

## Perspectives on Longitudinal Surveys

Randall J. Olsen

In discussing longitudinal studies, by far the most difficult problem we need to solve is exactly how to pronounce longitudinal. Janet, being a right thinking person and thoroughly American will, I am confident, provide excellent instruction on this point whereupon we can move to less difficult and contentious matters.

The National Longitudinal Surveys started in 1966 with the intention of following the respondents for five years. These respondents were drawn from different birth cohorts, young men, young women, middle-aged women and men nearing retirement, but frequently from the same families. The first four cohorts in Table 1 all started in the mid to late 1960s.

Table 1.

*The NLS: Survey groups, sample sizes, interview years, and survey status*

Survey group	Age cohort	Birth cohort	year	Original sample	Initial year/ latest year	Number of surveys	Number at last interview	Status
Older Men	45–59	4/1/06–3/31/21	1	5,020	1966/1990	13	12,092	Ended
Mature Women	30–44	4/1/22–3/31/37	1	5,083	1967/2003	21	2,237	Ended
Young Men	14–24	4/1/41–3/31/52	1	5,225	1966/1981	12	3,398	Ended
Young Women	14–24	1943–1953	1	5,159	1968/2003	22	2,859	Ended
NLSY79	14–21	1957–1964		12,686	1979/2004	21	<sup>3</sup> 7,724	Continuing
NLSY79 children	birth–14	—		<sup>4</sup> —	1986/2004	10	<sup>3</sup> 3,190	Continuing
NLSY79 young adults	and older	—		<sup>4</sup> —	1994/2004	6	<sup>3</sup> 4,238	Continuing
NLSY97	12–16	1980–1984		3,984	1997/2004	8	<sup>6</sup> 7,756	Continuing

<sup>1</sup> Interviews in 1990 also were conducted with 2,206 widows or other family members of deceased respondents.

<sup>2</sup> After dropping the military (in 1985) and economically disadvantaged nonblack/non-Hispanic oversamples (in 1991), the sample contains 9,964 respondents eligible for interview.

<sup>3</sup> The latest sample size available is from the 2002 survey.

<sup>4</sup> The size of the NLSY79 child sample depends on the number of children born to female NLSY79 respondents, attrition over time, and the gradual aging of the children into the young adult sample. The size of the young adult sample depends on the number of children who reach age 15 in each survey year. Information about the number interviewed in each survey is available in chapter 4.

<sup>5</sup> In 1998 only, the young adults eligible for interview were limited to those ages 15 to 20.

<sup>6</sup> The latest sample size available is from round 7.

The usefulness of longitudinal data became apparent, leading the Department of Labor to not only continue the effort, but expand it, first in 1979 with the addition of a new youth cohort, in 1986 with the Children of the NLSY and again in 1997 with another new youth cohort. For example, the original cohorts of the NLS were started to help understand why youth unemployment was so high (at least by the standards of the 1950's). What followed was a variety of reports that documented the differences in

labor market turnover that foreshadowed labor market success. That is, some youths had careers that churned among low skill jobs with little earnings growth whereas others exhibited trajectories of steadily improving job descriptions and earnings. For the cohort of adult women, the study documented the dynamics of their children growing older and the mothers making a transition back into the labor force. For the older men, the issue was early retirement, and the surveys revealed the interplay of health, income, family composition and wealth in the decision to retire. The early years of the NLS corresponded to a time when racial differences in labor market outcomes attracted both policy and academic interest. Some of the early reports from the NLS focused on how factors such as health, veteran's status, education and training had differential impacts on Blacks and whites and how their work lives changed over time.

By now, the NLS bibliography has over 5000 items and is growing at about 250 items per year. Given this number, speculating on what are the most significant findings will likely offend a lot of people. Instead, let us focus on the data elements that have been more useful to researchers and the design elements that were central to those elements. One of the distinguishing characteristics of the NLS data is its use of event history data.

## **1. Event Histories**

My reading of the history suggests that Burt Singer made a very strong case for the collection of data in event history format for the NLSY79. From the outset, the NLSY79 was designed to have event histories on employment, schooling, training, fertility and marriage. For all but schooling, we used bounded interviewing, that is, reminding the respondent where they stood with respect their employment status, etc. at the previous interview, obtaining either conformation or correction, and proceeding to extend the event history. When the survey was permanently shifted to computer assisted interviewing in 1993, we also collected the program reciprocity data in event history form and implemented bounded interviewing for the schooling event history. The NLSY79 followed a complementary field strategy of returning to non-respondents in subsequent years no matter how many rounds they had missed, unless of course they had to be dropped from the study due to hostility<sup>1</sup>. For the original cohorts of the NLS the Census Bureau had a policy of dropping respondents who had not been interviewed for two years – an inferior policy that we finally convinced them to change.

