Title

Defining Welfare Spells: Coping with Problems of Survey Responses and Administrative Data

Authors

Brady, Henry E.
Luks, Samantha

Publication Date

1995-09-01

Defining Welfare Spells: Coping with Problems of Survey Responses and Administrative Data

Henry E. Brady and
Samantha C. Luks
Department of
Political Science
University of California,
Berkeley

UC DATA
University of California
2538 Channing Way #5100
Berkeley, CA 94720-5100
FAX (510) 643-0663

September 3, 1995
Working Paper #6

Defining Welfare Spells:

Coping with Problems of Survey Responses and Administrative Data

Henry E. Brady
and
Samantha C. Luks

Department of Political Science
University of California, Berkeley
Berkeley. California 94720

hbrady@bravo.Berkeley.Edu
sam@ucdata13.Berkeley.Edu

Comments Invited

Prepared for delivery at the 1995 Annual Meeting of the American Political Science Association, Chicago Hilton and Towers, Chicago, Illinois, August 30 to September 3, 1995

Abstract

We explore both how to define a spell of welfare and how well surveys measure welfare spells. Comparing survey and administrative data on the receipt of AFDC from the Work Pays Demonstration Project in California, we find that a substantial amount of administrative churning occurs in administrative data and that a one-month definition for an interruption in welfare is probably not sufficient. Through a mixing model of several break lengths, we find that a single definition of a break in welfare is not applicable to all respondents and that two month breaks are correct for some respondents while three month breaks are best for other respondents. Finally, we conclude that the complexity of defining an accurate break in spells creates difficulties for detecting biases in survey responses.

I. How Should We Define a Welfare Spell?

A. Spell Durations and the Deserving Poor

Distinguishing between the deserving and undeserving poor has been a central theme of American and European poverty policy (Katz, 1986; Trattner, 1974). Torn between a desire to help those in need and worries about encouraging sloth or immorality, policy-makers have treated some groups as deserving ostensibly because they cannot help being poor and others as undeserving because they supposedly have the capacity to overcome their plight or because they have behaved in ways deemed to have brought their poverty upon themselves. Yet one era's deserving poor may be another's undeserving poor. The Aid to Families with Children (ADC) program of the 1935 Social Security Act offered a solution to this perplexing problem by singling out single indigent mothers, mostly widows with children, as the deserving poor. As times have changed, ADC (and its successor, AFDC, Aid to Families with Dependent Children) has enrolled increasing numbers of unwed mothers and doubts have developed about whether this group deserves the same treatment as widows with children. With the increasing proportion of women in the workforce, these doubts have been compounded by a feeling that even mothers with children should be working (Smolensky et al, 1992). Welfare mothers, never the most popular group in our society, have declined even farther in public esteem as they have been pictured as feckless and lazy.[1]

These changes have undermined the legitimacy of AFDC, and they have forced a reexamination of the basis for providing welfare benefits. Social science research has aided in this reexamination by developing new tools, the analysis of welfare durations, which increase our understanding of who gets on welfare, how long they stay on, and why they get off. One of the first findings of this research (Bane and Ellwood, 1983) was the realization that a majority of welfare recipients had short spells of welfare and that only a minority had spells extending beyond two years. This finding seemed paradoxical because in a cross-section of welfare recipients, the majority consists of those who have been on more than two years. There is no logical paradox, however, because long-termers are much more likely to show up in a cross-section than those with short spells. Yet there is a political cost because the visibility of these long-termers has been one of the major political liabilities of welfare, and much of the recent concern with welfare has focussed on reducing the length of their spells. The 1988 federal welfare reforms emphasized job training and job seeking as ways to

---

[1] In addition, the welfare population is now comprised of a large number of minority recipients. There is some persuasive evidence that racism has contributed to the unpopularity of welfare in America (Gilens, 1991).

provide opportunities for this group to leave welfare. This approach has had some limited success, but it has proven very difficult to move poorly educated women with one or more children off welfare. More recently, as a candidate for the Presidency in 1992, Bill Clinton proposed a two year limit on welfare, and others have proposed even more stringent limits on welfare recipients.

It would be foolish to think that social science research is responsible for the recent flurry of attention to welfare and for the increasingly harsh approach to welfare in the Republican Congress and in many states. But social science research certainly provided part of the terms of the debate with its findings about differences in spell durations. The duration of spells has become a convenient way to distinguish the deserving from the undeserving poor. Those who have short spells are thought to be more deserving than those who have longer spells. And presumably those who try to leave welfare by getting off for a while are more deserving than those who simply remain on welfare.

Yet the simple truth is that we do not have a good definition of what constitutes a spell of welfare and we do not have very good data on welfare spells. Despite the high-quality of the research on welfare spells, it is limited by the nature of the data sources and by a shaky definition of spell duration. The data sources are often high-quality national surveys that are forced to rely upon recall about the receipt of welfare benefits. In the earlier waves of the Panel Survey of Income Dynamics, for example, respondents are asked to recall if they received welfare in the past year and how much they received.[2] This approach is subject to the fallibility of memory and possibly social pressures to minimize the reported length of time on welfare. Furthermore, it only provides yearly data on receiving welfare. The Survey of Income and Program Participation undertaken by the Bureau of the Census does a somewhat better job by getting monthly data on program participation by interviewing respondents every three months for three years, but the problems of recall remain and the three year limit on observations create substantial censoring of the data.

Partly because of the limitations of the data sources, most research on welfare spells has relied upon flawed definitions of welfare receipt. Bane and Ellwood (1983) establish the criterion that an AFDC recipient must have received at least 250 dollars in aid payments in a calendar year in order to be considered on aid for that year. These payments, however,

---

[2]More recent waves of the PSID have added questions about monthly AFDC receipt. The series of questions in the 1987 PSID is as follows:

"Did you [head] receive any income from ADC or AFDC?"

[If yes]

"How much was it?"

"During which months of 1986 did you get this income?"

could have been received in one month, several months, or over a whole year. Ellwood's (1986) update of this article uses the measurement of total time on welfare during a fifteen year period. This method accounts for the fact that measuring a single welfare spell may not be adequate. In fact, more than one third of recipients have more than one spell (Bane and Ellwood). However this type of measurement falls short of capturing the true dynamics of welfare. Some individuals spend long periods of time on welfare with only sporadic breaks. Others have series of short spells. Substantively it is important to distinguish what underlies different patterns of welfare participation. For example, for those with long, uninterrupted spells of welfare one might want to investigate what prevents them from leaving welfare, whereas for cases with many short breaks, one might want to study what goes wrong every time they try to leave welfare.

Monthly data of welfare activity been utilized in some recent studies of welfare spells (Harris 1992, Blank 1989, Plotnick 1983, for example). This type of analysis allows one to observe spell lengths and repeated spells more accurately. However, the information is still obtained through surveys and the questions must rely upon recall---often for events as much as twelve months ago (as in the case of the PSID). As we show in the next section, this leads to serious worries about the reliability of the definition of a spell.

B. Memory of autobiographical events and facts

Survey respondents are often asked to recall events that happen or have happened in their lives. Often these questions revolve around matters of work and income (such as one's earnings or the amount of time one has spent looking for a job), or personal characteristics (such as the number of children one has, the length of time one has lived in one's current residence, or one's age). Respondents may be asked to recall events in their lives, such as the age one when they had their first child or whether they voted in the last election. Occasionally respondents are asked to describe behavior in which they have engaged that may reflect on them personally, such as church attendance (Hadaway, et al. 1993), condom usage (Williams and Suen 1994) or drug and alcohol use (Aquilino 1994).

Random error is to be expected in the responses to all these types of questions. We may also anticipate two types of bias to occur in the responses: telescoping and social desirability bias. When respondents are required to estimate when an event occurred in the distant past, telescoping may affect their responses, causing respondent to believe that events in the past have occurred more recently than they actually did. Moreover, when respondents are asked about behavior about which they are sensitive, their answers may be colored by social desirability bias. In this situation, respondents tend to overreport socially desirable behavior in which they have engaged, and underreport behavior which may cause them to be socially stigmatized.

All of these phenomena, random error, telescoping, or social desirability bias can seriously distort data on welfare spells. We have reason to believe that respondents may be unwittingly biasing distant breaks in welfare participation. Telescoping of both autobiographical and historical information has been documented in several studies of survey response error. What appears to occur is that respondents often allow the clarity of their memory of an event to bias the perception of when the event occurred. Events about which respondents remember many details are perceived to have happened more recently than they actually have. Similarly, events dimly remembered are perceived as occurring farther in the past than they have (Brown, et al 1985). Salient personal events also tend to be dated more recently (Wagenaar 1986). Given that respondents were aware that they were selected to be surveyed because they were on welfare and that they had already been asked a number of questions about current household and AFDC status, some respondents may have been thinking more about their welfare histories, thereby leading them to believe that their spells started more recently than they had.[3]

Social desirability may also play a factor in biasing the survey responses. Use of welfare is currently widely stigmatized, its recipients often being characterized as lazy, cheating the government, or having more children in order to get higher payments. Such stereotypical conceptions of AFDC recipients may cause respondents to underestimate the length of their current spells because of a social desirability bias, "in which cognitive dissonance can lead to a rather consistent distortion of memory in order to reinforce continued perception of oneself as a good citizen."[4] Theoretically, the extent of this bias would vary with the extent to which the respondent found AFDC use undesirable. Furthermore, longer spells may be more affected than shorter spells.

