>
<td>0.20</td>
</tr>
<tr>
<td>Black</td>
<td>0.27</td>
<td>0.28</td>
<td>0.25</td>
</tr>
<tr>
<td>White</td>
<td>0.18</td>
<td>0.18</td>
<td>0.19</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.41</td>
<td>0.39</td>
<td>0.42</td>
</tr>
<tr>
<td>Asian</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>Other race/ethnicity</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>Special Needs</td>
<td>0.09</td>
<td>0.11</td>
<td>0.08</td>
</tr>
<tr>
<td>Free/reduced lunch receipt</td>
<td>0.69</td>
<td>0.72</td>
<td>0.66</td>
</tr>
<tr>
<td>Male</td>
<td>0.51</td>
<td>0.52</td>
<td>0.50</td>
</tr>
</tbody>
</table>

*Note: one child born after the cutoff is missing all information in this table.*
Presentation #2

Title: Program evaluation of an Early Reading First project using regression discontinuity

Author(s): Sandra Jo Wilson
Peabody Research Institute, Vanderbilt University
sandra.j.wilson@vanderbilt.edu

Background / Context:
Preschool attendance has increased rapidly in the past 20 years and there are now state-funded programs in more than three-fourths of the states (Barnett et al., 2007a). With income requirements in over 70% of these programs, many more disadvantaged children have access to preschool than ever before. In spite of widespread preschool attendance by children from low income families, however, these children continue to enter kindergarten performing below their peers from middle income families in pre-academic skills such as early literacy and math as well as social skills (Breitenstein, Hill, & Gross, 2009; Chernoff, et al., 2007; O’Donnell, 2008; Zill & West, 2001). In this context, it is important that preschool programs demonstrate their effectiveness in bolstering the academic success of low income children.

The Early Reading First program of the U.S. Department of Education grew out of the need to provide support for the burgeoning preschool programs around the country. Early Reading First funding, in that regard, is intended mainly as programmatic funding rather than research funding; the evaluation requirements for applicants are minimal. Indeed, most Early Reading First grantees employ pretest-posttest designs without control groups for their evaluations. Grantees focus on providing curricula and materials for preschool classrooms, and implement extensive professional development and one-on-one coaching for the preschool teachers. Given the ubiquity and expense of public preschool, and the somewhat mixed results on the effectiveness of such programs, the research project described here was designed to serve two purposes. It needed to be an independent evaluation of an ERF program that met the needs of the project team for reporting to the funding agency and for providing feedback to teachers, but we also decided that it should incorporate an age cutoff regression discontinuity design to allow for examination of the impact of the program. In the course of designing, implementing, and analyzing the data from that design, various issues and questions arose about the way the design has been used in other studies and the specific application in this instance.

Purpose / Objective / Research Question / Focus of Study:
The main objective of this research was to provide a high quality quasi-experimental evaluation of an Early Reading First intervention program. Using the age cutoff regression-discontinuity design in this context allowed for an inexpensive but rigorous evaluation that was acceptable to the program team and local schools. This presentation will report the results of that evaluation but will also discuss the methodological issues associated with the design, their implications, and how they were handled. These issues include:

• Possible cohort differences across successive years of children.
• Using cohorts as control groups for one year and as treatment groups the next in multi-cohort sequences.
• Finding parallel measures that can be used with children of different ages.
• Defining the sample with regard to the kindergarten placements of treatment children after they complete their preschool program.
• Determining what constituted an intent-to-treat sample and which children should be identified as non-completers.

Setting:
The research is being conducted in 13 preschool classrooms in 7 public elementary schools in the Metropolitan Nashville area. The preschool program in the district is funded by the Tennessee Department of Education and serves mainly low income preschool children. To be eligible for the preschool program, children’s families must meet income requirements or the children must be English Language Learners, have very low school readiness scores, or have a disability that would require special education services upon entrance to formal schooling. As a result of these requirements, the elementary schools participating in the research are located primarily in low income and/or immigrant neighborhoods. Of the 7 schools in the project, three serve large populations of recently emigrated families, including families from Latin America, the Middle East, and Africa. The other four schools serve largely African American children.