Now, returning to respondents to pick up the event history from where ever it was left off required careful design and of the instrument, careful handling of the “information sheet” items that drive the bounded interviewing, and careful handling of the answer file so the public use file released these data in a usable format. Usable and simple to use can be two very different things.

---

<sup>1</sup> We have had several respondents' families claim he or she had died only to have interviewers document a remarkable improvement in health. Now we seek documentation of death.

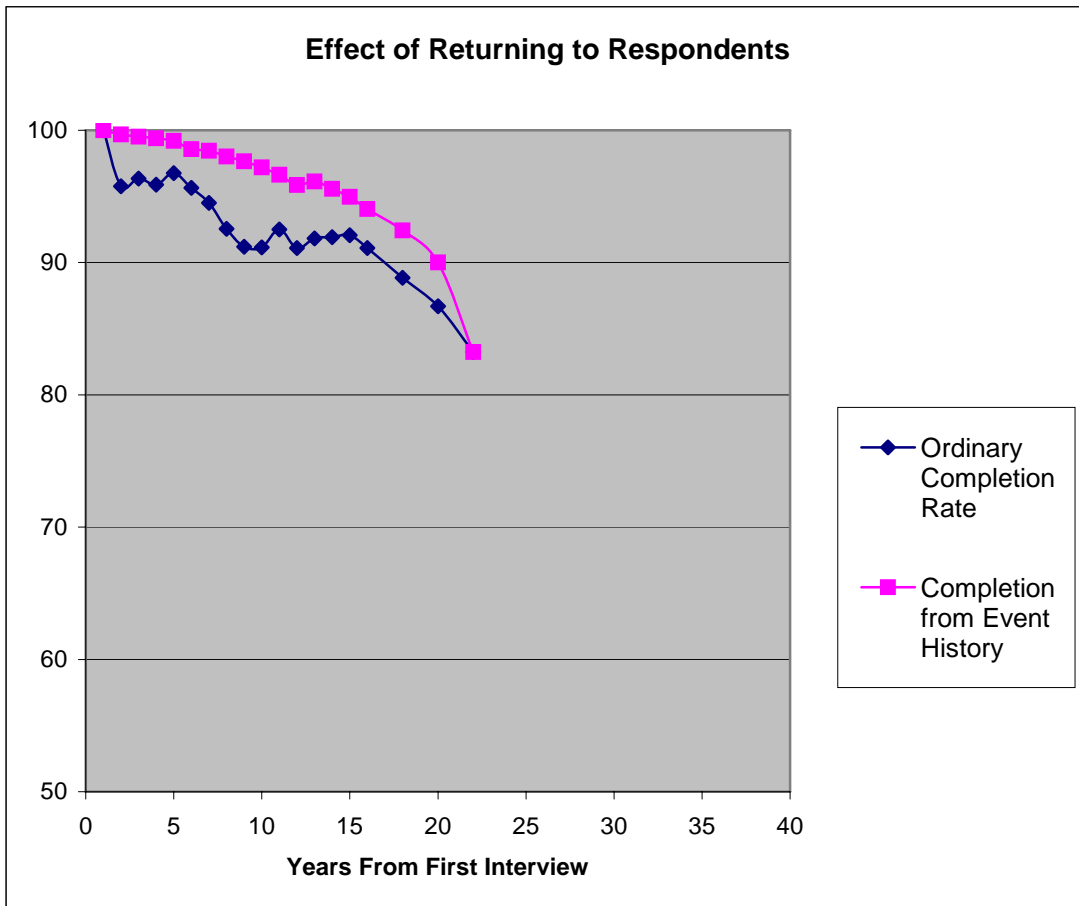


Figure 1 – Completion Rates and Bounded Event Histories for NLSY79

Figure 1 above shows the impact of utilizing an event history in conjunction with returning to past non-respondents. The darker, lower line shows the fraction of in-scope respondents interviewed at various years after the first interview in 1979 – often referred to as the completion rate. Interviews were yearly through 1994 and after that only occurred in even-numbered years. The lighter, higher line shows the fraction of the event history data from each year that we ultimately recovered from the respondent. That is, information from the tenth year after the baseline could be recovered either in the tenth round of the survey or in any round after the tenth due to the “since the date of the last interview” design. Looked at in this way, the high response rates obtained by NORC in the NLSY79 are, if anything, misleadingly pessimistic when it comes to the core areas of the survey that we support with event histories.

Tables 2 and 3 below provide a more detailed look at completion rates in the NLSY79 and NLSY97 and document how completion rates have fallen even though the broad designs of these two cohorts are very similar.

Table 2.

