

Longitudinal Data and Public Policy

Charles Brown

Department of Economics and Survey Research Center, University of Michigan
and
National Bureau of Economic Research

Prepared for a National Academy of Sciences Workshop
Access to Research Data: Assessing Risks and Opportunities
October 16-17, 2003

I am grateful to Rebecca Blank, Janet Currie, Emma Hutchinson, Bruce Meyer, and Olivia Mitchell for references to studies in the policy areas reviewed in this paper. Imperfections that remain despite their help are clearly my responsibility.

Readers should be aware that I was a member of the Panel Study of Income Dynamics team, and currently have similar status with the Health and Retirement Study.

The importance of longitudinal data in advancing our understanding of how labor markets function and the consequences of this functioning for firms, workers, and workers' families is well documented (Stafford, 1986 and Manser, 1998). In this paper, I attempt to assess the impact of research using longitudinal data on public policy. Assessing impacts on policy is more difficult than assessing impacts on academic research -- legislators rarely cite academic papers and when they refer to academic work at all it's a fair question whether the research changed the vote or a vote based on other considerations prompted reference to supporting research. One can, however, attempt to identify policy-related findings based on longitudinal data, and ask whether policy appears to have responded to such findings. Inevitably, this approach involves three judgments: what are "the facts"? to what extent was longitudinal data decisive in establishing "the facts"? to what extent did policy move in a direction that seems consistent with "the facts"?

In assessing the importance of longitudinal data, it is important to understand that such data have come from three sources: surveys, administrative records, and policy experiments (which typically combine "survey" data around the time of the intervention with follow-up surveys and/or administrative records to assess the its impact). Given my past involvement with the Panel Study on Income Dynamics (PSID) and current connections to the Health and Retirement Study (HRS), my first thoughts were of survey data from households. As I tried to move from my (natural) focus on the academic literature in general to papers with policy impact, I realized that administrative records and especially policy experiments have become increasingly important.

Longitudinal data can make two contributions to research. First, they allow more accurate reporting of transitions between states, durations in a particular state, and changes in variables of interest than is typically possible from a single cross-section data collection. Validation studies usually find that longer recall periods lead to less accurate reporting of retrospective information (Bound, Brown, and Mathiowetz, 2001) Thus, collecting information each year about the past year's activities will produce more accurate data than asking for a ten-year history. Second, longitudinal data allows the researcher to control for otherwise-omitted variation in outcomes among individuals, as long as this variation is constant for a given individual. Both of these contributions are evident in the examples discussed in this paper.

I begin by discussing five policy areas in which longitudinal data have made important contributions: welfare reform, job training, unemployment insurance, preschool programs, and retirement. I then focus briefly on ways in which longitudinal data could make further contributions to policy discussions. I conclude with a few thoughts on the ways in which the interplay between different types of longitudinal data have contributed – and could contribute – to informed policy discussions.

At the outset, I should stress that I have chosen topics and particular papers within each topic in order to illustrate the range of contributions from longitudinal data. I do not attempt to cover all substantive areas where I believe such contributions have been

largest. Indeed, I am confident that someone with a different background could make at least as strong a case as I have from five completely different policy areas; e. g., employer-provided health insurance, policies affecting the disabled, K-12 educational reform, occupational safety, and tax policy would form the core of an equally compelling analysis (written by a different analyst!). I've tried to rely on recent important survey papers, where available, as a check on my own idiosyncratic interpretations, and have not attempted to cite the underlying literature with the care that a paper on one of these topic areas would attempt. But I hope that revisiting the policy areas that I do consider, and reflecting briefly on the larger universe that I have left untouched, will convince readers both that the contribution of longitudinal data has been considerable, and that the payoff to helping researchers do better might be equally large.

1. Welfare Reform

Public assistance ("welfare") has been reformed and re-reformed over the past 40 years. The most discussed (and most "reformed") programs are those that provide cash assistance to families with dependent children. The fundamental dilemma is that changes in benefit levels or eligibility rules that provide more assistance to poor children tend to discourage work effort by their parents. Prior to 1996, Aid to Families with Dependent Children (AFDC) had undergone several important changes: a gradual reduction in the real benefit level in most states and federally-mandated changes in the rate at which benefits are reduced as parents' earnings increase. Some states received "waivers" from federal regulations to experiment with program design; often these took the form of more stringent work requirements. In 1996, AFDC was replaced with Temporary Assistance for Needy Families (TANF), which mandated "work requirements" as a condition for benefit eligibility and a lifetime limit of five years of benefits for each recipient, while giving states greater freedom to experiment in designing their own TANF programs.

If we go back to the early 1960s, benefit levels varied dramatically by state, and benefits were reduced dollar-for-dollar with earnings, providing recipients no incentive to take jobs that paid less than their AFDC benefit and little incentive to take jobs that paid somewhat more. Two-parent families were rarely eligible for any benefits, which generated concern in some quarters that the program was encouraging family break-up. In order to assess the impact of changes in the benefit level and/or the benefit reduction rate, and of extending benefits to two-parent families, the Office of Economic Opportunity (part of what is now the Department of Health and Human Services) funded several "negative income tax" (NIT) experiments. Separate experiments were run in New Jersey, Gary, Seattle, Denver, and rural areas in Iowa and North Carolina in the late 1960s and early 1970s. The key data were reports of earnings and hours worked by participants and control group members; these are "survey" data in the sense that they are obtained from workers rather than from payroll records, but they are "administrative" in the sense that, for members of the experimental group, payments were based on these reports. There seems to be general agreement that the behavioral responses were smaller than might have been expected based on non-experimental studies, but considerable disagreement about the practical implications of the estimates within this range. It is difficult to infer the effects of a permanent negative income tax from responses to a

temporary experiment, and the actual response of recipients appears to have been overstated by under-reporting of hours by those in the treatment group (understating hours and earnings increased their benefit payments). Burtless (1986, p. 456) concluded that the earnings reductions caused by such a program would be "only moderate", but that the resulting earnings reductions "would offset a substantial part of the income gain from more generous transfers." Ashenfelter (1986, pp. 53-55) was more agnostic, based on weaker evidence of any labor supply effects from simpler estimation strategies and greater concern about the under-reporting of earnings.

Whatever the policy message of the NIT experiments, the direct policy impact seems to have been small. The NIT experiments were initiated during the War on Poverty, when NIT-like reforms were under serious discussion. By the time the experiments could be fielded and the results analyzed, policy discussions had shifted away from using simple non-categorical tax-like alternatives to AFDC (Hall, 1986).