Population / Participants / Subjects:
Three annual cohorts of preschool children are participating. The project is currently in the fall of its third year; two cohorts of children have completed the preschool program described below. As the first cohort of students began kindergarten, they served as the treatment group in the age cutoff regression-discontinuity design. As the second cohort of children began the ERF program, they served as the control group in the design. Demographic information for both groups of participating children is shown in Table 1 (Table 1 here).

Intervention / Program / Practice:
The intervention used in the program is a multi-service package that includes a language and literacy oriented preschool curriculum and extensive training and coaching support for teachers. Each teacher is assigned a coach who spends one day per week in the classroom to observe lessons, demonstrate lessons and instructional techniques, and provide ongoing feedback to teachers. The preschool curriculum selected for implementation in this program is Opening the World of Learning (OWL). It systematically teaches vocabulary, phonemic awareness, and letter knowledge in varied settings. Each thematic unit introduces and systematically develops core concepts over 4 weeks, enough time for children to encounter new ideas and words on multiple occasions and to apply new language and concepts as they engage with theme-related questions. Teachers learn to support vocabulary and language learning through large group read aloud techniques that vary across the multiple readings of narrative and expository books. Teachers also learn to plan and implement small group lessons that select and reinforce key concepts and to use informal talk during learning center activities, meals, and transitions to support oral language comprehension, production, and vocabulary learning.

This curriculum is supplemented with two additional components. An emergent writing strand includes teacher demonstrations during large group interactive writing activities and child opportunities for writing at the classroom writing center. An English language learner strand includes professional development for teachers and assistants in basic techniques for working with second language learners, a two week summer program for children aimed at helping
prevent summer learning loss, and a lending library program which allows children to use their home language resources to learn English.

Significance / Novelty of study:
This research adds to the growing body of studies using the age cutoff regression-discontinuity design to investigate the effects of preschool. It further highlights some of the methodological issues related to that design, some of which have not been well recognized or deeply examined in prior research and discussions on this design.

Research Design:
The research design makes use of the preschool age cutoff applied by the local school district to create a regression-discontinuity design. Each incoming cohort of students must meet the age eligibility requirement of being age 4 on or before October 1 of the current school year to be enrolled. Thus far, this requirement has been rather strictly applied by the school district, with only one child in the first two cohorts falling outside the age requirement. The treatment group is comprised of children who met the birthday cutoff in one year and received the preschool program in that year; the control group is comprised of children who did not meet the birthday cutoff and who did not enter the preschool program until the following year. The assessment schedule for the project allows for the outcome measures to be collected from the treatment and control groups during the same period at the beginning of their respective kindergarten and pre-kindergarten school years.

Data Collection and Analysis:
Three waves of assessments are collected on each participating child-- a pretest at the beginning of preschool, a posttest at the end of preschool, and a follow-up at the beginning of kindergarten. The pretest and posttest assessments are used primarily to provide evaluation results to the project team, teachers, and the Department of Education. The kindergarten follow-up assessment makes the regression discontinuity design possible because it is collected at the same time as the next cohort’s ‘pretest’ at the beginning of preschool.

Five standardized tests are given to the children. The Peabody Picture Vocabulary Test, required by the Department of Education, is a measure of receptive vocabulary. Four subtests of the Woodcock Johnson Test of Achievement were also given: Letter-Word Identification, Spelling, (expressive) Picture Vocabulary, and Oral Comprehension. Assessments are collected in individual sessions with the children and administered by trained examiners.

Analyses of the impact of the intervention on the student achievement outcomes were conducted using 3-level hierarchical mixed models in SPSS. The first cohort’s kindergarten assessment was the relevant dependent variable for the treatment group and the second cohort’s ‘pretest’ assessment was the dependent variable for the control group. These were administered during the same 10 week period in the fall of 2009 for all children. Each of the achievement outcomes was run as a separate model with the following covariates: time of test, gender, ethnicity, ELL status, and linear and quadratic variables for the selection continuum. The selection continuum variable was calculated as the number of days from the birthday cutoff (this ranged from 0 to 365 for the treatment group and from -365 to 0 for the control group). Children were nested within classrooms and schools.
Findings / Results:

Five models were run, each with one of the achievement outcomes as the dependent variable. The quadratic form of the selection continuum variable was not significant in any of the five models; in all cases, the linear form of the selection continuum was an adequate representation of the relationship of the selection variable to the outcomes.