NLSY79 response rates (excluding deceased respondents) by sample type

Year	Cross-sectional sample			Supplemental sample			Military sample			Total sample		
	Response Total	# deceased	rate <sup>1</sup>	Response Total	# deceased	rate <sup>1</sup>	Response Total	# deceased	rate <sup>1</sup>	Response Total	# deceased	rate <sup>1</sup>
1979	6,111	—	—	5,295	—	—	1,280	—	—	12,686	—	—
1980	5,873	4	96.2	5,075	5	95.9	1,193	0	93.2	12,141	9	95.8
1981	5,892	15	96.7	5,108	14	96.7	1,195	0	93.4	12,195	29	96.3
1982	5,876	24	96.5	5,036	19	95.5	1,211	1	94.7	12,123	44	95.9
1983	5,902	27	97.0	5,093	26	96.7	1,226	4	96.1	12,221	57	96.8
1984	5,814	30	95.6	5,040	33	95.8	1,215	4	95.2	12,069	67	95.6
1985	5,751	36	94.7	4,957	43	94.4	<sup>2</sup> 186	0	92.5	10,894	79	94.5
1986	5,633	43	92.8	4,839	51	92.3	183	1	91.5	10,655	95	92.6
1987	5,538	51	91.4	4,768	56	91.0	179	3	90.4	10,485	110	91.2
1988	5,513	56	91.0	4,777	68	91.4	175	3	88.4	10,465	127	91.2
1989	5,571	60	92.1	4,853	78	93.0	181	3	91.4	10,605	141	92.5
1990	5,498	67	91.0	4,755	82	91.2	183	3	92.4	10,436	152	91.1
1991	5,556	75	92.1	3,281	65	91.5	181	4	91.9	<sup>3</sup> 9,018	145	91.8
1992	5,553	81	92.1	3,280	71	91.6	182	4	92.4	9,015	156	91.9
1993	5,537	90	92.0	3,293	83	92.3	181	4	91.9	9,011	177	92.1
1994	5,457	104	90.8	3,256	96	91.6	178	4	90.4	8,891	204	91.1
1996	5,290	129	88.4	3,171	109	89.5	175	5	89.3	8,636	243	88.8
1998	5,159	152	86.6	3,065	118	86.7	175	5	89.3	8,399	275	86.7
2000	4,949	170	83.3	2,921	136	83.1	163	7	84.0	8,033	313	83.2
2002	4,775	188	81.2	2,792	151	80.6	157	7	81.6	7,724	346	80.9
2004	4,686	221	79.6	2,818	193	81.0	157	7	80.9	7,661	421	80.1

<sup>1</sup> Response rate is defined as the percentage of base-year respondents remaining eligible and not known to be deceased who were interviewed in a given survey year.

<sup>2</sup> A total of 201 military respondents was retained from the original sample of 1,280; 186 of the 201 participated in the 1985 interview. The total number of NLSY79 civilian and military respondents eligible for interview (including deceased respondents) beginning in 1985 was 11,607.

<sup>3</sup> The 1,643 economically disadvantaged nonblack/non-Hispanic male and female members of the supplemental subsample were not eligible for interview as of the 1991 survey year. The total number of NLSY79 civilian and military respondents eligible for interview (including deceased respondents) beginning in 1991 was 9,964.

Table 3.

NLSY97 sample sizes, retention rates, and fielding periods

Round	Fielding period	Cross-sectional sample		Supplemental sample		Total sample	
		Total	Retention rate	Total	Retention rate	Total	Retention rate
1	February–October 1997 and March–May 1998	6,748	—	2,236	—	8,984	—
2	October 1998–April 1999	6,279	93.0	2,107	94.2	8,386	93.3
3	October 1999–April 2000	6,173	91.5	2,036	91.1	8,209	91.4
4	November 2000–May 2001	6,055	89.7	2,026	90.6	8,081	89.9
5	November 2001–May 2002	5,919	87.7	1,964	87.8	7,883	87.7
6	November 2002–May 2003	5,899	87.4	1,999	89.4	7,898	87.9
7	November 2003–July 2004	5,783	85.7	1,973	88.2	7,756	86.3
8	November 2004–June 2005 <sup>1</sup>	2__	2__	2__	2__	2__	2__

NOTE: Retention rate is defined as the percentage of base-year respondents who were interviewed in a given survey year; deceased respondents are included in the calculations.

<sup>1</sup> This is the planned fielding period.

<sup>2</sup> The sample sizes for round 8 are not yet available.

An important issue in labor economics is the effect of tenure with a firm versus experience, which can be had with any firm, on the rate of pay. This goes to the heart of the concept of human capital, and especially specific human capital, which is skill accrued with an employer that is valued more highly by the incumbent employer than elsewhere in the labor market. Studying the role of specific human capital requires

accurate data that identifies for whom each person has worked over their entire working life. When it comes to understanding how the market rewards skills, characteristics and experience, research is indivisible. That is, if one has erred in determining the part of earnings due to specific or general human capital, one is at substantial risk of mis-estimating other important magnitudes, such as the return to education, Black-White earnings gaps, or the union-nonunion wage differential. Earnings from work dominate economic magnitudes and deserve all the study they receive. The NLS data have been especially powerful for studying labor force dynamics in large part because of the scope and completeness of the employment event history, which goes back to basic design decisions.

