



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)
- Human Resource Expert Panel (August 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

State	State observed at least once (weighted %)			
	1993-2003	2003-2013	1993-2013	2008-2013
For the overall sample				
Employed	97.8	95.8	99.6	92.5
Unemployed	5.4	6.5	9.4	4.6
Retired	16.0	17.1	25.9	13.4
Not in labor force & not retired	4.9	4.5	5.9	3.8
Any functional limitation	21.1	20.4	27.9	13.6
Employed part-time (principal job)	19.4	22.8	30.2	17.1
Residing out of the U.S.				9.8
For those employed at least once				
Worked non-S&E job	48.4	31.9	54.2	26.6
Received Federal support	46.3	48.3	55.4	40.2
Job is not related to field	18.3	17.7	22.8	7.9
Worked supervisory role	83.0	78.2	88.8	68.8
On tenure track	13.2	11.4	16.3	10.6
Worked postdoc position	6.5	3.8	7.0	4.8

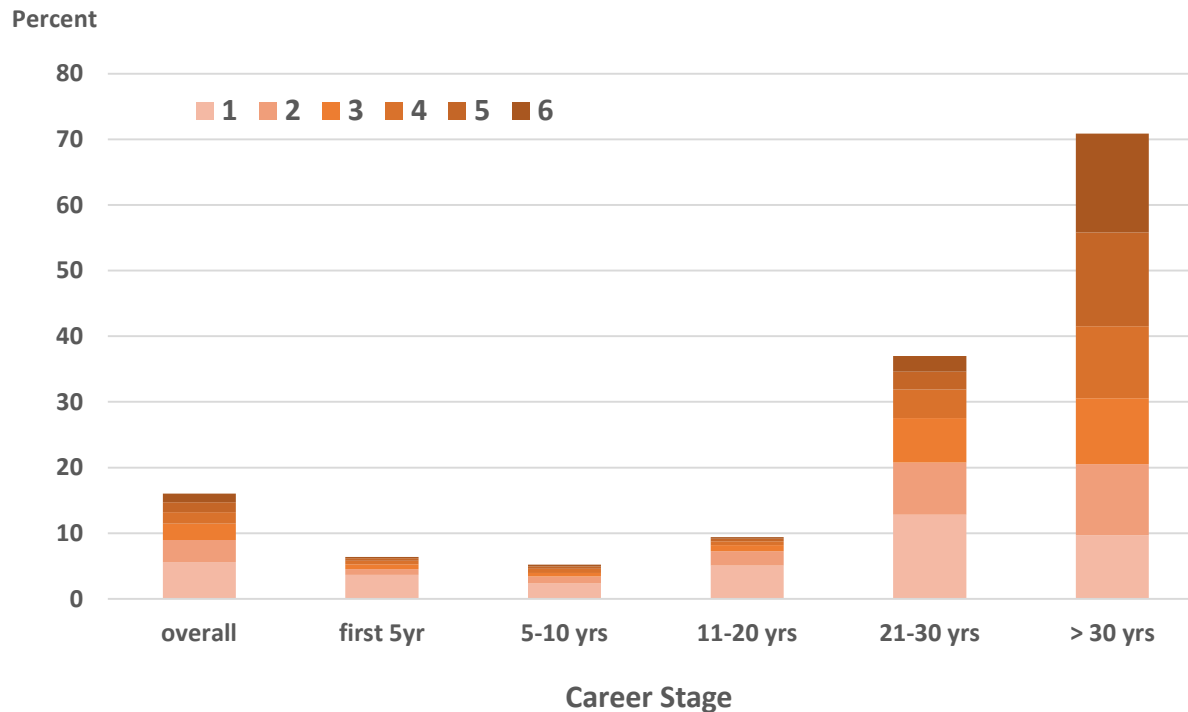
Tracking Transition of States

Weighted estimates of transition of labor force and employment status

Outcome Transition	Transition observed at least once (weighted %)			
	1993-2003	2003-2013	1993-2013	2008-2013
For the overall sample				
Labor force status (3 categories)	21.4	20.9	34.6	12.3
Labor force status (4 categories)	21.7	21.5	34.8	12.7
Response location (region)	21.6	16.7	29.0	14.9
Marital status		18.7		12.7
U.S. citizenship status	6.3	8.7	7.8	11.5
Residence location (US vs. non-US)				3.6
For those employed at least once				
Employment sector (3 categories)	37.0	34.5	52.7	24.5
Employment sector (7/8 categories)	44.0	41.1	59.9	27.4
Job major group	58.1	53.8	73.6	38.7
Employer location (State)	41.9	38.8	58.1	26.7
Salary increased	99.3	99.2	99.9	90.8
Primary activities	69.5	67.0	83.6	50.6
Changed employer	32.3	39.0	44.1	30.4
Changed job	38.6	39.0	50.7	26.9
Tenure track moved to tenured	8.8	7.4	12.2	4.1