Much of the research done about social desirability biases concerns the validity of self-reports about voting behavior, the assumption being that voting is a desirable behavior and

---

[3]The series of survey questions on our survey is as follows:
"Are you [or name of R's spouse/partner] currently receiving AFDC either for you or any of the children living with you?"
"Are you receiving AFDC for your own children (including biological and step-children), someone else's children, or both your own and someone else's children?"
"How old were you the first time you received AFDC?"
"Have you been getting AFDC ever since then or has there been at lease one time when you were off AFDC?"
[if off at least once] "How long have you been receiving AFDC this time?"

[4] Don Cahalan, 1968, "Correlates of Respondent Accuracy in the Denver Validity Survey," *Public Opinion Quarterly*, Winter, Volume 32, p. 621.

nonvoters will claim to vote when they have not. Evidence of this behavior has been limited. Traugott and Katosh (1979) find that nearly 15 percent of survey respondents misreport voting, those most likely to do so being in young, nonwhite, and low income groups. In most cases they report voting when they did not. Roos and Grant (1985) point out that part of the reason that these groups appear to misreport more is that there is a larger group of true nonvoters among these subgroups. Hence, there is a greater chance for people in these groups to misreport in the socially desirable direction. Some studies find that those who vote and those who profess to vote are similar. Volgy and Schwarz (1984) find that in terms of both background and attitudinal characteristics, "...both undocumented and documented participants may come from a similar 'pool' of potential activists, and that in these characteristics both groups can be distinguished from nonparticipants."[5] If, indeed, demographic differences in misreporting have been overstated, and participants and misreporting nonparticipants have similar values, these findings may be relevant to predicting misreporting the length of time one has spent on welfare.

Weiss's (1968-69) study of vote misreporting among women on welfare studied shows many of the same patterns as the other vote misreporting studies. The biased respondent is:

> ...high on community integration. That is, she has lived in the neighborhood longer, is older, better educated, has had longer work experience, and spent her childhood in an urban place. Secondly, she is oriented to middle class values.... The biased respondent is likely to reject welfare status for herself and her children, to value education, and to indicate a future time orientation.[6]

Her findings imply a relationship between one's aspirations and one's claims beyond the typical demographic differences found in other studies. Furthermore, when only certain groups misreport, the perception of an association is created. For instance, if we were to look only at survey data, and respondents who lived in urban areas were more likely to display a social desirability bias than respondent who lived in rural areas, we might come to the incorrect conclusion that urban welfare recipients have shorter spells.

At a minimum, these effects might lead to incorrect estimates of the length of welfare spells. Telescoping can lead to reports of shorter spells than are actually the case. Social desirability bias might lead to the same result as welfare recipients try to minimize their

---

[5]Thomas J. Volgy and John E. Schwarz, "Misreporting and Vicarious Political Participation at the Local Level," *Public Opinion Quarterly*, Volume 48, p. 760. See also Hill and Hurley (1984).

[6]Carol H. Weiss, "Validity of Welfare Mothers' Interview Responses," *Public Opinion Quarterly*, Winter 1968-1969, Volume 32, p. 626.

length of time on welfare. These effects, however, might not have any impact on multivariate analyses of spell lengths unless the extent of telescoping or the amount of social desirability bias is correlated with factors, such as education or age, which might be considered determinants of spell length. Unfortunately, we know very little about the correlates of telescoping or social desirability bias so that we cannot know whether they pose problems for multivariate analyses.

II. Description of Data from the Study and descriptive statistics of data.

A. Data Sources for our Study

The obvious solution to these problems is to get better data on the receipt of welfare and to compare it with survey responses. This is exactly what we have done in our study. We have both administrative data on welfare status and a survey in which respondents were asked about the length of their last welfare spell.

Administrative Data---Surely an excellent definition of being on welfare in a particular month is receiving a check from the welfare office. We have two kinds of administrative data of this sort from the California Work Pays Demonstration Project (WPDP). In this project, a stratified random sample of approximately 15,000 AFDC cases was selected from four California counties: Alameda, Los Angeles, San Joaquin, and San Bernardino counties. All cases were on aid during the sampling month (October, 1992), and a case always consists of one or more children plus a mother, father, guardian or some other adult. For each case, two types of administrative data were collected: (1) monthly aid code and payment information from the county welfare office records from December 1992 to June 1994, and (2) statewide MediCal MEDS data containing monthly aid code records from January 1987 to December 1992. The county data are probably the best source of information about welfare status because it is the counties that write the checks to welfare recipients. The MEDS database keeps track of all people who are eligible for California medicaid (MediCal) benefits. Welfare recipients are automatically eligible for MediCal. It is possible that there might be a lag before welfare recipients enrolled in the counties are recorded in the MediCal database, and it is possible that there might be a lag before those going off welfare are removed from the MediCal database.

Table 1 provides some information on the relationship between the two administrative databases by comparing MediCal status with country welfare status for December, 1992. We have four possible statuses for the MediCal database. A case could be in the "Family Group (FG)" program which is primarily for single mothers, in the "Unemployed Parent (U)" program which is for primarily for households with two parents, in the transitional welfare program (T) which is primarily for cases that are not receiving AFDC benefits but are still eligible for MediCal, or in some other status (Other). The county data base does not have

Table 1

Comparison of AFDC codes for County and State MediCal Data
December 1992
(cells are percentage of total)

		MediCal Data			
		FG	**U**	**T**	**Other**
	FG	67.6% (9795)	0.3% (45)	0.2% (25)	0.2% (35)
County Data	**U**	0.1% (21)	30.9% (4485)	0.1% (9)	0.1% (16)
	Other				0.5% (67)

100%
(14.498)

any cases in a transitional status but has the three other categories.

The most restrictive definition of concordance between the two databases is that there should be no off-diagonal entries so that a person in FG on the MediCal database is in FG in the county data (and not in U or Other). By this strict definition, 99.0 percent of the entries agree. A somewhat looser definition is that an FG classified as a U or a U classified as an FG should not be counted as an error, and those on in FG or U in the county data but T in the MediCal data should not be counted as errors. By this looser definition, 99.6 percent of the entries agree. These results strongly suggest that the two databases yield very similar results. This is fortunate because we construct our administrative measure of being on welfare by splicing the two databases together. We use the MediCal database before December, 1992 and we use the county data from December, 1992 onwards. Eventually we will get MediCal data for the period after December, 1992, and then we will do additional comparisons of these two methods of defining welfare participation.

Survey Data---Of the 14,447 cases on aid in December, 1992, 2,214 were interviewed between October, 1993 and September, 1994 about their personal and family characteristics and their work and welfare histories. The surveys were conducted in either English or Spanish. Excluding cases that were off aid at the time of the interview,[7] or were interviewed after June 1994,[8] 1730 cases were available for analysis.

Table 2 shows various characteristics of the remaining 1730 cases. Because of our sampling design, we have representative samples of AFDC cases within each program (FG or U) and within each county. We tried to get equal numbers of respondents (about 450 we hoped) in Alameda, San Bernardino, and San Joaquin and about twice this number from Los Angeles County. We also tried to get about one-third U cases, and two-thirds FG cases. This amounts to an oversampling of U cases within each county. The respondents are predominantly female (98%) because we sought to interview the female head of household in every case unless there was no such person.

The characteristics of the sample reflect the populations from which they came. The respondents are predominantly non-White (72%). About a quarter are African American and 42 percent are Latino. Over 85 percent of them are between the ages of 20 and 44. Approximately one-third of the cases are from Los Angeles County, while the rest are divided among Alameda, San Joaquin and San Bernardino counties. Alameda County with

---

[7] If the person was off aid at the time of the interview then their previous welfare experience was explored using a different set of questions which are not comparable to the ones we discuss in this paper.

[8] We exclude cases interviewed after June, 1994 because our administrative data stops in that month.

Table 2

Characteristics of Survey Sample

		Percent	Frequency	N
Race	Black	26.1%	450	1726
	Native Amer.	1.8	39	
	Latino	41.0	707	
	Asian, Filip.	3.0	51	
	White, Other	28.1	487	
Age	16-19	2.9	50	1730
	20-24	16.5	286	
	25-29	23.9	414	
	30-34	21.8	378	
	35-44	25.5	442	
	45-54	7.0	121	
	55-80	2.3	39	
Sex	Female	97.8	1694	1730
	Male	2.2	36	
County	Alameda	22.6	390	1730
	Los Angeles	33.9	587	
	San Bernardino	18.8	325	
	San Joaquin	24.7	428	
AFDC Type	FG	70.3	1217	1730
	U	29.7	513	

twenty-one per-cent of the respondents comprises much of the highly urbanized East Bay of the San Francisco Bay Area, and it includes the cities of Oakland and Berkeley. San Joaquin is to the east of Alameda county and is in the rural and agricultural central valley of California. Its largest city is Stockton. Los Angeles County comprises the City of Los Angeles, the San Fernando Valley, and many other communities such as Santa Monica, Beverly Hills, Long Beach, and Claremont, California. San Bernardino borders Los Angeles County on the east, and it extends all the way to the Arizona border. In its western regions San Bernardino County is part of the vast urban sprawl of Los Angeles, but in the east it is mostly sparsely settled desert. These four counties comprise about half the welfare population in the state of California.