While broad-based changes of the sort "tested" in the experiments did not occur, the benefit-reduction rate under AFDC was reduced in 1967, increased in 1981, but reduced again in some states in the 1990s. Meanwhile, researchers turned their attention to poverty and welfare "dynamics". Cross-section data can tell us what fraction of the population is poor or receiving welfare at one point in time, but a given poverty rate (or AFDC caseload) is consistent with very different dynamics. If the same families are poor (or on AFDC) year after year after year, the case for interventions to lift them out of poverty (or nudge them off of AFDC) is stronger than if spells of poverty and welfare dependence are short and transitions to "adequacy" and "self-reliance" are quick.

Some of the earliest and probably most influential of these studies by Bane and Ellwood (summarized in Bane and Ellwood, 1994) used longitudinal household data from the Panel Study of Income Dynamics. These provided nationally representative samples and a potentially useful set of recipient characteristics that might be used to "predict" which new recipients were destined to become long-term recipients; they could also follow recipients as they moved from one state to another. An alternative strategy was to use administrative records (e.g., Gleason, Rangarajan, and Schochet, 1998), which offered potentially more accurate dating of spells¹ but fewer interesting variables for predicting spell length.

One gets a different picture of welfare durations depending on whether one asks about the ultimate duration of newly-begun spells or the ultimate duration of spells currently in progress. Focusing on new spells puts greater weight on short spells, while focusing on spells-in-progress puts greater weight on the longer ones. One gets a still different picture if one focuses on incomplete rather than completed spells. Overall, these analyses show clear evidence of both short-term "safety net" and long-term

¹ Bane and Ellwood (1994, p. 33) note that their PSID data record the number of years in which individuals receive benefits, so that someone who had separate "spells" in consecutive years would be recorded as having a single two-year spell. But they note that administrative records will show someone who lost benefits for a month or two due to some administrative complication as having had two spells, while for policy purposes one would want to think of this sequence as one welfare experience.

"dependence": "The vast majority of people starting welfare at a point in time and the vast majority of those who ever have spells on welfare stay only a short time. Yet the majority of recipients at a point in time are in the midst of a much longer spell, and most welfare funds are spent on them." (Bane and Ellwood, 1994, p. 36) Half of new spells will end within two years, and seventy percent within four years; yet nearly half of those receiving benefits at one point in time will be on the rolls for more than 10 years.

A second key fact, which emerged more slowly in the literature, is that some families suffer multiple spells: exits from AFDC are not necessarily permanent. Of those leaving welfare for a full year, 32 percent have begun a new spell in the next six years. Blank and Ruggles (1994, p. 50) used data from the Survey of Income and Program Participation, and focused on a shorter "window"; they found that "20 percent of all post-AFDC and food stamp spells end with a relatively quick (within 6-9 months) return to public assistance."

The complexity of the resulting picture leaves the "fact" to be emphasized to the indiscretion of the advocate, and on an intensely partisan topic like welfare reform the advocates were more than up to the task. On balance, those concerned with dependence gained the upper hand in the political debate, leading to state experimentation with stronger work requirements and time limits, and ultimately to requiring such measures with TANF in 1996.

While the negative income tax experiments were designed to measure "supply responses" to changes in the rules determining benefits, they were not very helpful in assessing work requirements or time limits for benefits. However, when states received "waivers" from federal AFDC rules to implement such policies, HHS required serious evaluations, typically involving a classical design -- randomized assignment to treatment and control groups -- and outcomes were then followed longitudinally. The evaluations blended baseline information (essentially, a survey) with administrative records from the unemployment insurance (UI) and AFDC programs. "Over time, this generated a body of literature about welfare-to-work programs that was crucial in convincing people that such programs could have positive effects on earnings and labor supply, and negative effects on welfare spending." (Blank, 2002, p. 1107)

The most common waiver experiment was mandatory employment-related services. Some were "job-search-first" programs in which job-search activities were required of new recipients, and those who did not find jobs were then referred to education or training programs. Other "education-first" programs reversed this sequence. A third group assigned different recipients to different treatments (usually those with more education being assigned to the "job-search-first" component). Impacts varied widely across programs, but overall it is clear that such programs did increase earnings (the unweighted median gain across programs was about \$400/year) and reduced welfare benefits. However, unless the experiment also included a lower benefit-reduction rate or similar supplement, the employment services did not increase participant income (Bloom and Michalopoulos, 2001). Little could be inferred about time limits, because fewer states adopted experimental time limits and (at least at the time of Bloom and

Michalopolous's report) too little time had passed since their introduction. At a smaller number of experimental sites, effects on children were also studied (Morris et al, 2001). My reading of these experiments is that the evidence is more tentative (much weight is put on parents' or teachers' assessments of achievement, behavior, and health because more objective measures were not always collected, and longer-term effects could not be measured), but that interventions that raised parental income tended to be good for their children, whereas work-only interventions had no clear effect (positive or negative).

An alternative focus is on the characteristics of those who remain on welfare, despite policies designed to nudge or force them off of the rolls. Danziger (2001) followed a group of particularly disadvantaged recipients. Not surprisingly, she found relatively low exit rates, and that the welfare reliant often had multiple barriers to finding jobs. Significantly, some of these (e.g., lack of a high school diploma) can be measured in standard surveys, but others (e.g., learning disabilities, illiteracy, mental health problems would not).

2. Job Training

Given a strong preference for increasing earnings rather than transfers in dealing with poverty (and job displacement among those with above-poverty earnings), job training and related services have been an important focus of both policy and of policy evaluation. Unfortunately, the fraction of the population receiving publicly funded job training in any one year is smaller than the number of families on AFDC/TANF (for example, in recent years the number of families receiving TANF is at least ten times the number of disadvantaged youth and adults receiving job training). As a result, most of the evaluation of these programs has begun with program records identifying recipients, with their pre-program history and post-program experiences measured either with survey data and/or administrative earnings records. With the exception of a few programs where applicants were randomized into treatment or control groups, the control group for most job-training evaluations consists of eligible (or at least reasonably disadvantaged and so near-eligible) non-participants. Here, external data are needed both to select members of the control group and to track their earnings. Typically, earnings data are from administrative records. In some cases, a blend of administrative and survey data (e.g., the CPS/SSA matched file which located Social Security earnings records of 1976 CPS respondents) are used.

Like AFDC, the job training programs to be evaluated have undergone significant changes. The Manpower Development and Training Act (MDTA) was passed in 1962. MDTA provided classroom or on-the-job training to a client group with relatively stable employment interrupted by job loss. In 1973, MDTA was replaced by the Comprehensive Employment and Training Act (CETA), which focused on economically disadvantaged workers and labor market entrants. Classroom and formal on-the-job training was somewhat downplayed, and "work experience" to provide a transition to stable employment was an important component of CETA. The perception that CETA provided dead-end makework rather than work experience leading to regular employment resulted in its being replaced in 1982 by the Job Training Partnership Act (the eyebrow-

raising collaboration of Senators Kennedy and Quayle). JTPA formalized a focus on preparing its clients for private-sector employment by including Private Industry Councils (PICs) in the governance structure of local programs, and provided job training, job-search assistance, and related services. JTPA was, in turn, replaced by the Workplace Investment Act of 1998, but this most recent training initiative has not (yet) generated the volume of evaluation that are available for its predecessors.