Relatively large, statistically significant effects favoring the treatment group were found on all five outcome variables, controlling for the covariates mentioned above. The unstandardized regression coefficients for the treatment variable and the adjusted group means are shown in Table 2 (Table 2 about here).

Girls and boys performed about equally on all assessments except the Spelling subtest, with girls outperforming boys on that subtest. ELL children’s performance was significantly lower than their English-speaking peers on the PPVT, and the Woodcock Johnson Picture Vocabulary and Oral Comprehension subtests.

Further analysis focusing on methodological issues are underway and include detailed examination of cohort differences, sensitivity analysis for variation in the sample definition, investigation of the implications of nonparallel measures, and related matters.

Conclusions:

Results indicate that the Early Reading First intervention package had a statistically significant effect on children’s receptive and expressive vocabulary, letter-word identification, spelling, and oral comprehension. Discussion of these results will include the treatment effects, but will also illustrate specific methodological issues associated with the age cutoff regression-discontinuity design using this project as an example and report on some of the sensitivity tests and empirical explorations that have been made on these matters. The following briefly summarizes the issues that have become apparent in the current project.

- Cohort differences across successive years of children. The project is designed to include three annual cohorts of children, thereby making possible two treatment groups and two control groups. Demographic characteristics of the two cohorts participating thus far differ in ways that raise questions about the equivalence of the treatment and control groups at the cutting point where the design assumes a smooth distribution. Various ways of testing and controlling for cohort differences will be explored.

- Using cohorts as control groups for one year and treatment groups in the next in multi-cohort designs. The second cohort of children serves as the control group for the results presented here, but will also serve as the treatment group when compared to the third cohort. There are some ways to take advantage of this situation, which allows for replication of the design and results, but also issues of overlap, statistical dependency, and appropriate analytic models.

- Finding parallel measures that can be used with children of different ages. Some of the standardized achievement measures used in this study have different starting points for children at different ages. We will discuss how this was handled in the present project.

- Defining the sample with regard to the kindergarten placements of treatment children after they complete their preschool program. The preschool children in the treatment group do not all go to kindergarten in the same school as their preschool. We will use the pattern of kindergarten attendance in the present project to discuss how this circumstance might affect the results of the design.
Determining what constitutes an intent-to-treat sample and which children should be identified as non-completers. There are various ways to define what is presumed to be the initial sample that was presumably divided between treatment and control groups only on the basis of the age cutoff, especially with regard to the control group. The implications of those variations that can be compared using the data from this project will be discussed further.

Appendix A. References
### Appendix B. Tables and Figures

#### Table 3. Demographic Characteristics of the ERF Students

<table>
<thead>
<tr>
<th></th>
<th>Cohort 1 (Treatment)</th>
<th>Cohort 2 (Control)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Male</td>
<td>85</td>
<td>47%</td>
</tr>
<tr>
<td>Female</td>
<td>96</td>
<td>53%</td>
</tr>
<tr>
<td>ELL</td>
<td>43</td>
<td>24%</td>
</tr>
<tr>
<td>Not ELL</td>
<td>138</td>
<td>76%</td>
</tr>
<tr>
<td>Black/African American</td>
<td>119</td>
<td>66%</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td>27</td>
<td>15%</td>
</tr>
<tr>
<td>Caucasian</td>
<td>13</td>
<td>7%</td>
</tr>
<tr>
<td>Arabic</td>
<td>10</td>
<td>6%</td>
</tr>
<tr>
<td>African</td>
<td>9</td>
<td>5%</td>
</tr>
<tr>
<td>Asian/Asian American</td>
<td>3</td>
<td>2%</td>
</tr>
<tr>
<td>Other</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Age at pretest</td>
<td>4.7 years</td>
<td>56m</td>
</tr>
<tr>
<td>Age at posttest</td>
<td>5.3 years</td>
<td>63m</td>
</tr>
</tbody>
</table>

