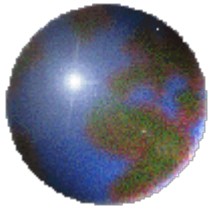
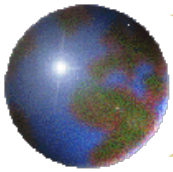


Building a Hitotsubashi Panel Study: Lessons from the U.S. Experience; Prods for Your Design

Clifford Adelman, Institute for Higher Education Policy; 6 March 2013



What will we do today? Review and offer options for:

- U.S. longitudinal studies
- Education to labor market design
- Sampling options and weighting
- Logistics of longitudinal studies
- Labor market variables
- Issues for the release of data and findings

For the U.S., “longitudinal” means *nagai aida*

- Labor Department studies:20 years
- NCES grade-cohort studies: 10-14 years
- NCES event-cohort studies: 6-10 years

The length of the study depends on both the population and what you want to know about them—and why?

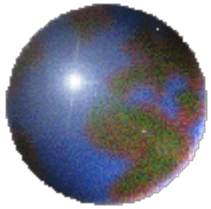
It also depends on how much money you have and are willing to spend of this type of inquiry, and the infrastructure necessary to execute the study. This is true even for an individual university.

NCES Grade Cohort Studies

Study	Beginning Grade	Years	Number of Surveys	Median Age at Conclusion
NLS-72	12	1972 – 1986	5	32
High School and Beyond/So	10	1980 – 1993	4	29
NELS-88	8	1988 - 2000	4	26
ELS-02	10	2002 - 2012	4	26

In all of these cases

- Secondary school and college transcripts are included, which means course taxonomies and student-level analysis files.
- Parents and teachers are interviewed in the base year.
- Sampling is based on the census of schools and school populations, yielding a Stratification Cell and Primary Sampling Unit with which to weight.
- Hitotsubashi does not need any of this.



These studies are not *yasui*

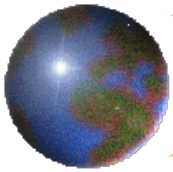
In current dollars, a 12-year longitudinal study with 4 student surveys, and transcripts will cost $80-$100 million, depending on your other specifications.

Completed NCES Postsecondary Event-Cohort studies

Study	Years	Number of Surveys	College Transcripts	Age Range
Beginning Postsecondary Students	1989-1994	3	No	16-68
Beginning Postsecondary Students	1995-2001	3	No	15-76
Beginning Postsecondary Students	2003-2009	3	Yes	15-79
Baccalaureate and Beyond	1993-2003	4	No	18-74

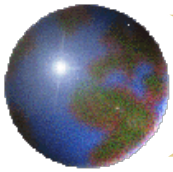
Common and distinctive characteristics

- The sampling source for all of these studies was the National Postsecondary Student Aid Study (NPSAS), an undertaking mandated by law.
- NPSAS is a full universe, hence weighting is determined by participant history in the surveys subsequent to the initial event (starting any kind of college in the BPS; earning a bachelor's degree in the B&B).
- Imputation procedures began only with the BPS 1995-2001.
- College transcripts were collected for a different B&B in 2000, but that study was dropped after one year. A new B&B started in 2008, with completed surveys in 2009 and 2012, but how long it will continue depends on funding. It must be judged, at present, as incomplete.
- None of these studies include secondary school records.



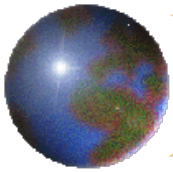
Study design for Hitotsubashi "Education to Labor Market" panel

- **WORK BACKWARDS!!!!**
- **Assume that your universe is 28 or 30 years old.**
- **What do you want to know about their current status?**
- **What do you want to know that brought them to their current status?**
- **Ask those questions first, and then design the study so that it produces the variables you need for analysis.**



Selecting and recruiting the panel

- Initial universe: all entering Hitotssubashi students in "base year"
- Descriptive portrait of this universe (so you know your ultimate sample can be weighted to reflect this universe)
- Ask all students to volunteer during orientation. Tell them they will be receiving Internet surveys for the next 10 years, and that the way they respond will help Hitotsubashi improve what it does for students.
- Assure students of anonymity. They will receive random ID numbers to use throughout their participation.
- Assess the signed volunteer forms you receive. You know you will lose some of these people. How many do you think you will need at the beginning?



Weighting the panel

- Since you will not have everybody in the entering class, your volunteer sample must be weighted to match the initial universe.
- So what variables would you use in the weighting procedure?
- Future weightings will be by panel participation: 4-year, 7-year, 10-year, so, for example, 1+4+10 participation weight.