An early and influential analysis of MDTA was Orley Ashenfelter's (1978) evaluation of classroom training, for those receiving such training in 1964. Earnings data for trainees and controls were taken from Social Security earnings files, which meant that the control group could not be matched on education, hours worked prior to training, or local labor market conditions. Ashenfelter found that trainees suffered a significant drop in earnings in period prior to training, and the estimated effect of training was quite sensitive to whether one assumed that that drop would have been persistent or transitory in the absence of training.²

Anxiety about the adequacy of non-experimentally chosen control groups and the range of estimates of MDTA and CETA program effects was intensified by Lalonde's (1986) evaluation of the National Supported Work (NSW) demonstration. Perhaps because the demonstration was an experiment rather than an ongoing program, a decision was made to randomize eligible applicants into treatment and control groups. Thus, simple treatment-control differences in post-training years provide unbiased estimates of the program's effects, which were significant but not enormous (\$800-\$900 per year, for both female and male participants). Lalonde also generated control groups from PSID and from one wave of CPS to which respondents Social Security Earnings had been appended. Thus, he could ask whether non-experimental control groups would give estimates similar to the preferred experimental estimates. They did not, and the range of resulting estimates was very wide.

One reading of Lalonde's results might be: experimental=good, non-experimental=bad. Heckman, Lalonde, and Smith (1999) reached a different conclusion: that non-experimental control groups need to be selected with great care. Using data from the National JTPA Study (which, like Lalonde's earlier study, had both a random assignment of some applicants to a control group and other control groups) they argued that matching the controls group to the treatment group based on recent labor force history and local labor market (as well as standard "demographic" variables) is important. So is getting data for both treatment and control groups from the same survey or same administrative data source. "Comparing comparable people [with comparable data] goes a long way toward reducing the bias in non-experimental methods reported by LaLonde." They noted that the administrative records which play a very large role in MDTA and CETA evaluations do not provide much detail on labor force status, and confidentiality restrictions that limit the geographic identifiers preclude careful matching of local labor markets for trainees and controls.

² Heckman, Lalonde, and Smith (1999, Fig. 1-6) show that this same pattern can be found in other programs (CETA and JTPA) in the US, and for similar programs in Sweden and Norway as well.

Heckman, Lalonde, and Smith summarized the results of evaluations of MDTA, CETA, JTPA, and several demonstration projects such as NSW. Of these, the JTPA and demonstration project evaluations are based on randomly assigned control groups, while the others evaluations are based on alternative (and generally less reliable) control groups. For the most part, the literature has focused on economically disadvantaged workers, rather than displaced workers receiving retraining. On the whole, they find what I would describe as a discouraging variety in the estimates in just about every dimension -- differences across programs, across cohorts for a given program, across sites for the same program at one point in time, and across broad demographic groups. Moreover, when two studies differ, they typically differ in so many dimensions that the reasons for the differences can't be isolated with confidence. Nevertheless, some patterns emerge. The programs studied do seem to raise earnings of adult female participants; particularly if (as very limited evidence suggests) these gains persist well beyond the 2-3 year post-training period that is typically studied, the training would pass standard cost-benefit tests. For adult males, the evidence is less clear -- training seems to raise earnings, but by less for men than for women. For youth, there is little evidence that such programs raise earnings.

From the standpoint of data collection, these studies illustrate a rich potential interplay between various data sources, and the potential for disaster when such interplay is not aggressively pursued. Unless the control group is randomly assigned, selecting a reasonable control group requires more detailed information than is typically available from administrative data. On the other hand, tracking non-participants for several years is needed to see if the gains from training persist, and this is difficult to do without access to administrative earnings records.

3. Unemployment Insurance

For those with sufficient recent labor force attachment, workers who become unemployed through temporary or permanent layoff are eligible for unemployment insurance (UI) benefits. Typically, benefits are equal to half of previous earnings, but there is a relatively low benefit ceiling, so that the replacement rate for those with above-average earnings can be significantly less than .5. Ordinarily, benefits are available for only 26 weeks. It has long been recognized that such benefits are likely to extend the duration of unemployment -- being paid half one's previous earnings while unemployed will lead some individuals to search longer for their next job, and lead others to search less intensively. The policy issue is the size of these responses.

A number of papers have used longitudinal data, from household surveys or especially from the UI system's administrative records, to gauge the effect of these benefits on the duration of unemployment. A typical study design focuses on a time period in which either the maximum benefit was increased (which is relevant only for those whose benefits were limited by the old ceiling) or the duration of benefits was extended for some unemployed but not others. Here the role of longitudinal data is to provide an accurate measure of completed duration; once the completed durations have been established, they become the dependent variable in what is essentially a sequence of

cross-sections (one cross section per time period). Krueger and Meyer (2002, Table 2.5) provide a recent survey of these studies, and conclude that a ten percent increase in the benefit level (or the replacement rate) increases completed duration by at least 5 percent. Effects of benefit extensions are also positive, though more varied.

Longitudinal data also allow one to plot the hazard function for leaving unemployment. A common finding is that this hazard increases sharply as an unemployed worker approaches the point at which benefits are exhausted (typically 26 weeks). This provides simple but convincing confirmatory evidence that such benefits really affect re-employment decisions in a perceptible way, though of course this evidence is based on the minority of recipients who (nearly) exhaust their benefits.

The evidence that providing relatively generous benefits slows recipients' return to work was sufficiently strong that several states experimented with alternative mechanisms to counteract this effect. Fortunately, evaluation was built into these experiments, so that those filing new claims for benefits were randomly assigned to treatment and control groups. Longitudinal administrative data allow one to calculate completed spells of UI receipt; total benefits collected, and earnings after returning to work (though differences here were mixed and often imprecisely estimated).

One idea is to pay a bonus to those who find a new job (and so stop receiving UI benefits) quickly. Four states experimented with this option. The first such experiment, in Illinois, paid a bonus of \$500 (roughly four times the average benefit) to those who took a new job within 11 weeks of filing for benefits. This reduced average benefit duration by about a week (or about 5 percent of the average duration in the control group). This in turn reduced net costs for the UI system, especially given that only a quarter of those offered a bonus found a new job in time to qualify for it, and of these only 55 percent bothered to claim the bonus (Meyer, 1995, p. 109). Subsequent experiments in New Jersey, Pennsylvania, and Washington produced smaller reductions in weeks receiving benefits and generally less encouraging prospects for reducing costs for the UI system. Meyer (1995) argued that these prospects become even darker when one takes account of the likelihood that a bonus would encourage those who know they can find a job quickly (and so don't bother for filing for UI) to apply for UI benefits in order to get the re-employment bonus. He concludes that by the time of his survey, roughly four years after initial reports from the last experiment were available, "the initial optimism [of both policy makers and academics to the idea of a re-employment bonus] is fading."