B. Survey Measures of Spell-Length

In order to ascertain the reported length of the respondent's most recent spell, a series of questions was asked. First the respondent was asked how old she was the first time she first used AFDC. Then she was asked if she had been off of welfare since the beginning of her first spell. For those respondents who had not had a break in their spells, the length of their current spell was calculated by subtracting their age at first spell from their current age. Those respondents who did have a break in their spells were asked if they were currently on welfare and how long ago in years their most recent spell started. If the period of time was less than one year, the respondent was asked how many months she was using aid. Using the date of the interview and the reported spell length, the approximate number of months ago the spell began was constructed.

The survey responses show spells varying in length from one month to 32 years. Figure 1 shows a histogram of the reported spell lengths. Nearly 40 percent report having received AFDC for less than two years, and the mode of cases falls around the two-year range. Only 10 percent of the respondents claim to have been on welfare for more than 10 years and less than 1 percent report to have received welfare for more than 20 years.

C. Administrative Measures of Spell-Length

The definition of welfare spell from administrative data would seem to be straightforward because these are the data of record. All we need to do is to look for breaks in the receipt of welfare. But some recipients may appear to cycle on and off of aid due to administrative problems without actually going "off" aid in the sense of getting a job, getting married or having a child reach the age of 18. In this situation, using a definition of one month off aid as a spell break poses some risk, as cases occasionally appear to be off aid due to "administrative churning," in which case the recipient does not receive a check for that month, but her being off aid has nothing to do with need. Often the recipient will receive a

Figure 1

Histogram of Survey Responses for Most Recent Spell Length in Months

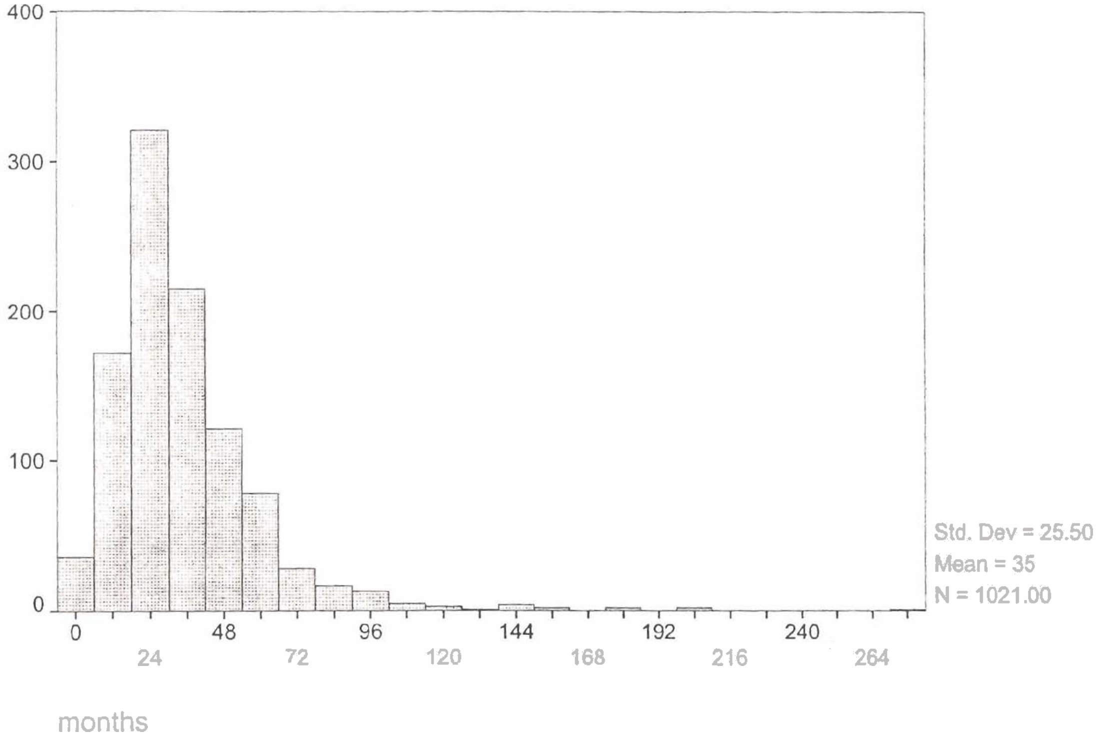
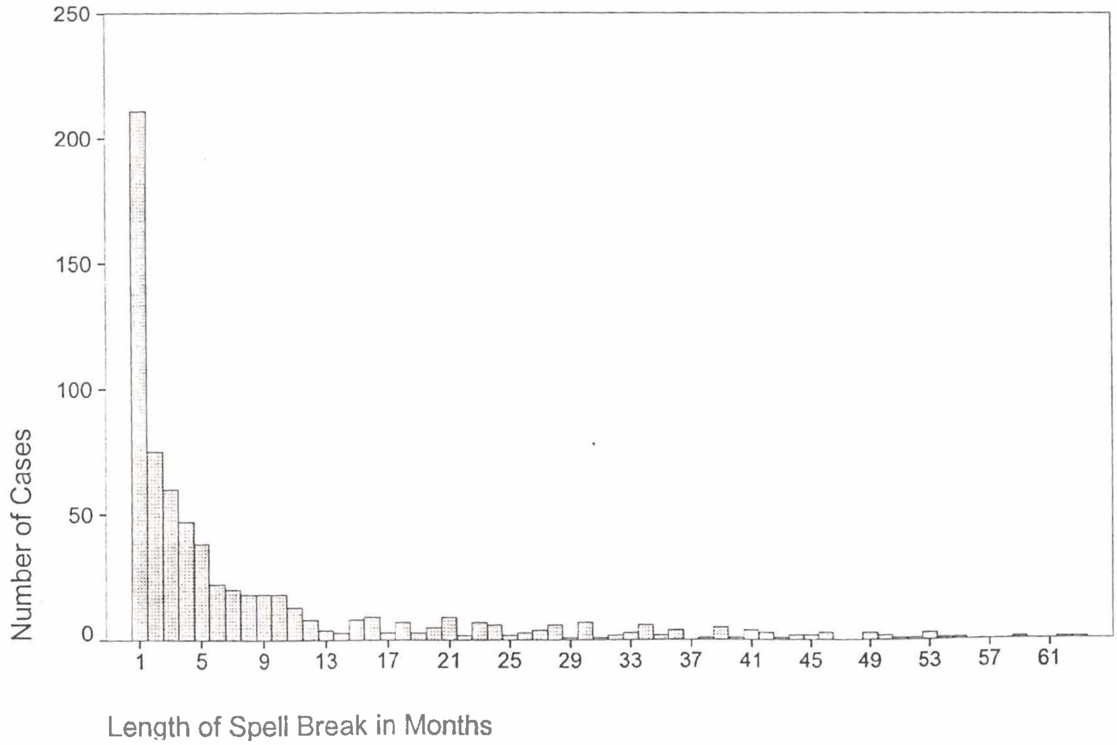


Figure 2

Length of Break in AFDC between Most Recent Spell

and Second Most Recent Spell



double payment the next month. Another scenario that occurs is when recipient may find short-term work or "permanent" work which did not last more than a few weeks. Such work could make the recipient ineligible for aid for that month, but it is less clear if we can conceptually consider the recipient to be "off aid." Furthermore, we do not know if the recipients themselves would *say* that they were off aid in any of these circumstances. All of these factors make the use of one-month spell breaks suspect.

Blank (1989) and Harris (1992) find limited evidence of "churning" in which short breaks of a month or two only amount to some administrative problem such as failing to get forms in or mistakes by the administrative system. This administrative churning may be the exception rather than the rule. Blank also notes that of the few cases with short breaks, many of the breaks appear to have economic causes:

> ...If one looks at the 50 short spells off AFDC [out of a sample of 605], 29 of them are readily explainable: The household clearly has increased income in these months, either through employment, or because of increases in other income sources. Thus, there is little evidence of 'welfare churning in this data, and little justification for assuming that brief periods off welfare are not real spell endings.[9]

Yet other evidence suggests that there might be a substantial amount of administrative churning. A histogram of the length of interruptions (in months) of AFDC spells based upon a 10 percent sample of welfare recipients in the state of California between 1987 and 1992 yields a large percentage of one-month interruptions (over 25 percent of the interruptions) and much smaller percentages of two month, three month, four month, and longer interruptions. Indeed, the spike at one month seems so large that it is does not seem to fit well with the rest of the histogram.

A similar histogram for 692 cases in the APDP survey sample but using the monthly administrative data is presented in Figure 2. This presents the length of the most recent spell break in the administrative data (among survey respondents) for those recipients who have more than one spell during the 90-month period.[10] The spike at one month does not fit well with the rest of the curve. Nearly one third of these cases show a one-month break in aid. If we are to assume that these are the only cases in the entire sample where this occurs, at

---

[9]Rebecca M. Blank, "Analyzing the Length of Welfare Spells," *Journal of Public Economics* 39(1989), p. 257.

[10]For this Figure we can only look at cases with more than one spell during the 90 month period because otherwise we do not know the length of the break before the most recent spell.

least one-fifth (211 out of 1020) show this pattern. This is a much different result from that reported by Harris.

There are several possible reasons that our data differs from Harris's. First, the PSID, which Harris uses, relies on self-reports. Certain types of one-month breaks may be ignored by respondents. Also, the administrative quality of our data may produce some lags in information. The data used for the first six years of our study comes from the state MEDS file. Counties initially collect this data (and all counties have different systems) and then the data is sent to the state. It is possible that some counties have a better success rate at getting the information to the state.

Finally, Susan Ayasse (1994), finds some indications that cases with teenaged eldest members seem to experience more short breaks in aid. If teenagers with children are more likely to have administrative problems because of unfamiliarity with bureaucratic procedures or other administrative difficulties, then we would expect to get this result. None of these results are definitive in either direction. We cannot conclude that churning is not a problem and we cannot conclude that it is a problem. How then are we to use the administrative data to define the length of welfare spells?