Turning back to the lower line in Figure 1, it may be useful to say a few words about how that line got to be where it is and why it looks different from Figure 2 below, which shows similar information from the Young Women's cohort of the NLS – a study began in 1968 when the women were at similar ages as the NLSY79 when it was started. The lower line for the Young Women falls at a faster rate overall even though it also had a high completion rate early on. The Census Bureau collected the data for the Young Women and it has real advantage at “getting the foot in the door” inasmuch as far more people have heard of the Census Bureau and know what it does than is the case for NORC.

First of all, returning to lapsed respondents accounts for much of the difference. Virtually anyone can pick out the year when we persuaded Census to return to past non-respondents, although by then many of the past non-respondents were unlocatable with the passage of time. Second, the length of the field period matters. The Census Bureau has responsibility for so many surveys that each one has a time slot. When it comes to persuading respondents, sometimes it simply takes time. A longer field period albeit at a lower intensity allows a lot of respondents to be converted. Finally, respondent fees. My rule of thumb is that 50-60% of the respondents will do a new wave of a longitudinal survey no matter what. Another 10% will not do that survey wave no matter what. The rest of the respondents are for sale and the price, and unit of payment, is highly variable. Sometimes the price is in small gifts, sometimes bending to idiosyncratic respondent desires about how and when the interview is to be done. Other respondents can be “sold” on the survey by being persuaded that they are part of something larger and more important. After all, often people with Ph.D.s can be recruited to highly prestigious universities or government agencies at a lower price than at organizations with substantially inferior reputations. Why should we suspect respondents are any different? Good respondent materials and a skilled approach to a respondent can cater to an otherwise reluctant person's ego with some success.

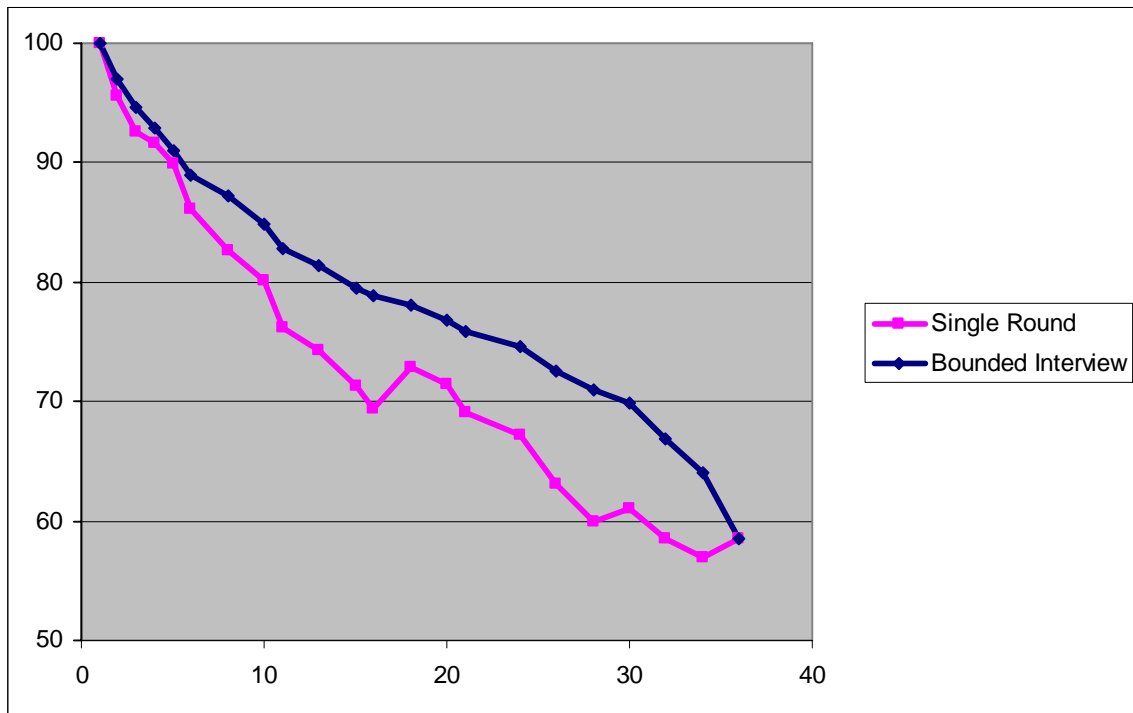


Figure 2 – Completion Rates and Bounded Event Histories for Young Women’s Cohort

However, for some respondents the price is money, but even here the price is not uniform. Luckily for longitudinal surveys, repeated contacts with a respondent frequently gives us information as to what price may be necessary to convert a non-respondent. That is not to say that a non-extortionate fee can convert anyone, but the judicious application of respondent fees can raise response rates from five to ten percentage points. In fact, frequently response fees, judiciously applied, can *reduce* costs. Ironically, survey programs often spend large amounts of time on callbacks, personal visits, respondent materials and other high-cost ways of converting reluctant respondents but they can be unwilling spend less by offering the money to the respondent instead.