An alternative way of reducing the cost of UI benefits is to provide greater job-finding services, require participation in a job-search workshop, or more rigorously enforce requirements that UI recipients be actively searching for work. Various treatments of this sort reduced average duration by 0.4-.1.1 weeks, and were inexpensive and so resulted in net savings to the UI system.³ Meyer (1995) noted that while the bonus

³ See Meyer (1995) and, for a more recent pair of experiments, Decker, Olsen, Freeman, and Keplinger (2000). Ashenfelter, Ashmore, and Deschênes (forthcoming) report that while more intensive verification of claims eligibility can result in a modest reduction in benefit payments (though, as they emphasize, the

experiment results are largely consistent with the thrust of the non-experimental literature (that unemployed workers respond to financial incentives), the positive results of the job-search experiments contradict non-experimental findings, and he concludes "in this case, non-experimental methods are not as reliable" (p. 127).

A final example of the use of longitudinal data in evaluating UI experiments is worth mentioning both for its substantive results and its methodology, which I hope can be much more widely applied. In Kentucky, new UI claims are classified according to their predicted benefit duration, and those forecast to have long durations receive mandatory re-employment services. Those with the highest priority are assigned first, and so on until available "slots" are exhausted. The classification by predicted duration is in 20 categories, rather than being a continuous variable, and often a local UI office can target only some of those in the last-served category for services. In this case, the assignment is randomized within the last-served (marginal) priority category, and those who are not assigned to mandatory services form a natural control group for those who are. Black, Smith, Berger, and Noel (forthcoming) find that those receiving mandatory services have shorter benefit durations, most of which is concentrated shortly after they receive notice that they will be required to participate, rather than after the services are actually received. They conclude that the threat of re-employment services is an important part of their observed impact. However, they find little evidence that the effect of treatment varies by priority class.⁴

4. Pre-School Programs

While the policies discussed so far are directed primarily toward adults, there is growing interest in educational reforms that directly affect children. As with many policy experiments, evaluation naturally takes the form of comparing changes in some outcome (e.g., test scores) for those who receive some treatment to those who do not.

The potential contribution of longitudinal data to evaluating these policies is well illustrated by the evaluation of Head Start, a program for disadvantaged pre-school aged children, and other similar demonstration programs. The need for a longitudinal design is evident: the treatments are administered to pre-school children, with the hope that they will improve outcomes in school and in later life. Treatments include different mixes of preschool education programs, day care, home visits, nutritional supplements. Outcomes include school readiness, emotional development, child health, cognitive development (e.g., test scores and school grades), educational attainment, employment status in young adulthood, and even avoiding arrest.

Currie (2001) provides an evaluation of "model" programs, emphasizing those that used control groups based on randomized assignment. While the individual

evidence here is not very strong) they found "no evidence that verification of claimant search behavior led to shorter claims or lower total benefit payments."

⁴ At some local offices in some time periods, the number of new filings is small enough relative to available resources that the marginal priority category would be those with short expected durations; in other locations/time periods, only those predicted to have fairly long durations would be included.

programs are often small (typically about 100 children--see NAS, 2000, Appendix A), and effect of a particular program on a particular outcome may not be statistically different from zero, the overall pattern of the results suggests that model interventions do have positive effects that last into elementary school and (we have much less evidence here) beyond. Nonetheless, there is also evidence of "fadeout" -- that impacts decline as one extends the period of evaluation -- particularly for measured IQ.

The generally favorable verdict on the model programs raises the question of whether "generic" childhood education programs are also effective. Here the available evidence is non-experimental, raising obvious concerns that participants in a program like Head Start -- and their families -- may be quite different from non-participants. Using data from the National Longitudinal Survey of Youth, Currie and Thomas (1995) followed children who participated in Head Start and compared them to siblings who did not. For black children, they find that initial gains in test scores fade out during elementary school; for white children, they do not. In other work, they find that black (but not white) Head Start children are likely to attend lower-quality elementary schools, and this helps explain the fade-out experienced by black Head Start children. Garces, Currie, and Thomas (2000) used a retrospective question asked of PSID respondents under 30 about whether they enrolled in Head Start as a child, and linked it to the regular PSID longitudinal data on schooling, labor force participation, and another supplement on criminal activity. Perhaps surprisingly (given likely recall errors) they found evidence of positive effects on school attendance, employment, earnings, and lack of arrests, though these beneficial effects typically emerge only some groups and not others. To my eye, there is little pattern to these results.

The generally favorable evaluation of Head Start has had a positive impact on its continued existence; the most recent debates have been whether to extend low-cost versions to more students or offer higher-cost programs that serve fewer children, and whether to refocus the program to focus more sharply on cognitive outcomes. From another standpoint, however, the research record has been much less satisfactory. Those implementing programs would like to know not only whether Head Start produces measurable benefits, but also which components are most effective, and how this effectiveness differs across different groups of children. For a variety of reasons -- particularly the limited sample size of the model programs -- very little is known on this score (NAS, 2000, pp. 18 and 30).

An alternative to the direct provision of services under Head Start is the "case management" approach piloted in the Comprehensive Childhood Development Program (CCDP). Here the model was that services needed by low-income, at risk mothers and their children are often available in the community, and CCDP attempted to link these families to available services. While there was some evidence that CCDP succeeded in increasing access to parenting education and traditional educational opportunities for mothers, and child care and preschool for the children, there was no evidence of positive impacts on child outcomes (St.Pierre, Layzer, Goodson, and Bernstein (1997)).

5. Retirement

Given concerns about the future of the Social Security system and, more generally, the ability of various sources of retirement income to support the baby boom generation as it reaches retirement age, there has been an enormous volume of research related to these issues. Not surprisingly, longitudinal data has played a major role in many facets of these debates.

Two major longitudinal data sets combined household surveys with Social Security earnings records. The Retirement History Survey began with interviews of household “heads” in 1969, and re-interviewed them (or their surviving spouses) every two years until 1979. The Health and Retirement Study began with interviews in 1992 of those born in 1931-41 in 1992 (i.e., those age 51-61), and their spouses. This original sample has been re-interviewed every two years since, and other age cohorts were added to the HRS sample, so that those born in 1947 or earlier are now included. HRS also collects detailed data from employers on the pension plans of its respondents. Both RHS and HRS include measures of non-retirement assets, health, etc.

