All participants being unequal: A bias analysis of three contemporary strategies for locating longitudinal study participants after an extended hiatus

Celeste Stone, American Institutes for Research
Jenny Bandyk, University of Michigan, Survey Research Center

2012 AAPOR Conference
Orlando, Florida
May 20, 2012
Introduction

• Longitudinal study participants cannot participate in future waves if they cannot be found

• Long lags between waves and limited between-wave contacts make tracking participants difficult and costly

• Bias may result if location propensity is systematically related to outcomes of interest
Overview of Project Talent

• Representative 5% sample of U.S. students in grades 9-12 in 1960 (n=373,826)

• Conducted by the American Institutes for Research

• 1960: Students assessed on 940 measures, including:
  ▪ Cognitive abilities and aptitudes, personality, activities, health, family background, future plans

• 1961-1974: Follow-up surveys: 1-, 5-, and 11- years after high school graduation

...  

• Today: Renewed interest in Project Talent as a study of aging and the life course
Statement of problem

- Revitalize Project Talent as resource for studying the influences of early life conditions on aging and the life course
- Locate participants who are now 67-70 years of age and have not been contacted since 1976
  - For most, information needed to facilitate tracking has not been updated since 1960-78
  - 38% have not participated since they were in high school in 1960

1) **Can these participants be located after 37-51 year hiatus?**

2) **Are there any systematic differences in who is located?**
Methods

• Proof-of-Concept: Assess the feasibility of revitalizing study on hiatus since 1974

• Mode: Mail

• Conducted by AIR and the University of Michigan

• Tracking: 2011; Data collection: 2012

• 1% subsample of base year participants (n=4,159)
  ▪ From 130 of the original schools (10%) across 39 states

• Need: Mortality status, Addresses

• Locating strategies:
  ▪ Batch tracking
  ▪ Interactive tracking
  ▪ Outreach activities
Results: RQ1 - Can participants be located?

• Locating rates

Expected mortality rates for this cohort in pop.:
- Men: ~26%
- Women: ~16%
Results: RQ2 – Who is located?

• RQ2: Are there any systematic differences in who is located?
  - Dependent variable: Tracking success (1=Located; 0=Not located)
  - Examined 20 key variables collected in 1960:
    - Selected based on hypothesized relationships with outcomes of interest and previous nonresponse bias analyses
  - Used Chi-square Automatic Interaction Detection (CHAID) technique
RQ2: Differential tracking rates

Key variables examined (20) * = grouped in quintiles

• General (5)
  ▪ Sex, Grade level, High school type and size, Census region, Family socioeconomic status*

• Behavioral (1)
  ▪ Class rank*

• Cognitive ability/aptitude (4)*
  ▪ Verbal scores, Quantitative aptitude, Technical aptitude, Creativity scores

• Personality (10)*
  ▪ Sociability, Social Sensitivity, Impulsiveness, Vigor, Calmness, Tidiness, Culture, Leadership, Self-Confidence, Mature Personality
RQ2: Differential tracking rates

Analytical Technique

- Chi-square Automatic Interaction Detection (CHAID)
- Systematically identify most significant predictors of dependent variable
  - Dependent variable: Tracked/not tracked
- Successively partitions sample into homogeneous subsets (with respect to DV)
- Result: Tree-like structure – “Decision tree”
  - Branches: Variables with smallest p-value
  - Leaves/Nodes
- Specifications imposed:
  - $p \leq 0.05$ and cell size $\geq 35$ observations
CHAIID Results: Differential Tracking rates

Overall
85.4% (4159)

Sex: Male
93.2% (2014)

Tech: Q1
87.1% (186)

Mature: M,Q1-2
92.4% (118)

Mature: Q3-5
77.9% (68)

Tech: Q2-4
92.5% (1114)

Confid: Q1-4
91.3% (866)

Confid: Q5
96.8% (248)

Tech: Q5
95.8% (714)

Rank: M,Q1-3
94.0% (420)

Rank: Q4-5
98.3% (294)

Tech: Technical aptitude
Mature: Mature personality
Confid: Self-confidence
Rank: Class Rank

M: Missing
Q1: Bottom quintile (0-20%)
Q5: Top quintile (80-100%)
CHAID Results: Differential Tracking rates (con’d)

Overall 85.4% (4159)

Sex=Female 78.0% (2145)

Quant: M,Q1 67.4% (585)
Quant: Q2 73.4% (451)
Quant: Q3-5 85.6% (1109)

Impuls: M,Q1-4 69.9% (501)
Impuls: Q5 52.4% (84)
FSES: M,Q1-2 79.0% (290)
FSES: Q3-5 87.9% (819)

Region: 3,4,6,7,9 78.4% (194)
Region: 1,2,5 67.2% (250)
Region: 8 52.6% (57)

Mature: Q1-3 68.4% (38)
Mature: Q4-5 39.1% (46)
Leader: Q1 77.3% (132)
Leader: Q2 90.3% (72)
Leader: Q3 72.1% (86)
Leader: Q4 93.4% (258)
Leader: Q5 95.3% (85)

Schl*: V Sm/Lrg Pub 91.4% (477)
Schl: Other* 83.0% (342)
Tech: Q1-3 77.1% (201)
Tech: Q4-5 91.5% (141)