Survey funders or survey staff can find it offensive that respondent fees end up “rewarding” respondents for poor behavior. Sometimes people in the survey business prefer to spend more to get less rather than reward people who are motivated by different things than people whose careers are centered on public service. A commitment to recovering a highly complete event history spans everything from survey design to a tolerance for diversity in what motivates people. A strategy of paying all respondents the same fee will likely make it prohibitively expensive to pay the sorts of fees that will convert reluctant respondents in the longitudinal survey. When our studies employ unequal respondent fees, we pay the same to all persons in a family unit to prevent invidious comparisons. Fees are not only a cost effective tool to encourage survey response, but they can actually reduce the cost of converting difficult respondents. Government agencies are often reluctant to take such measures, but this is a straightforward case of which the sponsor wants more: higher response rates or intact sensitivities.

While we are on the subject of making fielding practices cost-effective, let me make a brief aside on the implications of the wide-open spaces of Canada. For the past four years we have pursued a strategy that enables us to utilize interviewers in remote locations as if the interviewers were in a virtual telephone shop. By giving them a broadband connection, interviewers in Hartland, New Brunswick (home of over 900 people and the world's longest covered bridge) operate as if they were sitting in a brick and mortar call center in Ottawa. We use a combination of Voice over IP and a web-based survey system to replicate all the capabilities of a call center. This approach would allow Statistics Canada to keep remote interviewers busy with conventional telephone work between surveys that require a local presence. If Statistics Canada does not have this capability it ought to consider the potential this approach would have for making its data collection operations more efficient.

## **2. Child Development**

One of the more influential components of the NLS program has assessed the biological children of the female respondents in the 1979 youth cohort as a child development survey. This effort, which started in 1986 and is led by Frank Mott at CHRR, demonstrated the power of longitudinal surveys of children as opposed to efforts that follow pupils in a narrow set of age ranges for just a few years.

A couple of features were central to the success of this study. First, because we followed the mothers since 1979, we had very comprehensive information about the mother and the family unit. This included a comprehensive battery of cognitive tests administered to most respondents in 1980. The mother's behaviors and attributes explain a large fraction of the variation in child development so any study of children that only pays cursory attention to the mother makes a serious error. A study of children that does not collect extensive data on the mother is a mis-allocation of resources. The mother is the 800 pound gorilla of child development.

Tables 4-7 below provide a brief illustration of the point. Table 4 gives variable definitions. All regression results use OLS. Table 5 shows the factors explaining variation on the Digit Span Memory test at age 4 - a test that is a good "ability" test in that race effects tend to be smaller than for other tests and relies on memory rather than exposure to vocabulary in the home - shows maternal characteristics matter a lot. Table 6 provides the determinants of a child's early Peabody Picture Vocabulary Test (PPVT), usually given when the child is about four. The metric for scores is percentage points, from zero to one hundred. For every five percentage points a mother's AFQT word knowledge subscale increases, her child gains about one point on the PPVT. Mother's word knowledge has a slightly smaller, but still significant, effect on Digit Span scores. The mother's education also has a substantial impact. One year of education for the mother and 10 percentile points on her AFQT work knowledge score is worth about \$10,000 per year of family income (1999 dollars) in terms of their relative effects on early child outcomes. Table 7 looks at the child's PIAT Reading Comprehension percentile ranking both at the initial test, which is about age 6 (first grade) and also the yearly gain between the first and last test (at about age 14). Unsurprisingly, children who score well

as preschoolers on the Digit Span and PPVT score well when they enter school - the effects are large and significant. Moreover, the effects of the child's early PPVT score and the mother's word knowledge score are significant factors in moving some children up their peer ranking in reading comprehension. Not only do the advantages at age four lead to a higher ranking in first grade, but the advantage continues to grow through age fourteen. It cannot be too much of a surprise that if the mother matters in early childhood she continues to matter as the child continues on into middle school.

These findings have been echoed over and over as analysts look at the Children of the NLSY. Recently researchers using these data to study child obesity found a similar impact - maternal obesity before pregnancy foreshadows obesity for the child. The lesson here appears to be that we should design child studies in a way that collects comprehensive data on the mother so we avoid attributing maternal effects to other variables.

The second important feature of the Children of the NLSY was following *all* the children of a woman. Once we have data on all a child's siblings it opens up a variety of analytic opportunities. Maternal and family effects are important to social, emotional and cognitive development. Being able to look at siblings and ask which are the factors that differ across siblings and contribute to different outcomes allows us to isolate what is important and what is not. While assessing all the children in a family raises burden, studies of child outcomes have a lot of face validity to respondents and interviewers can make a good case justifying the burden to the family.