It certainly makes a difference whether we use a one month or a six month break as a definition of the end in welfare. Figures 2a through 2f present histograms for spell lengths (in months) using one to six month definitions of spell terminations. Different definitions of a spell break affect the distributions of the spell lengths. Obviously, longer spell break definitions create the appearance of longer spells. When a one-month definition is used, 53 out of 1730 cases have had spells of less than one-half year. When the six-month definition is used, the number of cases in this category is reduced to 11 (See Table 3). For all definitions the mode of cases falls around the two year range, although the number of cases in this group is reduced to 19 percent of the sample from 23 percent when the six-month definition is used instead of the one-month definition.

The number of cases available for analysis is also affected by the definition of an interruption. Because the administrative historical data only goes back to January 1987, a number of the cases are lost to left-censoring. Using the one-month definition retains the largest number of cases. With this definition, only 20 percent of the cases began before January 1987. Because the six month definition is more strict, more than 35 percent of the cases are lost to left- censoring. Another problem that arises is when respondents who say they are on aid in the survey do not appear to be on aid in the administrative data. Shorter definitions of spell interruption increase the chance of this happening. For example, if a respondent appears to not be aid in the administrative data for the two months up to her interview, a one or two-month interruption definition will show that she has reported to be on aid when she was not. Definitions of three months or longer will not show this

Figure 2a

Histogram of Most Recent Spell Lengths in Months

Administrative Data -- 1 Month Definition

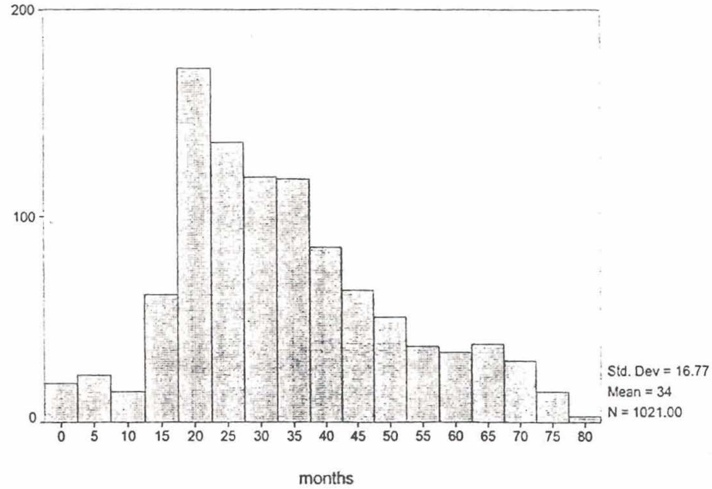


Figure 2b

Histogram of Most Recent Spell Lengths in Months

Administrative Data -- 2 Month Definition

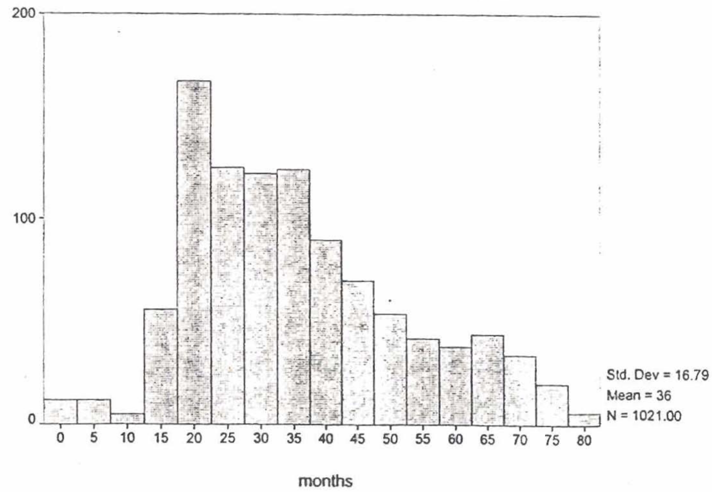


Figure 2c

Histogram of Most Recent Spell Lengths in Months

Administrative Data -- 3 Month Definition

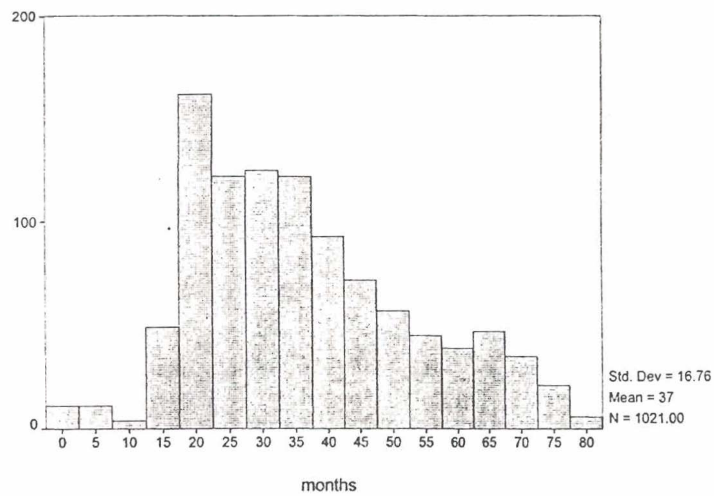


Figure 2d

Histogram of Most Recent Spell Lengths in Months

Administrative Data -- 4 Month Definition

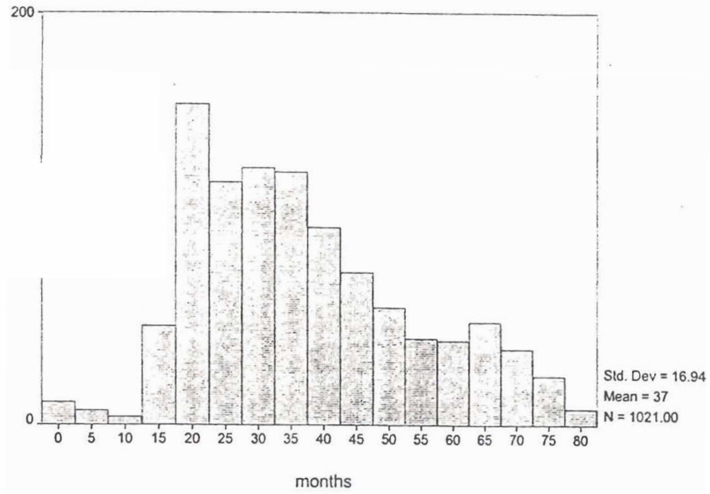


Figure 2e

Histogram of Most Recent Spell Lengths in Months

Administrative Data -- 5 Month Definition

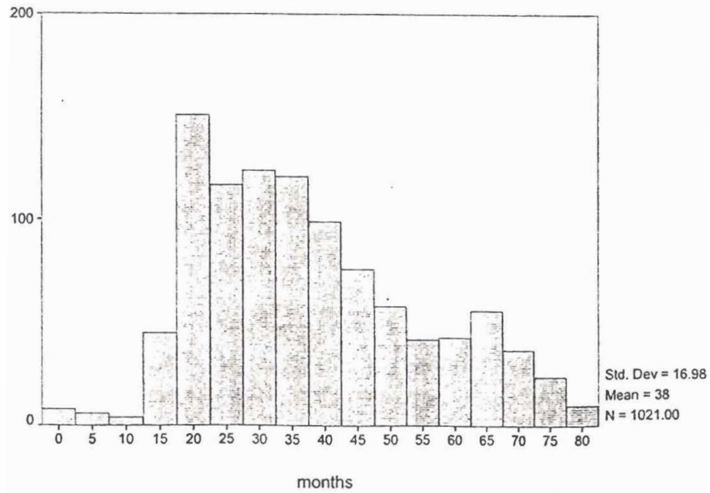


Figure 2f

Histogram of Most Recent Spell Lengths in Months

Administrative Data -- 6 Month Definition

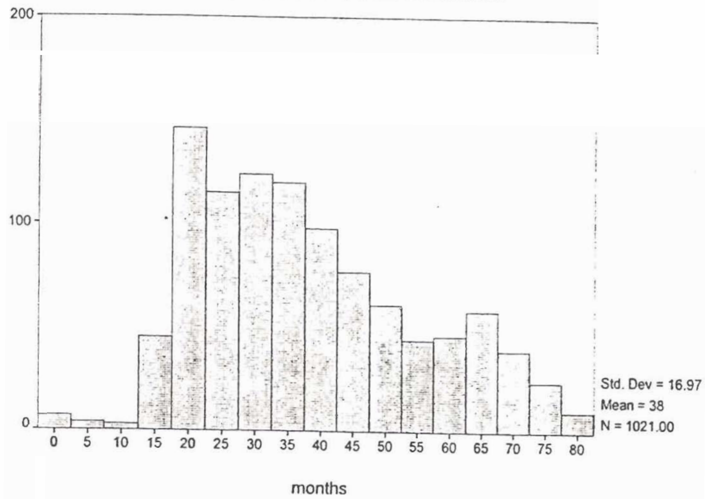


Table 3

Comparison of Six Administrative Definitions for a Spell Break

Definition (months)	1	2	3	4	5	6
Total number of cases	1730	1730	1730	1730	1730	1730
Percent of Cases lost to Left-Censoring	20.3 (352)	26.4 (456)	28.8 (498)	31.1 (538)	33.2 (574)	35.1 (608)
Percent of Cases that Appear to not Be on Aid at Time of Survey	6.7 (117)	5.6 (97)	5.3 (91)	5.0 (86)	4.4 (77)	3.8 (66)
Average Spell Length in months	35.8	37.1	37.5	38.1	38.5	38.4
Percent of Cases Available for Analysis (not left censored and on administrative data)	73.9 (1261)	68.0 (1177)	66.0 (1141)	63.9 (1106)	62.4 (1079)	61.0 (1056)

discrepancy. Nearly seven percent of the cases appear not to be on aid for the one-month definition, while four percent of the cases in the six-month definition show this happening. It is not altogether clear what to do with these cases, as all of them were on aid in the past, but do not appear to be on aid for the several months leading up to the survey. To compare the effectiveness of the different administrative definitions, we have taken a conservative approach and excluded all cases which are left-censored in the six- month definition and excluded all cases which do not appear to be on aid in the one-month definition.