#### Table 2. Hierarchical Mixed Model Results

<table>
<thead>
<tr>
<th></th>
<th>B for treatment effect</th>
<th>Treatment group adjusted mean (n)</th>
<th>Control Group adjusted mean (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPVT</td>
<td>9.5*</td>
<td>58.6 (186)</td>
<td>49.1 (240)</td>
</tr>
<tr>
<td>Letter-Word ID</td>
<td>26.4*</td>
<td>345.1 (186)</td>
<td>318.7 (237)</td>
</tr>
<tr>
<td>Spelling</td>
<td>13.5*</td>
<td>370.4 (186)</td>
<td>357.0 (237)</td>
</tr>
<tr>
<td>Picture Vocabulary</td>
<td>10.0*</td>
<td>452.0 (186)</td>
<td>442.0 (236)</td>
</tr>
<tr>
<td>Oral Comprehension</td>
<td>5.8*</td>
<td>443.9 (186)</td>
<td>438.1 (236)</td>
</tr>
</tbody>
</table>

*p<.05
Presentation #3

Title: Age cutoff regression-discontinuity evaluation of the Tennessee Voluntary Pre-Kindergarten program: Some conclusions and conundrums

Author(s): Mark W. Lipsey  
Peabody Research Institute, Vanderbilt University  
mark.lipsey@vanderbilt.edu  

Dale C. Farran  
Peabody Research Institute, Vanderbilt University  
dale.farran@vanderbilt.edu

Background / Context: Relatively few rigorous studies of the effectiveness of contemporary public pre-kindergarten programs have been conducted, though the recent spate of age cutoff regression discontinuity (RD) studies has added significantly to that body of work. The study on which this presentation will be based was launched in partnership with the Tennessee State Department of Education’s Office of Early Learning to provide an assessment of the effects of the statewide Tennessee Voluntary Pre-Kindergarten (TN-VPK) program on the readiness for kindergarten of the economically disadvantaged population it serves. TN-VPK has become a somewhat controversial program in Tennessee, with some legislators expressing doubts about its value in the context of severe budget shortfalls and still others referring to it even more skeptically as expensive babysitting.

Purpose / Objective / Research Question / Focus of Study: This study is part of a larger project investigating the effects of TN-VPK that also includes a randomized control experiment. That RCT, however, could only be implemented in a limited number of schools with many more applicants than seats in the pre-k program. Informative as that study component will be about the effects of TN-VPK, the participating schools do not provide a representative sample of the TN-VPK classrooms in Tennessee. To provide a more representative statewide picture, a stratified random sample of schools with TN-VPK classrooms was drawn and enrolled in age cutoff regression-discontinuity designs. This design was made possible because TN-VPK has explicit age requirements that are implemented as strict age cutoffs for TN-VPK eligibility in all schools.

The primary purpose of the RD component of the larger study is to determine the effect statewide of TN-VPK on the kindergarten “readiness” of the participating children. A secondary purpose is to investigate the relationships between those outcomes and the characteristics of the TN-VPK classrooms, e.g., the curriculum used, teacher credentials, and classroom quality.

Setting: TN-VPK supports over 900 pre-k classrooms that serve more than 18,000 at-risk children in 133 of the 136 Tennessee school systems. It is, therefore, truly a statewide program serving all areas of the state, both rural and urban.
Population / Participants / Subjects:
To be eligible for TN-VPK, children must be age four on or before September 30 of the respective school year. By statutory requirement, the program gives top priority to children who qualify for the Free or Reduced Price Lunch Program, and 86% of the children enrolled statewide meet that criterion. Statewide, 28% of the children enrolled in TN-VPK are African-American with another 9% other minorities, 4% are English language learners, and 10% have special education designations.

Intervention / Program / Practice:
TN-VPK operates through competitive grants to local school systems who apply for approval and funding. Those grants support only a portion of the necessary classroom funding, the balance must come from other sources. This arrangement permits and encourages collaboration between school systems and other organizations. In this “collaboration model,” school districts may operate their pre-k programs through collaborative agreements with local nonprofit and for-profit child care providers and Head Start programs so long as those agencies have attained the highest rating from the licensing system administered by the Tennessee Department of Human Services and meet the TN-VPK standards. Those standards, which are set by the State Board of Education, require the following:
• A state licensed teacher with an early childhood education endorsement in each classroom;
• A teacher assistant who holds or is actively working toward at least a CDA or associate degree in early childhood;
• Professional development support for teachers;
• An adult-student ratio no smaller than 1:10;
• A small class size (maximum of 20);
• An approved age-appropriate curriculum aligned with the Tennessee Early Childhood Education Developmental Standards;
• A family engagement component and a pre-k to kindergarten transition plan for each child;
• Vision, hearing, and health screening and referral services;
• A minimum of 5.5 hours per day, exclusive of nap time, for a minimum of 180 days per year within a calendar that includes 200 working days of 7.5 hours for teaching staff.