It turns out that my wife has been working the other side of this same street for even longer than I have. She has taught kindergarten and first grade for longer than she wants me to say. A staple of our conversations over the years has been a comparison of what she has learned by watching children in the classroom and getting to know their parents versus what the data in the Children of the NLSY have shown. Time and time again, the children in her classrooms conform to the pattern above, although she doesn't test the parents, at least in ways they ever recognize. When quantitative analysis matches up with decades of qualitative assessments we ignore these lessons at our peril. Looking for effects of vouchers, choice, bussing, class size and so forth are good sport. Unfortunately the dirty little secret is that these factors are most likely dwarfed by the mother and the home - of course that is precisely the message that, possibly for political reasons, no one wants to deliver.



Table 4

<u>Variable</u>	<u>Definition</u>
Age Mom at Birth	The age of the mother at the birth of the child. We anticipate that delays in child bearing will inure to the advantage of the child. This cohort of women is not old enough to encounter the higher risks of births over forty.
Black	A value of one, zero otherwise, indicates the mother's race is Black. In the NLSY79 Black Hispanics are counted as Hispanic and not Black.
Hispanic	A value of one, zero otherwise, indicates the respondent self-identified as Hispanic.
Mom HGC	Highest grade completed by the mother as of 1986
Sex	A value of two, one otherwise, indicates the child is female.
Birth Order	Birth order of the child relative to other siblings borne by the mother. We anticipate an advantage to the first born as this child does not have to share the mother's attention with siblings, at least for a while.
AFQT-AR	The arithmetic reasoning component of the Armed Forces Qualifying Test score. Expressed as percentile points from zero to one hundred.
AFQT-WK	The word knowledge component of the Armed Forces Qualifying Test score. Expressed as percentile points from zero to one hundred.
AFQT-PC	The paragraph comprehension component of the Armed Forces Qualifying Test score. Expressed as percentile points from zero to one hundred.
AFQT-NO	The numerical operations component of the Armed Forces Qualifying Test score. Expressed as percentile points from zero to one hundred.
Income	Total net family income in tens of thousands of dollars
Mother's Hours	Mother's hours of work averaged over the relevant time interval, expressed as a fraction of a 35 hour work week
Never Father	Dummy variable equaling one if the father was never in the child's household over the relevant time interval
Handicap	Dummy variable indicating the child has a physical or mental condition affecting either attendance at school or school work
Father Presence	percent of years the father was absent from the household over the relevant time interval, ranges from zero to one
In School	Percent of years the child was in school over the relevant time period
Private School	Percent of years the child attended a private (religious or other) school over the relevant time period. This variable is in addition to the "In School" variable and so shows the incremental impact of being in a private school over and above being in school at all.
Home Scale	A percentile scale rating the home environment for factors such as neatness, the presence of reading material, and so forth. Ranges from zero to one hundred.

*Table 5*  
*Early Digit Span Percentile Ranking on Memory*

<u>Variable</u>			
Constant	28.39 (5.96)	24.46 (5.04)	49.28 (28.81)
Mom Age at Birth	-0.16 (-0.87)	-0.10 (-0.53)	
Black	1.45 (1.04)	2.86 (2.01)	-2.94 (2.81)
Hispanic	-3.42 (-2.52)	-2.79 (-2.03)	-8.28 (7.15)
Mom HGC	1.02 (3.25)	0.80 (2.52)	
Sex	2.27 (2.34)	2.17 (2.22)	2.32 (2.58)
Birth Order	0.01 (0.02)	0.44 (0.76)	-1.33 (2.91)
AFQT-AR	0.01 (0.19)	0.02 (0.69)	
AFQT-WK	0.14 (4.69)	0.14 (4.66)	
AFQT-PC	-0.00 (-0.01)	-0.01 (-0.50)	
AFQT-NO	-0.02 (-1.06)	-0.03 (-1.31)	
Income(10,000s)	1.77 (3.99)	1.19 (2.62)	
Mother's Hours	-1.48 (-1.14)	-1.35 (-1.03)	
Never Father	2.15 (1.57)	3.23 (2.32)	
Handicapped	-12.60 (-7.44)	-12.03 (-6.99)	-12.75 (8.24)
Home Scale		0.11 (5.41)	

(t-statistics below coefficients)