One relatively simple factual issue that has come under re-examination is the extent to which Social Security redistributes income. At first glance, the answer is obvious: taxes are a flat proportion of earnings (up to a ceiling), monthly benefits relative to taxes paid are higher for those with lower earnings, so the Social Security redistributes income from rich to poor. However, especially when one uses the family as the unit of analysis, matters are more complicated. Redistribution from higher-wage husbands to lower-wage wives nets out at the family level; payment of spouse and survivor benefits redistributes from single workers to married couples, especially those in which one member has low earnings. Both the basic annuity form of benefits and the benefit for surviving spouses redistribute from those with lower life expectancy to those who live longer.

In order to evaluate these and other factors, one needs longitudinal data on earnings (in order to calculate lifetime earnings and benefits) and data on who is married to whom (in order to track the myriad of impacts of spouse and survivor benefits).⁵ Gustman and Steinmeier (2001) matched Social Security earnings records Health and Retirement Study individuals and couples; Laibson (2001) matched Social Security records to individuals and couples in the Survey of Income and Program Participation. Both studies concluded that while Social Security involves sizeable redistributions, the redistribution from rich to poor households is substantially less pronounced than a simple focus on the function relating benefits to individual earnings would suggest.

A broader set of policy concerns relate to the ability of the Social Security system to provide benefits promised under current law as the Baby Boom reaches retirement age, and the more general issue of the adequacy of retirement income for those that will soon be reaching retirement age. Reliable information about workers’ response to the incentives provided by Social Security and private pension plans, changes in incentives provided by private pensions, and the determinants of the non-pension assets of workers

⁵ Ideally, one could also incorporate individual characteristics that predict mortality, though the studies I refer to below have not moved far in this direction

approaching retirement age are clearly relevant to predicting the extent of the impending crisis for Social Security and in choosing among alternative policy responses.

With regard to pensions, there is clear evidence that most defined-benefit pensions provide sharp incentives to remain until early retirement and to leave no later than normal retirement age. Data from RHS (even with relatively imprecisely matched pension data—see Gustman and Steinmeier, 1986) and from company personnel records (surveyed in Hurd, 1990) showed that workers respond to these incentives in deciding when to retire. Anderson, Gustman, and Steinmeier (1999) assessed the importance of changing incentives under defined-benefit plans (toward early retirement) and the increasing relative importance of defined-contribution plans (which do not encourage early retirement) between 1969 and 1989. They combined data on changing pension provisions from a number of sources with a structural model estimated from the RHS. They estimate that changing pensions had significant effects on early retirement, though these account for only about 15 percent of the observed decline in full time employment over the period.

Estimating the effects of Social Security on retirement is more complicated, because the same “rules” apply to all workers, and benefits vary primarily because of differences in previous earnings – which one would expect to have independent effects on labor supply. Most authors calculate the incentives provided by Social Security and then, in effect, assume that workers respond to incentives from earnings, Social Security or pensions in a similar way. These studies usually found that the Social Security benefit rules penalize continued work at 65 (and so contribute significantly to retirement at that age) but not at 62. By process of elimination, the tendency to retire at 62 is ascribed to the fact that workers contemplating retiring earlier cannot borrow against their Social Security. Gustman and Steinmeier (2002) noted that age-62 retirees in the HRS have limited assets, even relative to previous earnings, which suggests an inability to borrow constrains their ability to retire earlier. Another potentially important incentive is provided by Medicare. Rust and Phelan (1997) found that RHS workers who retired at 65 were more likely to have employer-provided health insurance that would not be continued in retirement; those who retired earlier had either no insurance or insurance not contingent on continued employment. Mitchell and Phillips (2000) report that HRS respondents those who retired between 62 and 64 were less likely to have employer-provided insurance, but more likely to have retiree insurance, than those retiring at (or after) 65, when Medicare would be available

Given that a majority of the HRS cohort had reached age 65 as of the most recent (2002) interview, HRS has begun to be used to simulate the effect of specific policy proposals. For example, Gustman and Steinmeier (2002) estimated that raising the early retirement age from 62 to 64 would lead about 5 percent of the population to delay retirement. Coile and Gruber (2001) explored the effects of raising the normal retirement age from 65 to 67 and increasing the “credit” by those who delay retirement beyond the “normal” age. They conclude that while workers do respond in fairly predictable ways to the incentives provided by Social Security, neither of these changes would have very large effects on retirement behavior.

Linking HRS data on wealth to Social Security earnings records, Venti and Wise (1999) found that the dispersion in wealth accumulation among families with similar lifetime incomes is enormous – even among families with above-average lifetime earnings, a significant fraction approach retirement with little saved. One particular component of savings that has received much recent attention is defined-contribution pension plans. The trend from defined-benefit to defined-contribution pensions may either increase or reduce the wealth available at retirement, depending on participation, contribution, and asset allocation decisions that were “made for” the worker under defined benefit plans.

A series of papers (e.g., Madrian and Shea (2001) and Choi, Laibson, Madrian, and Metrick (2001) studied workers at cooperating firms that introduced automatic enrollment features into their 401K plans. When "enrolled" is the default rather than a chosen status, the firm must also define default "choices" for the contribution rate and how the money is invested. The studies found a very powerful effect of the defaults: automatic enrollment increases enrollment dramatically; the default contribution rate and default investment fund have powerful effects on worker "choices". For example, the typical low default contribution rate actually reduced the likelihood that a worker will choose a higher rate; a typical default investment fund (all contributions invested in a money-market fund or similar “safe” fund) greatly increased the proportion of participants who invest 100% of their contributions in such funds. Even those who are hired before automatic enrollment but first participate after it has begun (for whom the default choices have no direct impact) seemed to be influenced by these defaults. Over time, workers moved away from the defaults, but only gradually. Thus, defaults appear to be very important in many workers' choice of defined-contribution accounts, and one would expect that defaults would also have an enormous impact on the "choices" of workers if individual accounts are created under Social Security.

6. Extending the Impact of Longitudinal Data

Nearly anything that is done to facilitate the use of longitudinal data for “basic science” research will also contribute to informed policy discussion. That agenda is long, important, and well-covered by other papers at this conference. Here I want to focus on a shorter list of suggestions that I think would be particularly helpful for strengthening the impact of longitudinal micro data on policy analysis: greater persistence in studying long-run impacts, more thorough mining of regulatory data, and matching employer data to workers.

