Balancing Cross-sectional and Longitudinal Design Objectives for the Survey of Doctorate Recipients

FCSM Research and Policy Conference

March 9, 2018

Wan-Ying Chang (National Center for Science and Engineering Statistics)
Overview

- Background and motivation
- Sample design questions
- Mine the past survey data
- Findings & next steps
Background

- Survey of Doctorate Recipients (SDR): a biennial survey launched in 1973 to provide demographic, education, and career history information for U.S. research doctorate holders in a science, engineering, or health (SEH) field.

- Prior sample design (fixed panel plus births) was cost effective for collecting cross-sectional data, also generated panel data of various lengths.

- Redesign of SDR 2015 refreshed the entire sample, expanded population coverage, and increased the sample size to target estimation of fine field of study domains. As a result, only 1/3 of the 2013 panel sample was carried forward.
Planning & Outreach

To enhance SDR’s utility and meet dual cross-sectional and longitudinal goals, longitudinal panels within the refreshed sample need to be established formally and maintained over time.

- Outreach to SDR stakeholders to discuss the 2015 SDR sample expansion and initial results (October 2016; February-March 2017)

- Sample design expert panel & outreach emailing (May – September 2017)


- CNSTAT recommendations (January 2018)
Current Cross-sectional Design

- **Sample design**
  - 2015 SDR: stratified on field of study and oversampled women, underrepresented minorities and past panel sample
  - 2017 SDR: replenished with new PhDs sampled at the same rate

- **Questionnaire design**
  To collect employment characteristics on a short reference period

**A1.** Were you working for pay or profit during the week of February 1, 2015?

**E6.** On February 1, 2015, were you living in the United States or Puerto Rico, another U.S. territory, or were you living in another country?
Design Questions

- Does the current sample design embody subsets suitable for panel samples and sufficient for longitudinal analysis?
  - panel definition, sample size, length and frequency of follow up
  - analytical domains and longitudinal estimation reliability requirements

- Does the current questionnaire collect good data for longitudinal analysis?
  - outcomes tracked properly
  - sufficient duration and transition data for modeling longitudinal outcomes
Mine the Past Survey Data

SDR 1993-2013 data are used to construct four longitudinal panels

- 1993-2003 (6 waves, n=12,281)
- 2003-2013 (5 waves, n=15,808)
- 1993-2013 (10 waves, n=7,289)
- 2008-2013 (3 waves, n=23,502)

Methods:
- longitudinal weights created to account for wave nonresponse
- variables harmonized over time
- longitudinal outcomes measured for
  - counts of reported states and events
  - patterns of transition
  - duration

Limitations
Labor Force Status & Employment Outcomes

Labor force status
- Working
- unemployed
- Retired
- Not in labor force (not seeking work and not retired)

Employment outcomes
- employed full time or part time
- employment sector
- occupation group
- tenure status
- job relativeness to doctorate field
- changed job or employer
- received federal government support for work
## Tracking Reported States

Weighted estimates of reported labor force and employment states

<table>
<thead>
<tr>
<th>State</th>
<th>State observed at least once (weighted %)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>For the overall sample</strong></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>97.8</td>
</tr>
<tr>
<td>Unemployed</td>
<td>5.4</td>
</tr>
<tr>
<td>Retired</td>
<td>16.0</td>
</tr>
<tr>
<td>Not in labor force &amp; not retired</td>
<td>4.9</td>
</tr>
<tr>
<td>Any functional limitation</td>
<td>21.1</td>
</tr>
<tr>
<td>Employed part-time (principal job)</td>
<td>19.4</td>
</tr>
<tr>
<td>Residing out of the U.S.</td>
<td></td>
</tr>
<tr>
<td><strong>For those employed at least once</strong></td>
<td></td>
</tr>
<tr>
<td>Worked non-S&amp;E job</td>
<td>48.4</td>
</tr>
<tr>
<td>Received Federal support</td>
<td>46.3</td>
</tr>
<tr>
<td>Job is not related to field</td>
<td>18.3</td>
</tr>
<tr>
<td>Worked supervisory role</td>
<td>83.0</td>
</tr>
<tr>
<td>On tenure track</td>
<td>13.2</td>
</tr>
<tr>
<td>Worked postdoc position</td>
<td>6.5</td>
</tr>
</tbody>
</table>
Tracking Transition of States