III. Using Survey Data and Administrative Data to Determine Spell Lengths

A. The Basic Questions

The question remains as to which spell break length best represents being off welfare. To explore this, we now turn to a comparison of the survey results with definitions of spell duration based upon different notions of what constitutes an interruption. We have already noted that there is a debate in the literature about the possibility of administrative "churning," which might mean that interruptions of one or two months would simply be administrative mistakes and not real interruptions in welfare. In this situation, presumably welfare recipients would not think of these "interruptions" as true breaks in their welfare experience. Thus, in their survey responses, they would not define their last welfare spell in terms of these interruptions -- they would simply overlook them. This suggests that we could use our respondents' reported welfare duration from the survey as a validity check on the administrative data.

This presents several problems. One problem that we face is that while the administrative data is measured in months, most of the survey data measures spell lengths in years. We are therefore left with the option of whether or not to convert the administrative data into years. We have chosen to keep the administrative data in months because we believe that valuable information will be lost with a yearly measure. We cannot be sure that respondents will round the length of their spells to the nearest year. From interviewer comments, we know that a portion of the respondents have great difficulty quantifying time. Also, such a conversion would reduce the usefulness of the monthly information that we do have about the survey (i.e., the short spells). Therefore, from this type of specification, we should expect noise in the survey measure, but not necessarily bias.

We might also worry that the survey measure, like the administrative measures, has flaws which might make it a poor validity check on administrative measures. We have already noted that the literature on recall of events notes a number of phenomena including telescoping and social desirability bias that might affect a survey response. All of these might affect the reported measure of spell length by shortening the reported spell lengths. In the next section, we develop a model of these effects.

B. A Simple Model of Telescoping and Social Desirability Bias

With so many different phenomena going on in the survey and administrative data, the only way to sort things out is to build a model that simultaneously takes all of the phenomena into account. As with any modeling effort, this requires some assumptions about the way that the phenomena work. Our approach will be to build some simple models, to note the assumptions we have made, and to then go on to weaken the assumptions.

Let Y_i be the survey response for spell length for an individual i and let X_{ij} be the administrative definition of spell length for an individual i and an interruption of j months. Both of these are non-negative numbers because spell-lengths are measure on a time scale. This notation implies that the survey response is the dependent variable and the administrative definition is the independent variable. We make these assumptions for three interrelated reasons. First, there are good reasons to believe that there is a lot of error in a survey response including random errors, telescoping, and social desirability bias. By making the survey response the dependent variable we can regard this error as part of the error term in a regression equation. (If the survey response were treated as the independent variable, then we would have a classic "errors-in-variables" problem that cannot be easily solved by ordinary least squares regression.) Second, it seems possible that at least one of the administrative definitions is the correct definition of spell length so that a regression of Y_i on this X_{ij} will satisfy the normal assumptions for unbiased and consistent estimation using ordinary least squares. Third, we have no left-censoring on the survey response but we have censoring on the administrative data because it only goes back to January, 1987. Because of the left-censoring of the administrative data, we have excluded these cases from our analysis. Excluding these cases because of left-censoring does not create any serious problems for an analysis in which the administrative measure of spell length is the independent variable,[11] but it would create problems if the administrative measure were the dependent variable.

If X_{ij} were the correct definition of spell length and if the survey response Y_i were not subject to any error, then a regression of Y_i on X_{ij} would yield a unitary regression coefficient, a zero intercept, and an R^2 of one. If Y_i has random error, then this regression

[11] This is true if two conditions are met. First, the excluded cases must be from the same population as the rest of the data. This insures that the results we obtain will be true for the excluded cases in the range of the independent variables covered by the included cases. If we wish to be able to make inferences about the full range of the independent variables, then a second requirement must be met. The included cases must provide information about the full range of the independent variables. If the excluded cases are the only ones in some range of the independent variables and if the basic regression assumptions (especially linearity) do not hold in this range, then the exclusion of these cases can lead to incorrect inferences.

should still lead to something close to a unitary regression coefficient, although some complexities arise from the fact that Y_i is non-negative so that the error distribution may be truncated. We will ignore this problem in what follows because there are plenty of other more serious difficulties that must occupy our attention.

If Y_i is subject to telescoping or social desirability bias, then we might expect that the slope and intercept of a regression equation will be affected. Telescoping, for example, would probably reduce the slope coefficient so that spell-lengths as reported on the survey would be some fraction of their true length. Social desirability bias might shorten spell lengths in this way but it might also change the intercept as well as people reduce their reported welfare spell by a fixed number of months. One way to identify these effects, then, is to simply regress Y_i on X_{ij} as follows:

$$(1) \qquad Y_i = \beta_1 + \beta_j X_{ij} + \epsilon_i$$

and to examine the values of β_1 and β_j. Note that in this bivariate context, the regression coefficient β_j will be equal to $\mathrm{Cov}(X_j,Y)/\mathrm{Var}(X_j)$. This will work, however, only if X_{ij} is the correct measure of welfare spells.[12]

C. Empirical Investigation of the Survey Model

The preceding approach assumed a linear form for the relationship between the survey and administrative data, and it assumed that we know the correct administrative measure. In later sections we will develop methods for identifying the correct administrative measure. For the moment we will simply consider the six candidate measures based upon one to six month interruptions, and we will investigate the functional relationship between the survey response and each of these measures. Our goals are to increase our familiarity with the data and to provide a test of the linearity assumptions that we used above and will rely upon in subsequent section.

Our approach is to conduct a series of six bivariate LOWESS regressions[13] with the

---

[12] This is the simplest possible model of social desirability bias and telescoping. A more complex model might assume that α and/or β varied over respondents. Under these circumstances, we would have to make some assumptions about how the random variables α and β varied with true spell length and other characteristics of the respondents that might be correlated with true spell length.

[13] LOWESS initially estimates a local polynomial least squares fit to a subset of the points, then defines robustness weights to resmooth the fit. See Hardle (1989) pp. 192-193. For our data, we used 20 percent of the points for estimation.

survey response as the dependent variable and each of the six administrative measures as independent variables. LOWESS regressions have several advantages over other specifications (such as polynomial regression) we might have used. First, they do not require specifying a functional form. More importantly, they provide robust estimators which are resistant to outliers, unlike parametric functions and many other nonparametric smoothing techniques.

Figures 3a through 3f present scatterplots of the survey responses with each of the administrative definitions. The points are represented using "sunflowers" constructed so that each part of the sunflower represents a specific number of datapoints (five in our case). More complete sunflowers represent more points.[14] Superimposed upon these sunflower plots are LOWESS regressions (in a solid line) and a linear regression (a dashed line) for each one.

There are several notable features of these regressions. First, there is substantial scatter in these data. Second, the largest R^2 for the linear regressions is for the spells based upon a three month definition of an interruption. Third, the LOWESS regressions are very similar to their respective linear regressions, implying that a linear form is an excellent approximation to the data. This suggests that we can proceed by using linear models.

IV. Which Administrative Definition is Correct?

A. A Model with Multiple Administrative Definitions of Spells

What happens if we have a number of candidate measures X_{ij} and X_{ik}? How can we determine which is the best measure? Intuition suggests that a regression of Y_i on each of them might allow us to determine which is the correct, or at least better, measure of welfare spell length. Intuitions, however, can often be very misleading. Let us consider the issue in some detail.

Assume we are estimating a trivariate regression:

$$(2) \qquad Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i$$

where we have assumed without loss of generality that we have two and three month definitions of spell length. Let us further assume that X_{i2} and X_{i3} are related to the "true" spell length X_i as follows:

$$(3) \qquad X_{i2} = X_i + \eta_i \quad \text{and} \quad X_{i3} = X_i + \delta_i.$$

---

[14] Thus, a sunflower with just the central point represents five or fewer points; one with two petals (for a total of three parts) represents between 11 and 15 points.

