

# **SAMPLING WITH UNCERTAIN FRAME COUNTS: CHALLENGES IN SAMPLING HEAD START CHILDREN FOR THE FACES STUDY**

**Barbara Lepidus Carlson  
Mathematica Policy Research  
Princeton, New Jersey  
August 1, 2007  
JSM Salt Lake City**

# Head Start Family and Child Experiences Survey (FACES)

---

- **FACES is a repeated cross-sectional study of Head Start program quality and child outcomes (previous cohorts: 1997, 2000, and 2003)**
- **Gathers comprehensive data on the cognitive and social-emotional development of Head Start children**
- **Each cohort of children is followed from entry into Head Start, through 1 or 2 years of program participation, with followup in the spring of kindergarten**

# 2006 FACES Sample Design

---

- **Four stages (targeted completes)**
  - programs (60)
  - centers (2 per program)
  - classes (3 per center)
  - children (10 per class)
- **Home visitors included**
- **First 3 stages selected PPS (estimated number of eligible children)**

# Eligible Sampling Units

---

- **Programs**
  - 50 states plus DC
  - Excluding AI/AN and Migrant/Seasonal programs
  - Providing direct services to child in target age group
  - In good standing (financially, operationally)
- **Centers and Classes and Home Visitors**
- **Children Must Be:**
  - Age 3+
  - New to Head Start
  - Enrolled at time of site visit

# Sample Frames

---

- **Head Start Program Information Report (PIR) for programs**
- **List of centers and classes from sites on a rolling basis as recruited and provided**
- **Class rosters from sites on rolling basis, two weeks before site visit**

# Timing and Availability of Frames (First 3 Stages)

---

- **PIR available at time of sampling from 2004-2005 school year**
- **Center and class lists obtained in summer 2006, often before fall class enrollment known**
- **Estimated number of eligible children used as measure of size**
- **Estimates often way off (both directions)**

# **Child Rosters (Stage 4 Frames)**

---

- **Didn't want to get these more than 2 weeks before site visit**
  - **Dynamic nature of class composition**
  - **Especially in first few weeks of school year**
- **Still a number of changes during that 2 week period**
  - **Roster error**
  - **Change since roster**

# Grouping Centers and Classes Before Sampling

---

- More grouping needed than expected
- Center (fewer eligible children per center)
- Class (fewer eligible children per class—often selected all)
- Budgetary implications to have more centers and more classes
- Still ended up with sample shortfall at child level

# Final Fall 2006 Sample Sizes at Program, Center, and Class Levels

---

	Program	Center	Class
Sampled/Released	64	140	415
Eligible	63	135	410
Participating	60	135	410
Target	60	110 to 120	300 to 350
Sum of Weights	1,630	14,148	42,973

# Target Child Sample Over Time (Initial Design)

---

- **Select 3,274 children (fall 2006)**
- **Expected 90% consent and 45/55% split\***
  - 1,326 3yo cohort
  - 1,621 4yo cohort
- **Expected 95% completes among consented in fall 2006**
- **Expected completes in kindergarten year**
  - 766 3yo cohort (spring 2009)
  - 1,171 4yo cohort (spring 2008)

**\*From OMB package in prior round**

# Oversampling 3-Year-Olds (New to 2006 Cohort)

- Goal is to have comparable sample sizes between 2 age cohorts in kindergarten year
- Select **4,051** children (fall 2006)
- Expected 90% consent
  - **2,025** 3yo cohort
  - 1,621 4yo cohort
- Expected 95% completes among consented in fall 2006
- Expected completes in kindergarten year
  - **1,170** 3yo cohort (spring 2009)
  - 1,171 4yo cohort (spring 2008)

# Oversampling Three-Year-Olds

---

- When oversampling, one usually knows the population proportion in each stratum (3yo vs. 4yo cohort)
- This was not known for the current year, and the age mix proportion changing over time
- Used estimates from prior rounds of FACES

# Planned to Oversample 3-Year-Old Classes

---

- Thought most classes were single-age
- Turns out most were mixed-age
- Not an issue in prior rounds (not needed for sampling)
- At class level, ended up using synthetic MOS, giving classes with 3s a little boost in selection probability – but not enough to achieve sample targets

# **Oversampled 3s at Child Level Instead**

---

- **Wanted proportion of initial sample to be 56% 3yo and 44% 4yo (to end up with 50% each in kindergarten year)**
- **Within center, tried proportional allocation by age group first**
- **If  $< 56\%$  3yos, then figured out what was needed to get 56% within center**

# Original and Revised Strategy at Child Level

---

- **Select 20 children per class group**
- **Subsample 10 for main release**
- **Randomly order other 10 to release as needed (replicate of size 1)**
- **Modified after first 20 programs (due to emerging sample size shortfall)**
- **Last 40 programs**
  - **abandoned replicate release process**
  - **if <80 elig. children in program, selected all**
  - **if <40 elig. children in center, selected all**
  - **otherwise, basically same process as before**

# Final Fall 2006 Sample Sizes by Age Cohort

---

	All children in selected classes		Sampled/ Released		Eligible/ Consented		Target Consented	
	n	%	n	%	n	%	n	%
3-yr-olds	2,460	58.2	2,256	59.1	2,017	60.8	2,025	55.5
4-yr-olds	1,765	41.8	1,561	40.9	1,298	39.2	1,621	44.5
Total	4,225	100.0	3,817	100.0	3,315	100.0	3,646	100.0

# Conclusions

---

- **Lessons Learned**
  - **Estimates of class sizes and age mix unreliable before school year starts**
  - **Not sure would do things differently**
- **Difficulties encountered**
  - **Lack of accurate size estimates at each stage**
  - **Quick turnaround and rolling process – can't see the big picture**
  - **Sample allocation based on partial/inaccurate picture can lead to sample sizes over/under target**