Social: Q1-4 83.8% (111)
Social: Q5 68.9% (90)

* Schl: Other = Small & Medium Public; Private, Parochial, New York/Chicago Public

Quant: quantitative aptitude
Impuls: Impulsivity
Region: Census region
Mature: Mature
FSES: Family SES
Leader: Leadership
Schl: School type
Tech: Technical aptitude
Social: Sociability

M: Missing
Q1: Bottom quintile (0-20%)
Q5: Top quintile (80-100%)
Summary of CHAID results

• Higher tracking rates for males

• Higher tracking rates for participants with higher cognitive ability scores
  ▪ Technical aptitude (males)
  ▪ Quantitative aptitude (females)

• Differences in tracking rates by personality and behavioral indicators that are conditional on (and differ by) the key cognitive ability scores
Summary of CHAID results (continued)

- Observed differences in tracking rates by personality and behavioral indicators that are conditional on (and differ by) the key cognitive ability scores

<table>
<thead>
<tr>
<th>Cognitive ability group</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowest scores</td>
<td>Mature personality (-)</td>
<td>Impulsivity (-) Region Mature personality (-)</td>
</tr>
<tr>
<td>Middle</td>
<td>Self-confidence (+)</td>
<td></td>
</tr>
<tr>
<td>Highest scores</td>
<td>Class rank (+)</td>
<td>Family SES (+) School type Impulsivity (*)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Technical aptitude (+) Sociability (-)</td>
</tr>
</tbody>
</table>
Summary & Concluding Remarks

• Pretty good tracking rates overall for males, but still missing significant proportion of deaths
• Finding females is particularly challenging
• Significant differences in locating rates by variables likely to be correlated with outcomes of interest
  - Cognitive scores, SES background, impulsivity, confidence, high school achievement → Educational attainment, Economic situation, Employment, Lifestyle
• Regional differences → General mobility, Urbanization, Availability of jobs
• Some types of participants may not be represented as well before data collection even starts
  - Development of other strategies and methodologies for improving representativeness
Next Steps

• Areas for future research
   Conditional on having lower cognitive scores, why would the tracking success be lower for more mature males and females?
   Why might the impact of impulsivity be so uneven for females with particular types of characteristics?
   Other 1960 measures: Health, extracurricular participation, environmental stability, correlates of family/community attachment

• Evaluation of tracking quality – Did we find the correct people?

• Explore relationship between location propensity (difficulty) and response propensity
Questions?

Celeste Stone: cstone@air.org
Jenny Bandyk: jbandyk@umich.edu
Additional Information
CHAID Results: Differential Tracking rates

Overall
85.4% (4159)

Sex: Male
93.2% (2014)

Tech: Q1
87.1% (186)

Mature: M,Q1-2
92.4% (118)

Mature: Q3-5
77.9% (68)

Tech: Q2-4
92.5% (1114)

Confid: Q1-4
91.3% (866)

Confid: Q5
96.8% (248)

Tech: Q5
95.8% (714)

Rank: M,Q1-3
94.0% (420)

Rank: Q4-5
98.3% (294)

Technical aptitude
Mature personality
Self-confidence
Class Rank

M: Missing
Q1: Bottom quintile (0-20%)
Q5: Top quintile (80-100%)
**CHAID Results: Differential Tracking rates (con’d)**

Quant: Quantitative aptitude
Impuls: Impulsivity
Region: Census region
Mature: Mature
FSES: Family SES
Leader: Leadership
Schl: School type
Tech: Technical aptitude
Social: Sociability

- **Overall**: 85.4% (4159)
- **Sex=Female**: 78.0% (2145)

**Quant: M,Q1**
- **67.4%** (585)

**Quant: Q2**
- **73.4%** (451)

**Quant: Q3-5**
- **85.6%** (1109)

**Impuls: M,Q1-4**
- **69.9%** (501)

**Impuls: Q5**
- **52.4%** (84)

**FSES: M,Q1-2**
- **79.0%** (290)

**FSES: Q3-5**
- **87.9%** (819)

**Region: 3,4,6,7,9**
- **78.4%** (194)

**Region: 1,2,5**
- **67.2%** (250)

**Region: 8**
- **52.6%** (57)

**Mature: Q1-3**
- **68.4%** (38)

**Mature: Q4-5**
- **39.1%** (46)

**Leader: Q1**
- **77.3%** (132)

**Leader: Q2**
- **90.3%** (72)

**Leader: Q3**
- **72.1%** (86)

**Schl*: V Sm/Lrg Pub**
- **91.4%** (477)

**Schl: Other***
- **83.0%** (342)

**Tech: Q1-3**
- **77.1%** (201)

**Tech: Q4-5**
- **91.5%** (141)

**Social: Q1-4**
- **83.8%** (111)

**Social: Q5**
- **68.9%** (90)

* Schl: Other = Small & Medium Public; Private, Parochial, New York/Chicago Public

---

**Legends:**
- **≥ 90%**
- **80-89.9%**
- **70-79.9%**
- **60-69.99%**
- **50-59.99%**
- **40-49.99%**
- **< 40%**

---

**Notes:**
- M: Missing
- Q1: Bottom quintile (0-20%)
- Q5: Top quintile (80-100%)