Figure 3a

LOWESS and Linear Regressions of Self-Reported Spell

Lengths on One-Month Administrative Definition

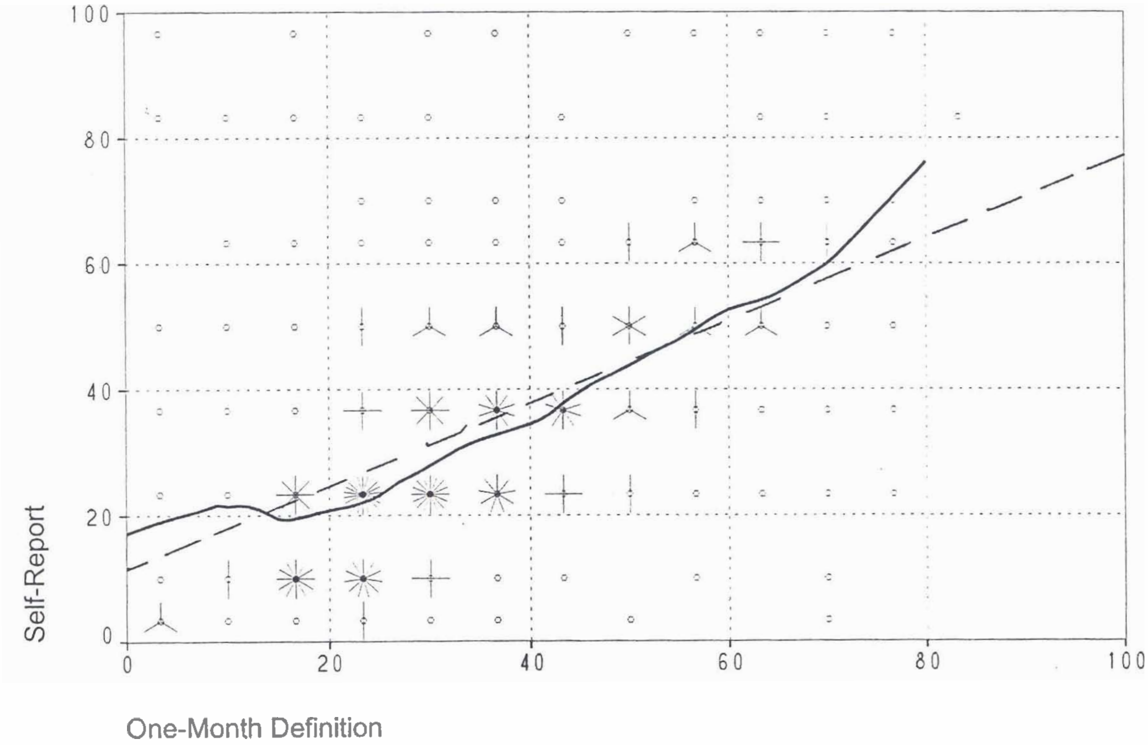


Figure 3b

LOWESS and Linear Regressions of Self-Reported Spell

Lengths on Two-Month Administrative Definition

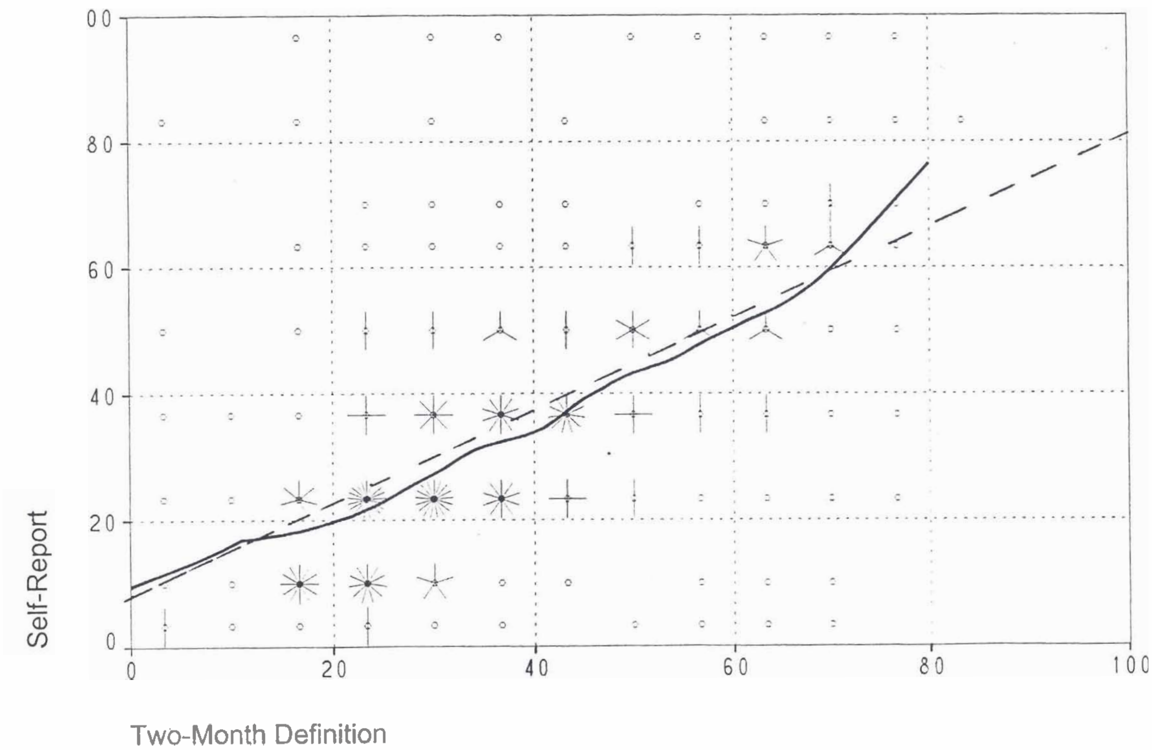


Figure 3c

LOWESS and Linear Regressions of Self-Reported Spell

Lengths on Three-Month Administrative Definition

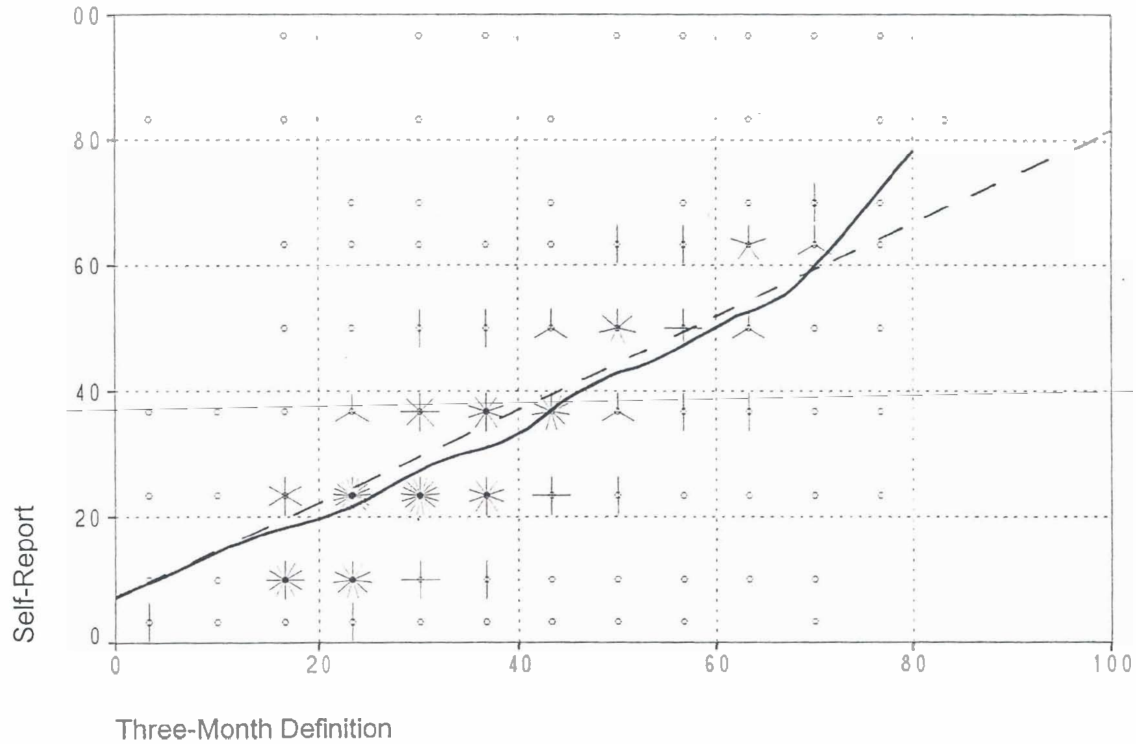


Figure 3d

LOWESS and Linear Regressions of Self-Reported Spell

Lengths on Four-Month Administrative Definition

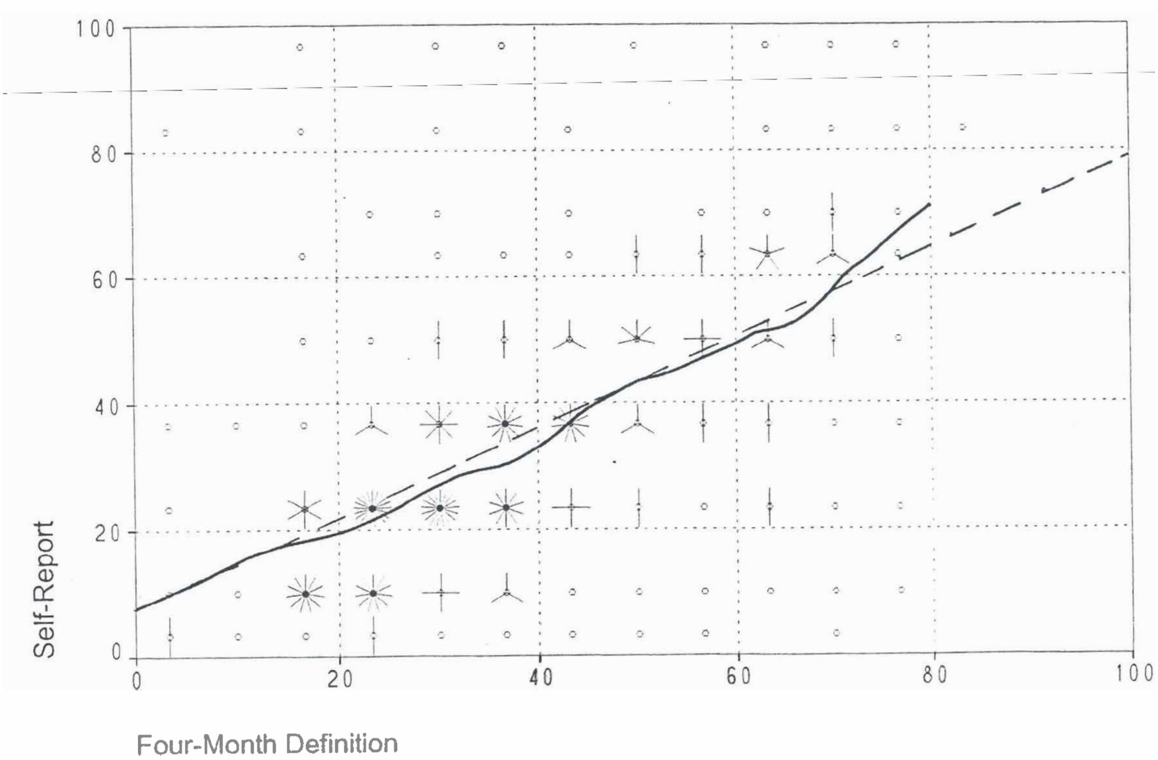


Figure 3e

LOWESS and Linear Regressions of Self-Reported Spell

Lengths on Five-Month Administrative Definition

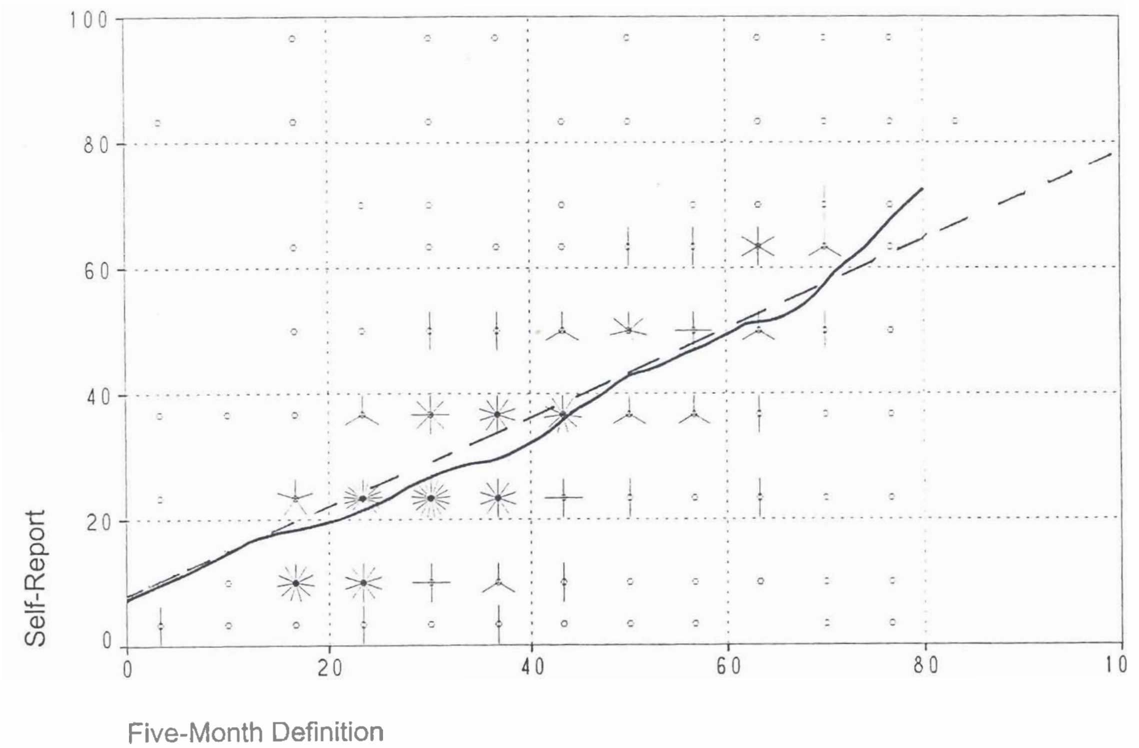
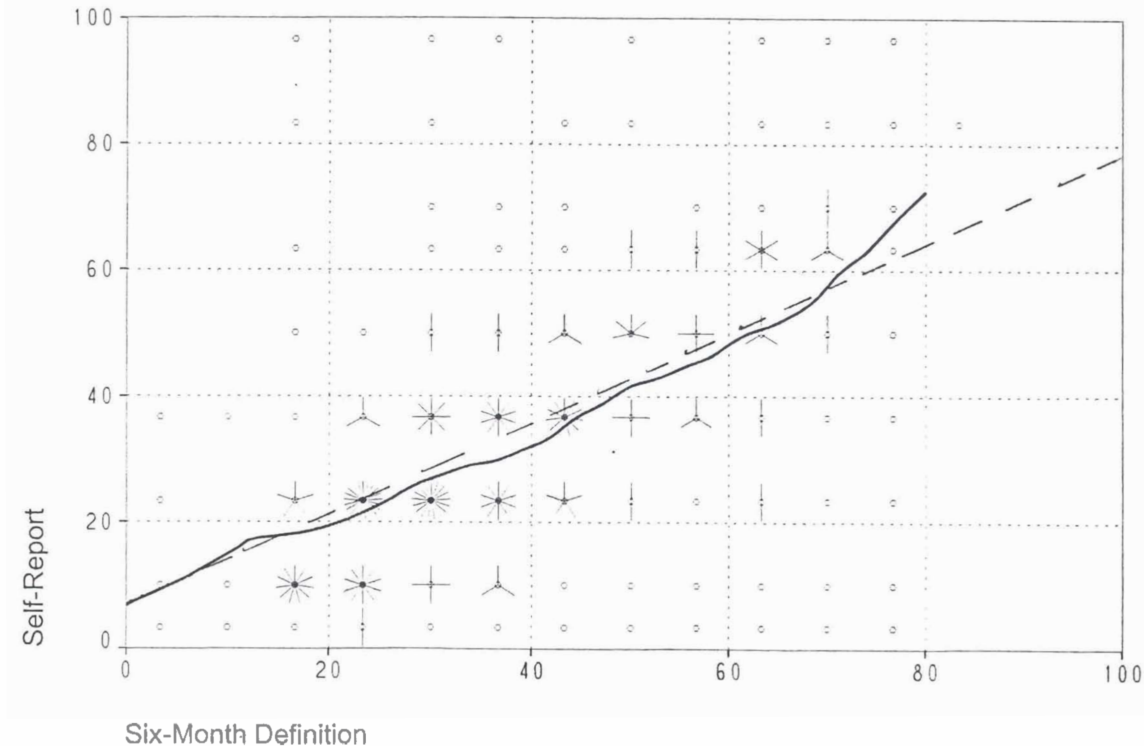


Figure 3f

LOWESS and Linear Regressions of Self-Reported Spell

Lengths on Six-Month Administrative Definition



Without any assumptions about the relationship of η and δ to X, this is trivially true. Let us assume, however, that η and δ are uncorrelated with X. This amounts to assuming that errors introduced by using, say, a two-month definition instead of the correct one are not correlated with the length of the true welfare spell. This assumptions greatly simplifies the following analysis, but it may be unrealistic. It seems possible that those with longer true spells are going to have more chances for administrative churning to occur. Hence, the errors in (3) will be correlated with X. A better assumption might be that errors occur multiplicatively as follows:

$$(4) \qquad X_{i2} = X_i \, \eta_i \quad \text{and} \quad X_{i3} = X_i \, \delta_i$$

with the logarithm of X uncorrelated with the logarithm of the errors. This assumption would complicate the following analysis, although it seems likely that the basic results would survive intact.[15]

For this preliminary analysis, we shall assume that (3) is approximately correct, and we shall ask about the impact of regressing Y on X_2 and X_3 when the true variable is X. The standard formulas for estimates β_2^* and β_3^* of the regression coefficients β_2 and β_3 equation (2) are as follows:

$$(5) \qquad \begin{aligned} \beta_2^* &= [\mathrm{Var}(X_3)\,\mathrm{Cov}(X_2,Y) - \mathrm{Cov}(X_2,X_3)\,\mathrm{Cov}(X_3,Y)]\,[\mathrm{Var}(X_2)\mathrm{Var}(X_3) - \mathrm{Cov}^2(X_2,X_3)]^{-1} \\ \beta_3^* &= [\mathrm{Var}(X_2)\,\mathrm{Cov}(X_3,Y) - \mathrm{Cov}(X_2,X_3)\,\mathrm{Cov}(X_2,Y)]\,[\mathrm{Var}(X_2)\mathrm{Var}(X_3) - \mathrm{Cov}^2(X_2,X_3)]^{-1}. \end{aligned}$$

A bit of computation yields the following values for the variances and covariances in these formulas:

$$(6) \qquad \begin{aligned} \mathrm{Var}(X_2) &= \mathrm{Var}(X) + \mathrm{Var}(\eta) \\ \mathrm{Var}(X_3) &= \mathrm{Var}(X) + \mathrm{Var}(\delta) \\ \mathrm{Cov}(X_2,Y) &= \mathrm{Cov}(X+\eta,Y) = \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X_3,Y) &= \mathrm{Cov}(X+\delta,Y) = \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X_2,X_3) &= \mathrm{Cov}(X+\eta,X+\delta) = \mathrm{Var}(X) + \mathrm{Cov}(\eta,\delta) \end{aligned}$$