- **Persistence in studying long-run impacts**

In several of the areas discussed in the preceding sections, there is a real tension between the promise of longitudinal data and the short-term focus of much policy analysis. Longitudinal data has, for the most part, produced estimates of only shortitudinal impacts. I understand that policy cannot wait for science, and that decisions based on credible estimates of short-term impacts and educated guesses about how those

impacts might evolve in the long run are likely to be better than decisions that do not benefit from even decent estimates of short-run impact. Nevertheless, my reading of the evidence is that we often have at least hints that impacts decay over time, and it would be worthwhile – for fighting the next war, if not this one – to keep our eyes more firmly on the longer-run impacts that longitudinal data can identify.

In cases where data are obtained by special-purpose longitudinal surveys of participants and controls, such a long-term focus would be costly. One reason that we know too little about the long-run impacts of early childhood interventions is that the costs of following participants (and controls) into adulthood are so large. But in other areas, where impact estimates are based on “looking up” participants and controls in administrative data files, the benefits of doing so seem well worth the cost. How did Ashenfelter’s MDTA class of 1964 fare a decade after training? Do the job search components of the AFDC and UI experiments have any long-term benefits for participants? Failure to ask these questions probably tilts the policy agenda toward short-run fixes that get people off of welfare, or through “training”, or off of unemployment insurance with measurable short-term “benefit” while doing very little to improve the longer-run prospects of those being served.

- **Mining regulatory data**

Labor market regulations often require that firms complete reports that facilitate the regulatory process. In a few cases, these data are publicly available. For example, firms that offer pension plans are required to file form 5500, which provides information on numbers of workers covered, balance sheet information for the plan, and other (limited) information on plan provisions. These data have begun to be used (e.g., Papke 1995 and 1999), but with the solvency of many plans – and the appropriate policies to resolve insolvency – in the public eye, these data are likely to make a large contribution to informed policy over the next few years.

More often, the data are confidential (either by custom or by statute) and are shared with researchers (if at all) on a restricted basis. For example, medium and large firms (all firms with over 100 employees, Federal contractors with over 50) are required to file EEO-1 reports on the race-sex-occupation composition of their work force. In the mid-1970s and early 1980s, several studies merged these data with results of contract-compliance audits to study the impact of the affirmative action obligation of federal contractors. Along the way, they discovered that the audits were not targeted at firms which seemed to have significant minority under-representation, and that, for all the controversy about goals and quotas, contractors often failed to meet their goals. Despite these failings, the balance of a relatively sparse research literature was that the short-term impact of compliance reviews on minorities and women was positive. (Leonard, 1990)

While the relative progress of minorities and women has remained a hot research topic, I have not been able to locate any follow-ups to these early studies. We don’t know whether the contract compliance program continues to exert any effect (beyond the general pressures that apply to all firms covered by Title VII) on employment in the short

run; nor (to return to an earlier theme) do we know what were the long-term impacts of earlier rounds of compliance activities. I know of no recent evidence on how reviews are targeted. Moreover, while Blumrosen and Blumrosen (2000) argue that a non-negligible minority of firms ignore their obligation to file the forms at all, we do not know whether this represents a relatively high level of year-missing or a smaller core of deliberate, serial non-compliance.

The experience of the early days of the contract compliance program illustrates the more general point that academic researchers can make important contributions to policy debates about regulatory activities if the data can be made available. My understanding of current EEOC policy is that the EEO-1 micro data can be obtained without identifying firm information; but it is exactly such information that would be needed to link the data in a useful way to information on contract compliance or other enforcement activities.

Much of the establishment data collected by federal agencies is obtained by BLS or the Census Bureau. The data are confidential, and that confidentiality is seen by those in the agencies as a sine qua non for obtaining voluntary compliance. My sense is that the agencies have been extremely cautious in sharing such data, even on a restricted basis, not just to meet their legal obligation but because they fear that even restricted release of the data to the research community would endanger voluntary compliance. Regulatory data are not provided voluntarily by firms – it is a legal requirement that they do so. In cases where the relevant statute does not require confidentiality, as with ERISA data, there is a real opportunity to make the data freely available. Even where, as with EEO-1, firms are promised confidentiality, fear that even restricted release of the data to the research community would endanger voluntary compliance should be less compelling.⁶

- **Matching**

One theme that emerges often in the five policy areas discussed above is that matching administrative data to survey data often has high payoffs. Extending such matching to allow researchers to match survey data to firm data would likely have comparable payoffs. These raise two kinds of confidentiality issues: protecting the confidentiality of the survey respondent (if I can identify your employer, I have a good chance of identifying you), and protecting the confidentiality of the firm (if I can deduce the identity of the firm from some of the information provided, I can learn other things about the firm that it regards as none of my business). Nevertheless, if such concerns can be addressed, the potential payoff is high. For example, matching pension plan characteristics (firm data) to HRS respondents gives us a much clearer view of the retirement incentives faced by HRS workers than direct questions to respondent do. And, in the HRS example, such data have been made available on a restricted basis, but researchers are able to access the data at their own university.

⁶ Those of us who are used to nearly all establishment data being confidential may find the experience with regulatory data in other fields inspirational. For example, an EPA database called PCS (Permit Compliance System) has information on firms that have Clean Water Permits. It provides the name and address of firm, SIC code, name of manager (!), permit conditions, actual versus allowed discharges of every controlled parameter, and information about inspections.

A relatively uncontroversial margin for extending the matching of employer and worker data is to take advantage of data that is already available by firm or establishment. For example, the ERISA data mentioned above could be matched to survey data to see how problems with firms' defined-benefit pension plans have impacted workers. As another example, one can "look up" employers of survey respondents to obtain employment at the worker's place of business, at the broader "firm" level, a rough measure of recent employment growth, company sales, and detailed industry – all from Dun and Bradstreet's – at \$5 per observation.⁷ This would, e.g., let one identify workers in trade-impacted industries more accurately than matching to 3-digit industry coded from workers' descriptions of what their employer produces.

Whether this can be used as a model for federal agencies remains to be seen, but it might be useful to sketch out an example where the disclosure risks would be small and the benefits significant. Commercial vendors already provide current employment and sales data, as noted above. Researchers often need longitudinal data at the establishment or firm level, which commercial vendors provide much less adequately. If researchers would often benefit from "old" establishment or firm data, and these data are much less sensitive than data that are already in the public domain, permitting links to old data on a carefully monitored, restricted basis seems worth serious consideration.

7. Conclusions

Longitudinal data have played an important role in the five policy areas considered in this paper. Nevertheless, the impact of research based on longitudinal data is constrained by a fundamental and nearly-inherent tension – policy research often demands prompt "answers", and longitudinal data are collected in real time. The examples presented above include a range of responses – and of successes – in dealing with this tension.