Subpopulation with High Prevalence of Retirement

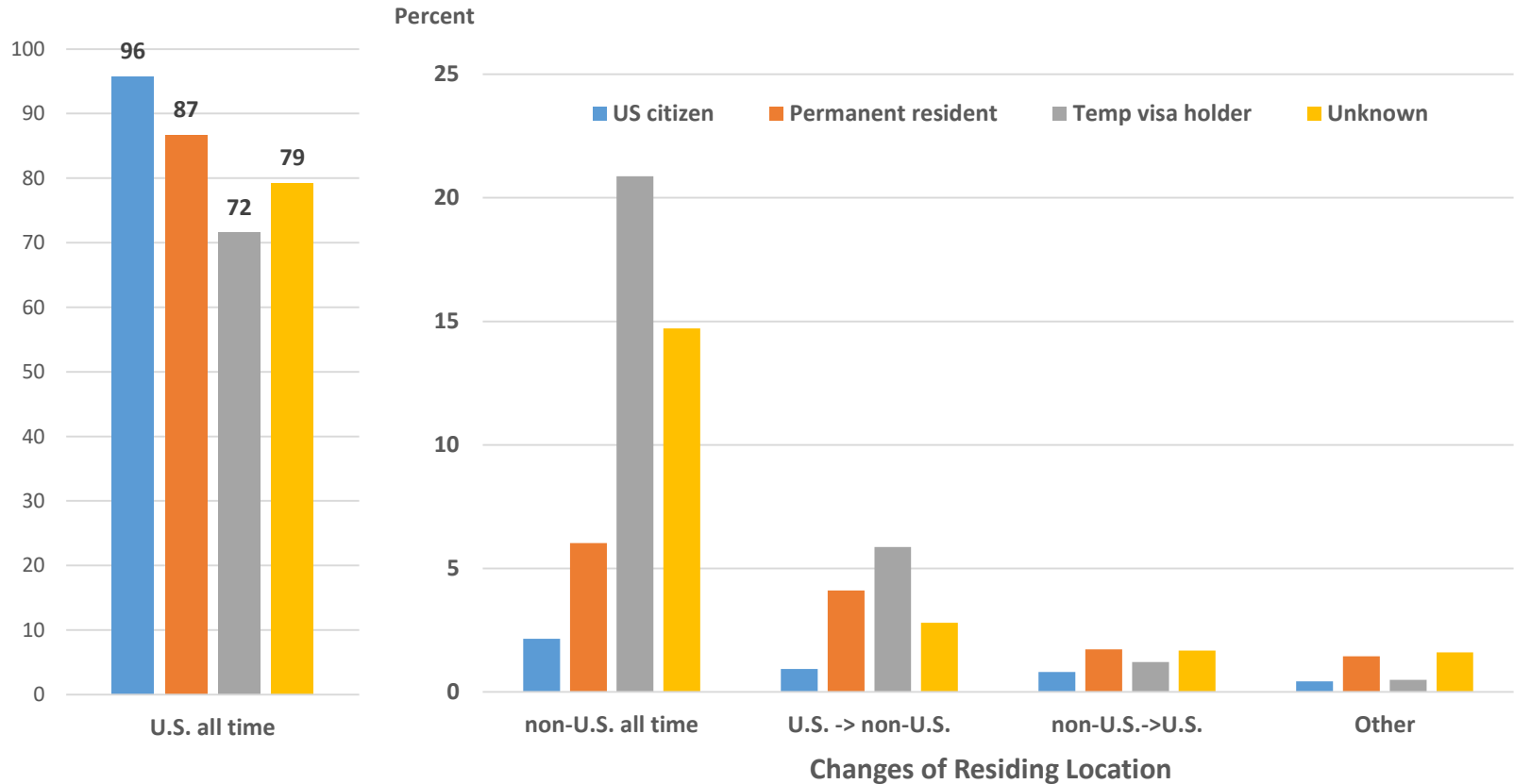
Number of times reported retired by career stage
1993-2003 panel