Table 6  
*Early PPVT Percentile Ranking – Verbal “Ability”*

Variable			
Constant	11.18 (3.57)	6.22 (1.76)	51.70 (40.75)
Mom Age at Birth	0.47 (4.14)	0.52 (4.56)	
Black	-17.12 (-17.31)	-14.54 (-14.46)	-26.60 (33.31)
Hispanic	-11.55 (-11.91)	-10.81 (-11.13)	-19.17 (21.68)
Mom HGC	0.87 (3.95)	0.53 (2.40)	
Sex	2.01 (2.93)	1.17 (1.70)	1.14 (1.69)
Birth Order	-4.09 (-10.55)	-3.35 (-8.58)	-4.72 (14.23)
AFQT-AR	0.07 (2.93)	0.08 (2.77)	
AFQT-WK	0.18 (8.83)	0.19 (7.62)	
AFQT-PC	-0.01 (-0.46)	-0.03 (-1.34)	
AFQT-NO	0.01 (0.96)	-0.00 (0.97)	
Income (10,000s)	1.00 (6.32)	0.98 (3.35)	
Mother's Hours	-3.23 (-3.64)	-2.80 (-3.16)	
Never Father	-2.31 (-2.33)	-1.58 (-1.59)	
Handicapped	-5.11 (-3.71)	-4.50 (-3.28)	-6.46 (4.86)
Home Scale		0.19 (13.37)	

(t-statistics below coefficients)

Table 7  
PIAT Reading Comprehension

Variable	Initial	Yearly Gain	Initial	Yearly Gain
Constant	22.03 (7.19)	-3.41 (-2.80)	20.29 (6.41)	-3.46 (-2.73)
First BPI	-0.03 (-2.38)	-0.00 (-1.88)	-0.02 (-1.88)	-0.01 (-1.84)
Digit Span	0.22 (17.66)	-0.00 (-0.82)	0.21 (17.11)	-0.00 (-0.75)
PPVT	0.21 (14.19)	0.01 (2.26)	0.20 (13.50)	0.01 (2.35)
Mom Age at Birth	0.44 (3.76)	0.02 (0.39)	0.48 (3.97)	0.02 (0.54)
Black	4.25 (4.39)	-1.26 (-3.61)	4.76 (4.81)	-1.31 (-3.64)
Hispanic	1.45 (1.51)	0.30 (0.85)	1.51 (1.55)	0.29 (0.82)
Mom HGC	0.19 (0.86)	0.01 (1.06)	0.00 (0.02)	0.32 (0.40)
Sex	4.42 (6.62)	-1.24 (-5.13)	4.38 (6.47)	-1.21 (-4.87)
Birth Order	-3.28 (-7.57)	-0.17 (-1.04)	-3.10 (-7.00)	-0.23 (-1.40)
AFQT-AR	0.02 (0.77)	0.00 (0.44)	0.02 1.02	0.00 (0.33)
AFQT-WK	0.04 (2.02)	0.01 (1.93)	0.04 (1.65)	0.02 (2.03)
AFQT-PC	0.04 (1.97)	-0.01 (-0.97)	0.04 (2.09)	-0.01 (0.99)
AFQT-NO	0.06 (3.61)	0.01 (1.05)	0.06 (3.53)	0.01 (1.08)
Income (10,000s)	0.44 (1.29)	0.13 (1.38)	0.22 (0.61)	0.15 (1.53)
Mother's Hours	-0.05 (-0.05)	-0.07 (-0.69)	-0.08 (-0.09)	-0.02 (-0.06)
Never Father	-1.35 (-1.59)		-1.073 (-1.24)	
% Father Absent		-0.07 (-0.23)		-0.15 (-0.48)
Handicapped	-9.38 (-8.04)	-0.19 (-0.45)	-9.40 (-7.98)	-0.10 (-0.22)
% Years in School		2.55 (4.53)		2.56 (4.44)
% Years in Priv Sch		-0.20 (-0.38)		-0.28 (-0.52)
Home Scale			0.06 (4.37)	-0.01 (-1.58)
R-squared	0.27	0.04	0.27	0.04
F	99.60	8.74	93.09	8.08

(t-statistics below coefficients)

### 3. Geographic Data

The NLS program releases its data in two formats - a public use version on the Web and a "geocode" version that requires a restrictive agreement with the Bureau of Labor Statistics. The geocode version of the data provides information on the respondent's county. There are over 3,000 counties in the U.S., so while the level of detail is not terribly fine for large metro areas, one certainly has a wealth of contextual data with the inclusion of the county codes. Not surprisingly, almost two hundred users order the geocode data each year. With roughly 900 users who are registered to download data from our Web site, this is a good fraction of users. The lesson here is that researchers value geographic data and it is a real mistake not to make these data available.

In addition to the county level data, we have gone back to the late 1970's to geocode each respondent's address - latitude and longitude - to a precision that in most cases places the respondent within about 50 feet of their true location. Naturally, these data cannot be made publicly available, but geographic data at this level opens up many opportunities for analysis that we did not think of twenty years ago.