One best-case scenario is where the intervention is brief, the impacts at issue are short-run, and the data needed for evaluation are routinely-collected administrative records. The experiments that preceded TANF are a good example of this happy coincidence. Job search assistance is provided promptly, the main intended effect was to quickly move its recipients off of welfare, and the record-keeping needed for the program goes a long way toward providing the data needed for evaluation. If the goal of the program had been to inspire the children of recipients to seek PhDs, the decision to adopt TANF would have been completely unencumbered by relevant data. Similarly, creation of new job training programs tends to precede careful evaluations of the program being replaced.⁸

Another favorable environment for longitudinal data having policy relevance is when a program manages to survive long enough that evaluation of its early incarnations can still be informative. Head Start might be an example here: the earliest cohorts of Head

⁷ See <http://www.zapdata.com> .

⁸ The Unemployment Insurance experiments seem to have had less impact than one might have expected, given that their hoped-for effects were quite immediate and the relevant data already routinely collected.

Start participants have reached adulthood, and later cohorts have progressed far enough in school that the medium-term impacts on achievement can be assessed. Nevertheless, we do not yet have evaluations of “typical” Head Start programs based on randomly assigned control groups—it seems to be easier to mount such evaluations for experimental programs than ongoing ones.⁹ If program administrators define priority classes of applicants and select randomly when they cannot serve everyone in the marginal priority class (as in the Kentucky re-employment service evaluation), this difficulty may be overcome.

A third example is where ongoing longitudinal data programs contain the data needed to address current policy questions. For example, PSID was originally created to study poverty-related problems. It survived long enough to permit the very informative work on duration of spells of welfare receipt, and is beginning to be used to assess the impact of more recent welfare reforms. By extending their original focus to the children of respondents, both PSID and NLSY provide evidence on consequences of poverty, welfare receipt, and maternal employment on children. Having survived long enough for its initial cohort to reach retirement age, HRS has become the weapon of choice in many policy discussions related to retirement.

Another tension between what policy requires and what researchers can deliver is the policy-maker’s interest in differential impacts: which TANF recipients are the most promising candidates for job-search assistance, and which should receive training? Does Head Start provide larger benefits for more disadvantaged children? Overall, I sense that these sorts of perfectly sensible questions have often been beyond the reach of researchers, even when otherwise-adequate longitudinal data are available.

Different “types” of longitudinal data have contrasting strengths and weaknesses. Surveys provide considerable flexibility of content, but they are expensive and sometimes slow. Administrative data from social programs can deliver large samples, often provide more reliable data, can sometimes be mined quickly, but often lack information that is more important to the analyst than to the ongoing operation of the program. Data from personnel records are often much more detailed than other sources, but leave one wondering about the representativeness of the findings. Moreover, their availability usually depends on a serendipitous relationship between the firm and a researcher.

Experimental evaluations provide a more complicated mix of strengths and weaknesses. Because they often assign eligible applicants to treatment or control groups randomly, the appropriateness of the control group is not in question, and transparent statistical methods can recover the program’s impact. However, impacts of a large-scale, permanent program may well differ from those of a temporary, small-scale experiment. In the context of the Negative Income Tax experiment, for example, a temporary program generates a temporary “sale on leisure”, so that recipients who are tempted to withdraw from the labor force at some point in their foreseeable future are encouraged to bunch

⁹ An evaluation of short-term impacts (school readiness), led by WESTAT, was begun in October 2000, but at least as of January 2003 the selection of sites was “ongoing” (Blum, 2003, p. 4). The National JTPA Study had serious difficulties recruiting sites willing to cooperate with the evaluation.

such withdrawals in the period of the experiment. On the other hand, participants might be reluctant to give up a good job that they would need again once the experiment was over, but be quite willing to sever the tie if a permanent program were in place. In the first case, the experiment exaggerates and in the second case understates the effect of a permanent program. Moreover, to the extent that a permanent program led to reduced labor supply, wages would presumably rise, tempting some of the participants (or those not eligible for the program) back into the labor force; a small experiment would not have these effects. Another difference is that the effects of experimental treatments given to those receiving transfers (welfare or UI) on those who might apply for benefits cannot be deduced from the behavior of those who receive the treatment (because, by definition, they have already applied).

That different empirical strategies offer different strengths and weaknesses, with no one strategy dominating, for generating policy-relevant research should come as no surprise. Ideally, different sources of data can be combined, and conclusions from competing strategies contrasted. Ultimately, the quality of policy-related research depends on the data that are available, and so is itself a policy decision.

References

Anderson, Patricia M.; Gustman, Alan L; and Steinmeier, Thomas L. "Trends in Male Labor Force Participation and Retirement: Some Evidence on the Role of Pensions and Social Security in the 1970s and 1980s," *Journal of Labor Economics*, vol. 17, no. 4, part 1, October 1999: 757:783.

Ashenfelter, Orley. "Discussion" [of "The Work Response to a Guaranteed Income: A Survey of Experimental Evidence," by Gary Burtless] in Alicia H. Munnell, ed., *Lessons from the Income Maintenance Experiments*, Federal Reserve Bank of Boston Conference Series No. 30, 1986: 53-55.

Ashenfelter, Orley; Ashmore, David; and Deschênes, Olivier. "Do Unemployment Insurance Recipients Actively Seek Work? Randomized Trials in Four U.S. States," NBER Working Paper 6982, February 1999. Forthcoming in *Journal of Econometrics*.

Bane, Mary Jo and Ellwood, David T., "Understanding Welfare Dynamics," in Mary Jo Bane and David T. Ellwood, eds., *Welfare Realities* (Cambridge MA: Harvard University Press, 1994): 28-66.

Black, Dan A.; Smith, Jeffrey A.; Berger, Mark C.; and Noel, Brett J. "Is the Threat of Reemployment Services More Effective than the Services Themselves? Evidence from Random Assignment in the UI System," *American Economic Review*, forthcoming.

Blank, Rebecca M. "Evaluating Welfare Reform in the United States," *Journal of Economic Literature*, vol. 40, no. 4, December 2002:1105:1166.

Blank, Rebecca M. and Ruggles, Patricia. "Short-Term Recidivism Among Public Assistance Recipients," *American Economic Review*, vol. 84, no. 2, May 1994, pp. 49-53.

Bloom, Dan and Michalopoulos. *How Welfare and Work Policies Affect Employment and Income: A Synthesis of Research* (New York: Manpower Demonstration Research Corporation, 2001).

Blum, Barbara B. "Research on Welfare Programs Important During a Period of Uncertainty," *The Forum*, no. 6, vol. 1, January 2003. Available at <http://www.researchforum.org>.