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

Subpopulation with High International Mobility

International mobility by citizenship status at degree time

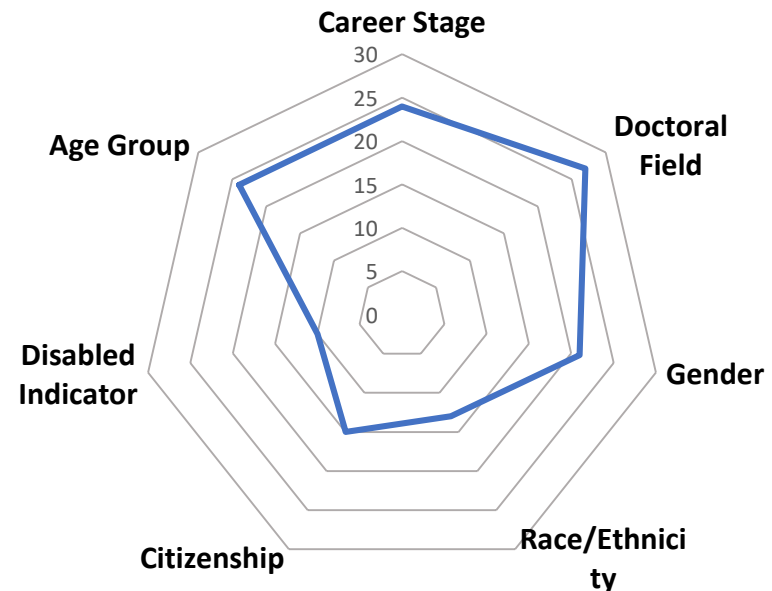


Identify Demographic Traits of Transition

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

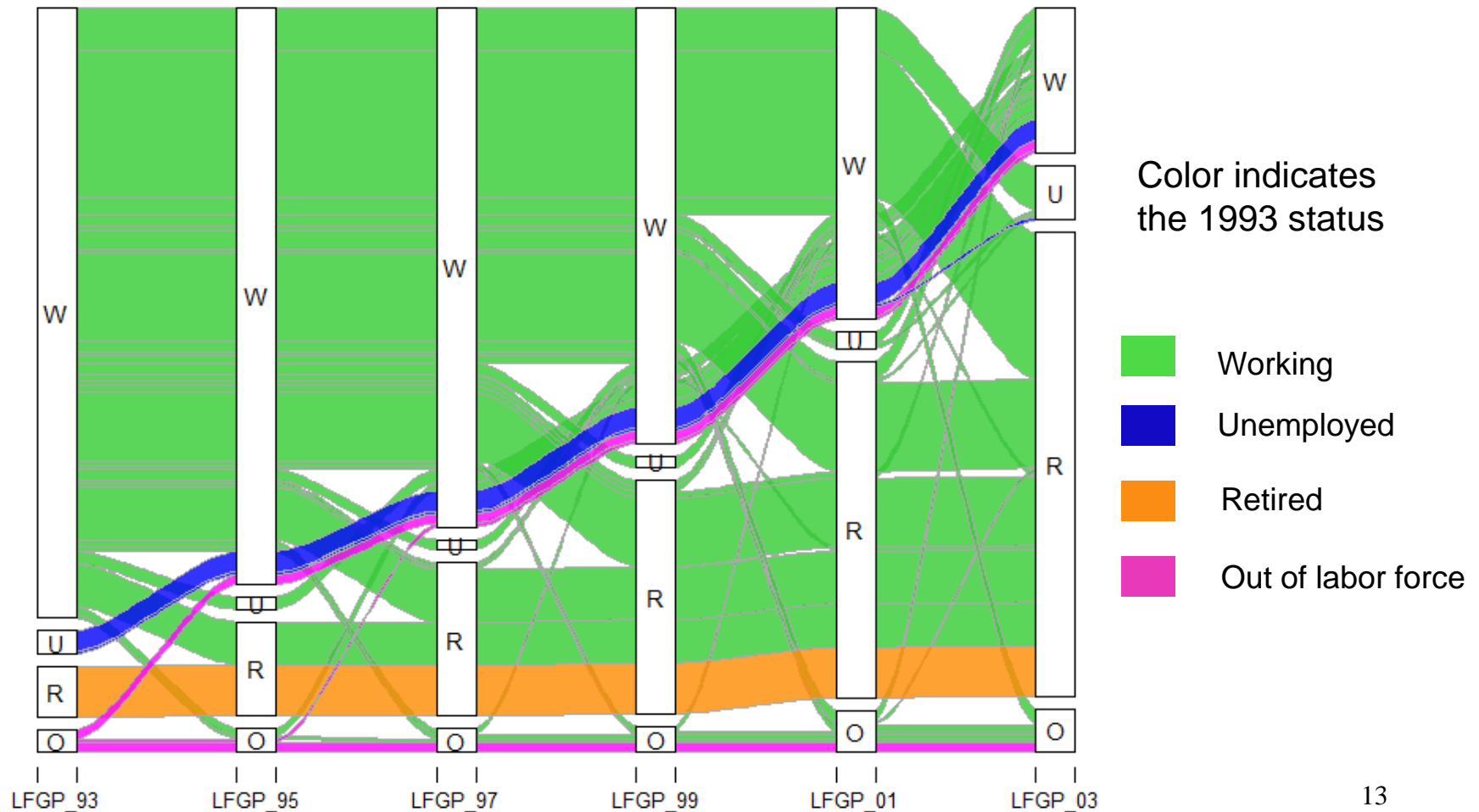
Important subpopulation

- 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

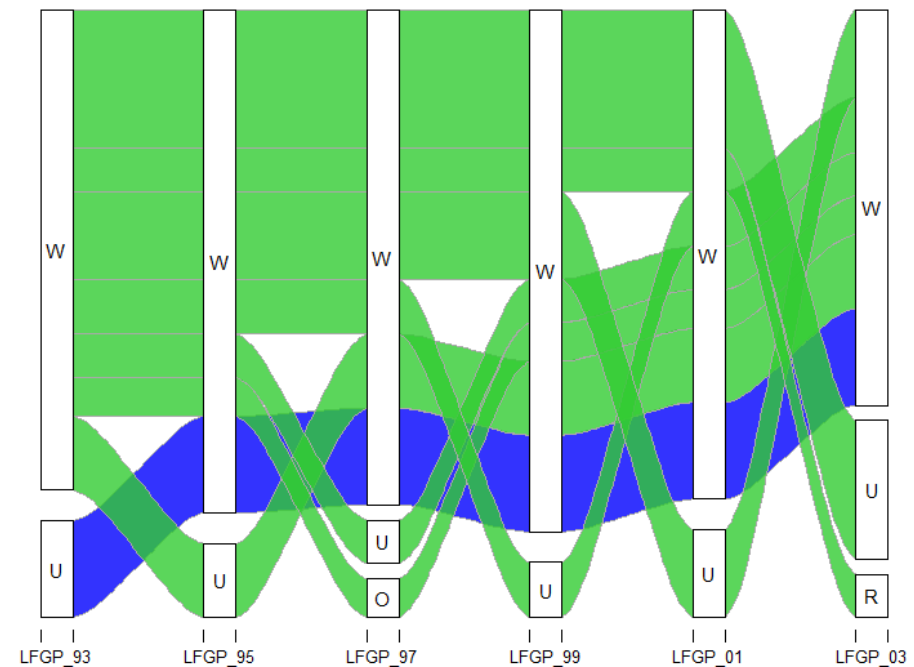
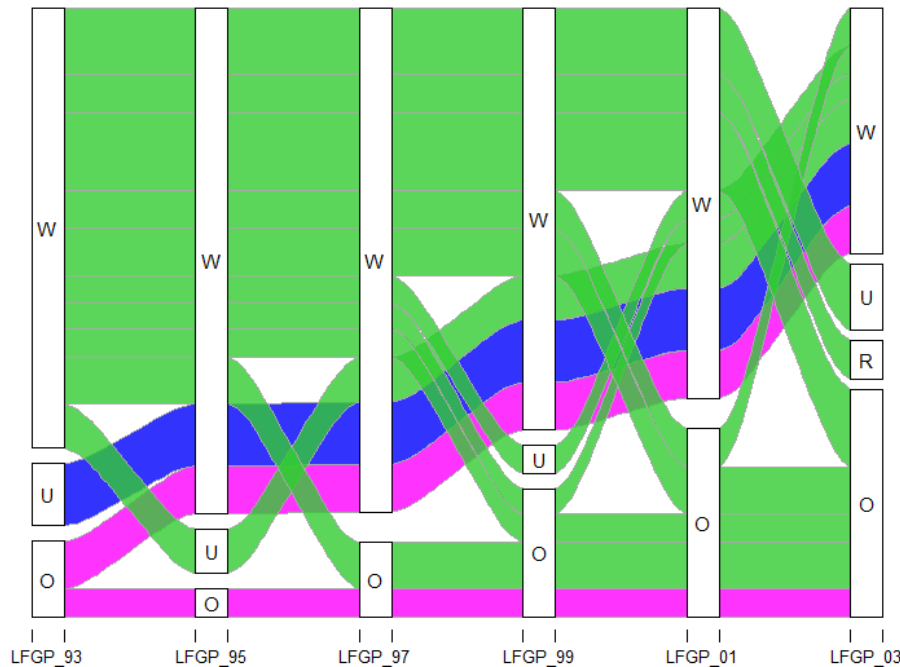


Labor Force Transitions by Gender

Early Career Doctorates (1993-2003)

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

A26. During what month and year did you start this job (that is, the principal job you held during the week of February 1, 2015)?

Month Year

PRINCIPAL JOB STARTED

A4. Prior to the week of February 1, 2015, when did you last work for pay or profit?

← **Mark this box if you never worked for pay or profit and then go to page 10, question D1**

Month Year

LAST WORKED

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!