Research Design:
A stratified random sample of 128 TN-VPK classrooms was drawn from the 2009-2010 state population of 942 classrooms in a manner that ensured adequate representation of the geographical regions of Tennessee, urban and rural schools, and public school based and ‘partnership’ programs implemented collaboratively with community agencies. No more than one classroom per school/site was selected and each classroom was designated a block in a blocked RD design; that is, each classroom provides both treatment and control cases from two successive cohorts of students. One of four designated Tennessee regions enters the design each year of the project, with the first region beginning in school year 2009-2010. Each cohort of students is assessed in the fall early in the school year with the treatment group assessed at the beginning of kindergarten after participating in the selected TN-VPK classroom the prior year and the control group of children assessed that same fall soon after they are newly enrolled in that classroom. During the school year when the treatment cohort is in the selected classroom, classroom observations measures are taken along with teacher questionnaires and other descriptive information about the respective classroom.
Data Collection and Analysis:
The two cohorts of children in the sampled classrooms are individually assessed using a set of Woodcock Johnson III achievement tests of pre-reading, language, and mathematic skills (Letter-Word Identification, Spelling, Understanding Directions, Applied Problems, Quantitative Concepts, Passage Comprehension, and Mathematics Calculation). Classroom observations are conducted using the ECERS and a locally developed instrument (‘narrative record’) on which classroom activities and the time spent on each are recorded. Analysis is done multilevel with children nested in classrooms. At the time of this writing, data are available from the first of the four Tennessee regions representing 32 classrooms and more than 1200 children in the two cohorts enrolled in those classrooms. Data collection is currently underway for classrooms in the second region, which has entered the design for the 2010-2011 school year.

Significance / Novelty of study:
Aside from ultimately producing a large, representative statewide sample for which pre-k effects can be analyzed in conventional form for the age cutoff regression-discontinuity design, this study has several features that raise methodological issues and/or allow exploration of their influence.

As in the other studies to be discussed in this symposium, efforts have been made to make the outcome measures parallel by using the same start and stop rules for both cohorts, in particular, also using the rules for 4-year olds when testing the 5-year olds. These data are collected in a manner that will allow examination of the difference it would make if the 5-year olds had, instead, been tested using the prescribed test rules for that age.

An especially distinctive feature of this study is the classroom level blocking. This focuses attention on the comparability of the successive cohorts coming through the same classroom and raises questions about the ways in which they might differ in that local context. To better address these questions, we are tracking the control students (second cohort through each classroom, assessed at the beginning of pre-k) through school records to determine their drop out status during their pre-k year and where they end up in kindergarten (or if they do—some are home schooled). This will allow us to conduct sensitivity tests using various definitions of the treatment and control children who constitute the respective study samples—all the children tested in each cohort, only those with similar exposure to the pre-k classroom, only those who continue into K in the same school or school district, etc.

A final feature of this RD study is that it is one component of a larger study that also involves a randomized control design. The latter is implemented only in willing schools with an excess of applicants, not all of whom can be admitted to the limited pre-k spaces available. However, there is some overlap between the schools participating in the randomized design and those participating in the regression-discontinuity design. These are not perfectly aligned so as to involve the same treatment children and school years but, in many cases, are close enough to permit potentially informative comparisons between the treatment estimates derived from the randomized design and those from the regression-discontinuity design.

Findings / Results:
Analysis of the data only recently collected for the first of the four Tennessee regions in the RD design is currently underway. Results on the issues raised above and others that may arise as we pursue those analyses will be available for this presentation at the time of the SREE conference.
Discussant

Title: Comments on the presentations and related remarks about the age cutoff regression-discontinuity design for preschool effects

Author: Thomas D. Cook
Northwestern University
t-cook@northwestern.edu