We often think of geographic clustering as a necessary evil in sampling designs. I believe using geocode data in the analysis can change this perspective. First of all, by including some fairly basic variables at the county level, one can significantly attenuate the intracluster correlation that makes the true standard errors in a regression larger than the naively computed standard errors. To first approximation, the "design effect", which measures the necessary adjustment to standard errors for clustered samples, is  $1+(k-1)p$ , where  $k$  is the cluster size and  $p$  the intra-cluster correlation. Simple demographic measures from the counties that make up the cluster explain a significant fraction of the "shared variance" for a cluster. Second, the behaviors of peer respondents within the same cluster can serve as instruments in more sophisticated analyses. For example, one can use detailed latitude and longitude data to compute the number of employers or child care providers within a predetermined radius, the availability of health care providers and other creative analytic constructs.

To be sure, there are confidentiality problems that arise with geographic data. That said, as institutional review boards become more assertive in how researchers use data, the balance of risk associated with inappropriate data use is shifting from data providers to data users as users, even of secondary data, are being forced to explain to their IRBs what measures they are taking to protect respondent confidentiality as a part of their human subjects protocols. At the risk of irrational optimism, if this shift allows us to release more detailed geographic data to users the research community will be better served despite the aggravations of IRBs. We need to be more creative in appending variables derived from geographic information systems into public use files in a manner that does not compromise confidentiality. This will require thought but will surely pay dividends. For example, we are constructing a matrix of unsigned distances between all our respondents' locations in all survey waves. This will support an examination of migration without referencing exact locations.

Under the heading of what we have learned, the ability of geographic data to enhance our understanding of which factors influence behavior appears to be steadily creeping up the list. Measures that suppress geography increasingly carry a serious loss to the research community, especially compared to the rather speculative disclosure risks. The law is the law, so for Stat Can the challenge is to develop institutions that allow these research benefits within their legal constraints.

#### **4. Missed Opportunities**

There are two areas where we missed big opportunities – schooling and health, although it is not yet too late for health. With the Child Survey's attention on social, emotional and cognitive development, perhaps the most important part of the process for which data are missing is schooling. It is not for want of trying. We collected data on the schools the children attended covering the years 1993-1995. However, changing standards regarding confidentiality made it impossible to collect detailed data on the children from their teachers, and the data we could release in a public use sample was heavily circumscribed. Given these limitations, we did not press ahead with additional schooling surveys for subsequent years. Comprehensive data on schooling processes and attributes together with detailed data on maternal and household characteristics would, in principle, support a more careful separation of the effects of schools versus the home environment – a difficult yet important research topic with significant policy implications.

On the health side, again the growing stringency of confidentiality rules and the justifiable unease, to say nothing of cost, about imbedding a detailed health study within a general purpose social science survey makes it difficult to combine these two strands of research. For the NLS, the sample starts off with many siblings and, in the case of the NLSY79, includes a second generation of respondents who are the children of the female respondents. Kinship networks will add a lot to health analyses by allowing researchers to control for family and hence at least some genetic effects, the opportunity loss of not mounting a serious effort in health becomes larger. Because the NLS is based in the Bureau of Labor Statistics, a combination of health and labor market analyses is a harder sell than, for example, with NIH alone. We all need someone to show the way and demonstrate that obtaining medical measurements from general survey respondents does not poison the well of cooperativeness.

The research gains from extended sib networks testify to another missed opportunity – establishing a matrilineal strategy for tracking both individuals and the biological children of the females – generation after generation. Every time users confront longitudinal data they almost always want data from earlier in the life cycle. By backward recursion one arrives at the conclusion that we want our samples to begin before the mother's pregnancy and continue for life. Assessing the offspring produced by the females in the Children of the NLSY would provide powerful data. Organizations such as Stat Can are ideally suited to pull off this strategy.

Perhaps the greatest unexploited opportunity for survey projects lies in administrative data. Again, the increased stringency regarding confidentiality and consent make this

more difficult than it once was. For agencies such as Stat Can this is not likely as big an obstacle as for those of use less firmly located inside government, but as a way to validate our data and explore the seriousness of non-response bias, administrative data can be a powerful tool. We are in the process of securing permission to match unemployment insurance employment data on earnings and employers from Ohio to our survey data. Being able to match from a wider variety of states would improve matters, but the permission process required makes this difficult. Knowing the measurement error process in our data will only improve analysis.

Finally, I will close with what I consider the largest error we have made. With the advent of computerized interviewing we have the tools to administer surveys that would have been technically infeasible in the days of pencil and paper if only because the sorts of check items we can execute on the computer are far more sophisticated than what interviewers could execute on a paper form. Despite our efforts to generate printed questionnaires that suppress much of the complexity in computerized questionnaires, when we have produced complex sections they have often been a nightmare for users. Our questionnaires have, in places, complex branching and rostering that defies comprehension by users not intimately familiar with the project. We collect data to serve researchers, and if we make the instruments opaque we suppress usage and ill serve the research community. Because of their inherent complexity, longitudinal data have never suffered fools gladly, but the punishment is, at times, extreme.