Blumrosen, Alfred W. and Blumrosen, Ruth G. *The Reality of Intentional Job Discrimination in Metropolitan America, 1999*. (Newark: Rutgers University Law School, 2002). Available at <http://www.rci.rutgers.edu/~nwklaw/blumrosen/Title.pdf>

Bound, John; Brown, Charles; and Mathiowetz, Nancy. "Measurement Error in Survey Data," in James J. Heckman and Edward Leamer, eds., *Handbook of Econometrics*, vol. 5 (Amsterdam: Elsevier, 2001): 3705-3843.

Burtless, Gary. "The Work Response to a Guaranteed Income: A Survey of Experimental Evidence," in Alicia H. Munnell, ed., *Lessons from the Income Maintenance Experiments*, Federal Reserve Bank of Boston Conference Series No. 30, 1986: 22-52.

Choi, James J.; Laibson, David; Madrian, Brigitte; and Metrick, Andrew. "For Better or For Worse: Default Effects and 401(k) Savings Behavior," NBER Working Paper 8651, December 2001.

Coile, Courtney and Gruber, Jonathan. "Social Security Incentives for Retirement," in David Wise, ed., *Themes in the Economics of Aging* (Chicago: University of Chicago Press, 2001): 311-341.

Currie, Janet. "Early Childhood Education Programs," *Journal of Economic Perspectives*, vol. 15, no. 2, Spring 2001: 213-238.

Currie, Janet and Thomas, Duncan. "Does Head Start Make a Difference?" *American Economic Review*, vol. 85, no. 3, June 1995: 341-364.

Danziger, Sandra. "Why Some Women Fail to Attain Economic Security," *The Forum*, vol. 4, no. 2, August 2001. Available at <http://www.researchforum.org>.

Decker, Paul T.; Olsen, Robert B.; Freeman, Lance; and Keplinger, Daniel H. "Assistent Unemployment Insurance Claimants: The Long-Term Impacts of the Job Search Assistance Demonstration," Mathematica Policy Research, February 2000. Available at http://www.ows.doleta.gov/dmstree/op/op2k/op_02-00.pdf.

Gleason, Philip; Rangarajan, Anu; and Schochet, Peter. "The Dynamics of Receipt of Aid to Families with Dependent Children among Teenage Parents in Inner Cities," *Journal of Human Resources*, vol. 33, no. 4, Fall 1998: 988-1002.

Gustman, Alan and Steinmeier, Thomas. "How Effective is Redistribution Under the Social Security Benefit Formula?" *Journal of Public Economics*, vol. 82, no. 1, October 2001: 1-28.

Gustman, Alan and Steinmeier, Thomas. "The Social Security Early Entitlement Age in a Structural Model of Retirement and Wealth," NBER Working Paper 9183, September 2002.

Hall, Robert E. "Discussion" [of "The Work Response to a Guaranteed Income: A Survey of Experimental Evidence," by Gary Burtless] in Alicia H. Munnell, ed., *Lessons from the Income Maintenance Experiments*, Federal Reserve Bank of Boston Conference Series No. 30, 1986: 56-59.

Heckman, James J., Lalonde, Robert J., and Smith, Jeffrey A. "The Economics and Econometrics of Active Labor Market Programs," in Orley C. Ashenfelter and David

Card, eds., *Handbook of Labor Economics*, vol. 3A (Amsterdam: North Holland, 1999): 1865-2097.

Hurd, Michael. "Research on the Elderly: Economic Status, Retirement, and Consumption and Saving," *Journal of Economic Literature*, vol. 28, no. 2, June 1990: 565-637.

Krueger, Alan B and Meyer, Bruce D. "Labor Supply Effects of Social Insurance," NBER Working Paper 9014, June 2002. Forthcoming in Alan Auerbach and Martin Feldstein, eds., *Handbook of Public Economics*.

Lalonde, Robert J. "Evaluating the Econometric Evaluation of Training Programs with Experimental Data," *American Economic Review*, vol. 76, no. 4, September 1986: 604-620.

Leonard, Jonathan. "The Impact of Affirmative Action Regulation and Equal Employment Law on Black Employment," *Journal of Economic Perspectives*, vol. 4, no. 4, Fall 1990: 47-64.

Madrian, Brigitte C. and Shea, Dennis F. "The Power of Suggestion: Inertia in 401(k) Participation and Savings Decisions," *Quarterly Journal of Economics*, vol. 106, no. 4, November 2001: 1149-1187.

Manser, Marilyn E. "Existing Labor Market Data: Current and Potential Research Uses," in John Haltiwanger, Marilyn E. Manser, and Robert Topel, eds., *Labor Statistics Measurement Issues* (Chicago: University of Chicago Press, 1998).

Meyer, Bruce D. "Lessons from the U.S. Unemployment Insurance Experiments," *Journal of Economic Literature*, vol. 33, no. 1, March 1995: 91-131.

Mitchell, Olivia S. and Phillips, John W.R. "Retirement Responses to Early Social Security Benefit Reductions," NBER Working Paper No. 7963, October 2000.

Morris, Pamela A., et al., *How Welfare and Work Policies Affect Children: A Synthesis of Research* (New York: Manpower Demonstration Research Corporation, 2001).

National Research Council and Institutes of Medicine. *Early Childhood Intervention: Views from the Field*, edited by Jack P. Shonkoff, Deborah A. Phillips, and Bonnie Keilty (Washington DC: National Academy of Sciences, 2000).

Papke, Leslie E., "Participation in and Contributions to 401(k) Pension Plans: Evidence from Plan Data," *Journal of Human Resources*, vol. 30, no. 2, Spring 1995: 311-325

Papke, Leslie E., "Are 401(k) Plans Replacing Other Employer-Provided Pensions? Evidence from Panel Data," *Journal of Human Resources*, vol. 34, no. 2, Spring 1999: 346-368.

Rust, John and Phelan, Christopher. "How Social Security and Medicare Affect Retirement," *Econometrica*, vol. 65, no. 4, July 1997: 781-831.

Stafford, Frank P. "Forestalling the Demise of Empirical Economics: The Role of Microdata in Labor Economics Research," in Orley Ashenfelter and Richard Layard, eds., *Handbook of Labor Economics*, vol. 1 (Amsterdam: North Holland, 1986): 387-423.

St.Pierre, R., Layzer, J., Goodson, B., and Bernstein, L. *National Impact Evaluation of the Comprehensive Child Development Program: Final Report*. Cambridge, MA: Abt Associates Inc., June 1997 available at:
http://www.acf.dhhs.gov/programs/core/pubs_reports/ccdp/ccdp00.html

Venti, Steven F. and Wise, David A. "Lifetime Income, Savings Choices, and Wealth at Retirement," in James Smith and Robert Willis, eds., *Wealth, Work, and Health: Innovations in Survey Measurement in the Social Sciences* (Ann Arbor: University of Michigan Press, 1999): 87-120.