Panel study: what do you want to know at the beginning?

- Secondary school academic records (Since Japan cannot get these, ask the students what they studied)
- Family composition, levels of parental education which is far more powerful than income!!!
- Intended future academic path: major field, graduate study.
- Intended future career. How would you ask?
- What else? What else do you ask now?

Grade Cohort Study: tracking for future surveys and unobtrusive data

- **Parent as the default for student's location;**
- **Student's e-mail address;**
- **Student's social media addresses;**
- **Student's mobile phone number;**
- **Commercial locating systems.**

These are part of the student's agreement to participate

For survey engines: Internet, telephone, in person.

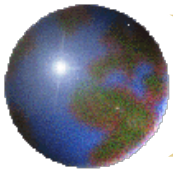
- Internet is the preference for students in all surveys.
- Internet is also preferred for parent interviews in the base year, but if no response, one can follow by telephone.
- Internet requires more careful advanced preparation, but execution is cheap. Telephone and in-person interviews are obviously more expensive to execute, but offer greater flexibility.

Undergraduate records and experience variables

- Are you able to get official undergraduate records (of courses, grades, GPA) from the University?
- If not, then the student must provide the answers in year 4.
- If student provides the answers, what variables do you want to see (you may be getting these already)?

Then, labor market experience in years 7 and 10: What do you want to know?

- Primary point of labor market entrance
- Monthly (or annual or quarterly) primary employment history: employed, unemployed, in-school, other. From this, Continuity of Primary Employment and continuity of *all* employment.
- For employment periods, average hours per week (in a range)
- Contiguity of primary employment, that is, to what extent is the job related to the student's primary field of study?



What do you want to know? Primary Employments Only, Part 1

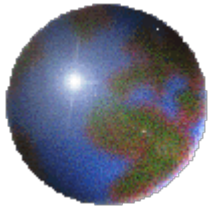
- Occupation (recommended taxonomy: about 50)
- Duties (short answer open-ended question)
- Industry (recommended taxonomy: about 25)
- Principal tools/software/etc. used in the primary position (see ONET)
- Nature and extent of employer-provided training

What do you want to know? Part 2

- Annual wages in the most recent year
- Annual total income in the most recent year
- Satisfaction with primary employment/career path
- Assessment of opportunities for advancement in primary employment/career path
- Assessment of need for further education and/or training
- Intention to change either primary employment or location of employment

Then, it depends what else you want to know, for example:

- Secondary employment, second jobs, internships while enrolled in school or college.
- Family status (definitely, “yes”!): marriage, partners, children, whether spouse/partner is also employed.
- Student assessment of the extent to which the primary workplace encourages innovation, creativity, entrepreneurialism
- A separate question stream for those who are self-employed (for example, artists, small business owners, consultants, etc.)
- A separate question stream for those who have been in the same occupation for 5-6 years, with attention to observes changes in skill and knowledge requirements



You may be doing all these things today---I do not know, but you will tell me.

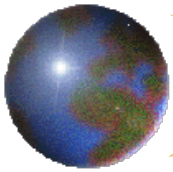
. . .but you may not, and our discussion should bring light to the positive and negative aspects of longitudinal design.

Release of data in 2 forms: public aggregate data and restricted. Require a license for the use of raw data (restricted)

- Even for University and Ministry employees
- And the license stipulates the same penalties for disclosure as noted for the external contractor.
- Produce public use versions of the data set composed entirely of aggregate derived variables, and following each of the scheduled surveys. There are no raw data in these releases.

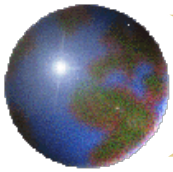
Examples of public and restricted data

- **Public: percentage of former students who had children during their first spell of employment: by gender, by undergraduate field of study, by type of industry in which former student was employed.**
- **Restricted: creation of an index of the academic intensity of undergraduate experience (credits in mathematics, extra enrollment sessions, culminating project, etc.) by independent researchers.**



Let me stop here, and listen to you.

- I am asking you, first, to work backwards, and tell each other what you want to know at the projected end of this study, when your subjects are 30 years old.
- Then build lists of education, labor market experience, and allied variables, that would help you answer those questions, and
- Then determine what survey questions would be asked at different stages of the longitudinal study, and at what points unobtrusive data would enter.



And as you ask your questions and build your lists, remember:

- **The data sets you produce will have a "descriptive window" for every variable that tells the user (along with the numerical distribution of values, whether continuous or categorical) how that variable was constructed, and provides advice on its use.**