Weighted estimates of transition of labor force and employment status

<table>
<thead>
<tr>
<th>Outcome Transition</th>
<th>Transition observed at least once (weighted %)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>For the overall sample</strong></td>
<td></td>
</tr>
<tr>
<td>Labor force status (3 categories)</td>
<td>21.4</td>
</tr>
<tr>
<td>Labor force status (4 categories)</td>
<td>21.7</td>
</tr>
<tr>
<td>Response location (region)</td>
<td>21.6</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
</tr>
<tr>
<td>U.S. citizenship status</td>
<td>6.3</td>
</tr>
<tr>
<td>Residence location (US vs. non-US)</td>
<td></td>
</tr>
<tr>
<td><strong>For those employed at least once</strong></td>
<td></td>
</tr>
<tr>
<td>Employment sector (3 categories)</td>
<td>37.0</td>
</tr>
<tr>
<td>Employment sector (7/8 categories)</td>
<td>44.0</td>
</tr>
<tr>
<td>Job major group</td>
<td>58.1</td>
</tr>
<tr>
<td>Employer location (State)</td>
<td>41.9</td>
</tr>
<tr>
<td>Salary increased</td>
<td>99.3</td>
</tr>
<tr>
<td>Primary activities</td>
<td>69.5</td>
</tr>
<tr>
<td>Changed employer</td>
<td>32.3</td>
</tr>
<tr>
<td>Changed job</td>
<td>38.6</td>
</tr>
<tr>
<td>Tenure track moved to tenured</td>
<td>8.8</td>
</tr>
</tbody>
</table>
Subpopulation with High Prevalence of Retirement

Number of times reported retired by career stage
1993-2003 panel

- Career stage is defined by years since degree time at the start of the panel observation window

- Percent
  - 1
  - 2
  - 3
  - 4
  - 5
  - 6

- Percent

- Overall
- First 5yr
- 5-10 yrs
- 11-20 yrs
- 21-30 yrs
- > 30 yrs

- Career Stage
  - 0
  - 10
  - 20
  - 30
  - 40
  - 50
  - 60
  - 70
  - 80
Subpopulation with High International Mobility

International mobility by citizenship status at degree time

Changes of Residing Location

- U.S. all time
- non-U.S. all time
- U.S. -> non-U.S.
- non-U.S.->U.S.
- Other

Percent

- US citizen
- Permanent resident
- Temp visa holder
- Unknown

- U.S. all time: 96%
- non-U.S. all time: 72%
- U.S. -> non-U.S.: 79%
- non-U.S.->U.S.: 87%
- Other: 11%
Identify Demographic Traits of Transition

Used regression models to summarize demographic traits associated with high likelihood of transition of selected outcomes

Important subpopulation
- Early career
- Physics & Biological sciences
- Women
- Age groups of <30 and >55
Discover Patterns of Transition
Transition of labor force status from 1993 to 2003

Color indicates the 1993 status

- **Green**: Working
- **Blue**: Unemployed
- **Orange**: Retired
- **Pink**: Out of labor force
Labor Force Transitions by Gender

20.7% of female doctorates not employed at least once

7.9% of male doctorates not employed at least once
Estimate Duration

Examine whether sufficient data were collected for estimating duration such as

- **time to event**
  - time to tenured, time to retirement, time to naturalization

- **persistence**
  - duration of employment episodes, spells of unemployment, persistence in sector/job type
Data on Job Start and End Time

- Among those reported working, job start time is asked

- Among those currently not working, job end time is asked

The reported times don’t necessarily correspond to a job and can’t be used to derive the length of a job
Consistency of Time Data

- Reported data for Job start time and year retired from the 2003-2013 panel were used with all imputed data removed

- Data consistency checked
  - For those worked the same job for all waves, 38.4% reported inconsistent year and 46.9% reported inconsistent month
  - For others worked on two consecutive waves, 33.7% reported inconsistent start year among those worked on the same job; 1.6% reported inconsistent start year of those reported changed employers
  - For those reported retired, the reported year last worked doesn’t coincide with the reported year retired, 39.8% of the time
  - Among those reported some data on year retired, 18.9% reported two or more different retirement year

- Need to implement changes to collect better duration data
Findings and Next Steps

- Analysis of the past SDR longitudinal data identified
  - demographic factors that should be considered as stratifying variables
  - small subpopulation with high level of transition should be oversampled
  - major transition patterns can be used to define “event” for longitudinal analysis
  - limitation and issues with duration data

- Next steps
  - Compare longitudinal sample design options and evaluate the impact to the overall sample size over time
  - Improve the questionnaire and data collection methods
Please direct questions and comments to…

Wan-Ying Chang
wchang@nsf.gov

Thank you!