---

[15] One way to think about the situation is that our analysis is approximately correct because a Taylor series expansion of logarithms yields a linear approximation as in (3). The problem, of course, is that we do not have a very good grasp of how good the approximation is.

where we have repeatedly used the assumption that the errors η and δ are not correlated with X or with Y.[16] When the results in (6) are substituted in (5), we obtain the following:

$$\beta_2^* = \text{Cov}(X,Y)[\text{Var}(\delta) - \text{Cov}(\eta,\delta)]\,[Q]^{-1}$$
(7)
$$\beta_3^* = \text{Cov}(X,Y)[\text{Var}(\eta) - \text{Cov}(\eta,\delta)]\,[Q]^{-1}.$$

Where Q is the following:

(8) $$Q = \text{Var}(X)\text{Var}(\delta) + \text{Var}(X)\text{Var}(\eta) + \text{Var}(\delta)\text{Var}(\eta) - 2\text{Cov}(\eta,\delta)\text{Var}(X) - \text{Cov}^2(\eta,\delta).$$

A particularly interesting case occurs when one of the error terms in (3) is zero. If, for example, δ is zero so that X_3 is the correct definition of welfare spell, then Var(δ) is zero and Cov(η,δ) is zero so that Q is Var(X) Var(η) and:

$$\beta_2^* = \text{Cov}(X,Y)[0]\,[Q]^{-1} = 0.$$
(7)
$$\beta_3^* = \text{Cov}(X,Y)\text{Var}(\eta)\,[\text{Var}(X)\,\text{Var}(\eta)]^{-1} = \text{Cov}(X,Y)/\text{Var}(X).$$

Therefore, in this case, the regression coefficient β_2^* on the incorrect measure of spell length is zero while the regression coefficient on the correct measure β_3^* is equal to the regression coefficient of Y on X. Thus, β_3^* is a statistically consistent estimate of the impact of telescoping and social desirability bias as in equation (1) above.

Another interesting case occurs when one definition is better than the other in the sense that it is less laden with error. Thus, if the true definition is one month, then the two month definition X_2 will be better than the three month definition X_3 because the two month definition does not include the additional error from assuming that three month interruptions are simply due to churning. We can model this by assuming that:

(8) $$\delta = \eta + \psi,$$

with ψ uncorrelated with η. Then we can write:

---

[16] The assumption that η and δ are not correlated with Y amount to the assumptions that they are not correlated with X and they are not correlated with the error term ϵ in equation (2).

$$(9)\quad \begin{aligned} Var(\delta) &= Var(\eta) + Var(\psi) \\ Cov(\eta,\delta) &= Cov(\eta, \eta + \psi) = Var(\eta), \end{aligned}$$

and Q is:

$$\begin{aligned} (10)\quad Q = {} & Var(X)[Var(\eta) + Var(\psi)] + Var(X)Var(\eta) \\ & + [Var(\eta) + Var(\psi)]Var(\eta) - 2Var(\eta)Var(X) - Var^2(\eta) \\ = {} & Var(X)Var(\psi) + Var(\eta)Var(\psi). \end{aligned}$$

Hence the regression coefficients are:

$$(11)\quad \begin{aligned} \beta_2^* &= Cov(X,Y)\ Var(\psi)/[Var(X)Var(\psi) + Var(\eta)Var(\psi)] = Cov(X,Y) \\ &= Cov(X,Y)/[Var(X) + Var(\eta)] \\ \beta_3^* &= 0. \end{aligned}$$

This means that the worse measure has a zero coefficient and the better one has a non-zero coefficient. Indeed, if the true value is β (equal to Cov(X,Y)/Var(X)), then:

$$(12)\quad \beta_2^* = \beta\ Var(X)/[Var(X) + Var(\eta)] = \beta\ R(X_2)$$

where $R(X_2)$ is the reliability coefficient for X_2. This is the same result as for a bivariate regression with error in the independent variable.

These results suggest that one strategy for finding the best administrative measure of welfare spells is to construct a number of candidate measures, say for interruptions of one month to six months and to undertake trivariate regressions with the survey response as the dependent variable and every possible pair of the administrative measures as independent variables. If one of the administrative measures is the correct one, then it should always prove to have a larger regression coefficient than any of the other measures. Moreover, for any pair of measures, the one that is farther from the true measure should have a zero regression coefficient while the one that is nearer to the true measure should have a non-zero regression coefficient (although the attenuation effect summarized in (12) above might cause it to have a statistically insignificant coefficient as well).

An alternative approach would be to try all the measures at once and to see which does best. The problem with this approach is that it will suffer from a great deal of multicollinearity and sampling fluctuation. The pairwise approach seems more robust.

B. Initial Data Analysis

Based upon the preceding model, it seems sensible to estimate a series of trivariate regressions of equivalent samples (n=1020) with the survey measure as the dependent variable and two of the possible administrative definitions as the independent variables. Table 4 presents the coefficient estimates for all 15 possible trivariate regressions, where the definition of a spell break ranges from one to six months. Note that the coefficients for the one-month and six-month definitions are substantially smaller than they are for the other measures; the largest coefficient for the one-month definition is 0.296 and the largest coefficient for the six-month definition is 0.239. Also, the three-month definition is the only one to consistently have a larger coefficient than any other definition with which it is paired. The only other measure to come close is the two-month definition, which is insignificant only when compared to the three-month definition. Not surprisingly, the four-month definition also performs fairly well, but also loses to the three-month definition. Finally, the three month definition has the highest correlation with the survey measure at .4951. While the differences in correlations are not great, the pattern concurs with the pattern of the trivariate regressions; the two-month definition has the second-highest correlation and the four-month definition has the third highest correlation.

Although these results appear fairly straightforward, there is reason to question them. Over 20 percent of the survey respondents indicated that their welfare spell began before January 1987 (See Table 5). Of these, for 10 percent (2 percent of the total) the administrative definition implies that the spells began after January 1987. This occurrence is not surprising. For example, if a 35 year old respondent has been on AFDC since she was 18, she may forget, or simply disregard, a break of as many as six months long. Also, the reason behind leaving welfare may affect whether or not the respondent sees herself as off welfare. Unfortunately we do not have information about past exits.

If, for whatever reason, a substantial number of respondents do not acknowledge short (i.e. several months) breaks in aid as actually being off welfare, and these respondents have lengthy histories of intermittent welfare participation, the regressions above should show better fits with the administrative data with longer break definitions. Because the second best fit to the survey data is the administrative definition using two-month breaks, it is worthwhile to explore the possibility that the three-month definition is too long. In the next set of trivariate regressions (See Table 6) we compare the same administrative break definitions. However, all cases where the length of the spell is overestimated in the survey by more than seven years are excluded, leaving a sample of 1006 respondents.

Table 4

Trivariate Regressions of Self-Reported Spell Length on Administrative Definitions

Definition	A	B	C	D	E	F	G	H	I
	.1026	.1369	.2325**	.2674***	.2940***				
	.6526***					.3006	.5558***	.5264***	.5232***
		.6296***				.4526*			
			.5279***				1942		
				.5001***				.2302*	
6					.4796***				.2398*
R^2	2401	.2433	.2357	.2365	.2365	.2429	2405	.2423	.2436

Definition	J	K	L	M	N	O	Correlation with Survey Response
1							.4449
2							.4889
3	.7701***	.6120***	.5841***				.4911
4	-.0234			.4418*	.4643***		.4780
5		.1406		.2851		.4802	.4751
6			1737		.2683	.2388	.4715
R^2	.2412	.2421	2431	.2305	.2316	.2268	

* $p < .05$
** $p < .01$
*** $p < .001$

N = 1021

Table 5
Self-reports versus Administrative Data as to Whether a Case Began Before January 1987

6 Month Administrative Definition

		Administrative data		
		Before Jan. 1987	**After Jan. 1987**	**Total**
Self-report	**Before Jan. 1987**	368 (27.2%) (51.9%)	983 (72.8%) (96.3%)	1351 (78.1%)
	After Jan. 1987	341 (90.0%) (48.1%)	38 (10.0%) (3.7%)	379 (21.9%)
		709 (41.0)	1021 (59.0%)	1730 (100.0%)

The results are only marginally different from the prior set of regressions. The three-month definition is still better than the other definitions at predicting the survey measure and has the highest correlation with the survey measure (0.6237). The magnitude of many of the coefficients and of all of the correlations is greater, as is their significance, but this is likely due to a better fit created by removing some of the outliers. The one-month definition appears to benefit most from this selection, but still is a far worse predictor than the other definitions.

Substantively, the results seem to point toward respondent bias. The additive value of the coefficients for all of the regressions is never much higher than 0.75. This indicates that spells longer than one year may be telescoped, if we are to assume a linear relationship between the administrative data and the survey data. But there is also another possibility because almost none of the coefficients ever go to zero as we would expect from the model described above. Although the preceding results are qualitatively similar to what we predicted from our model, they do not yield as many zero regression coefficients as we expected. For example, equation F in Table 6 has two roughly equal and statistically significant coefficients for two month and three month definitions. This suggests that no one administrative measure is the correct one. Maybe a two month interruption is the proper definition for some people while a three month interruption is the proper definition for others. This certainly seems possible on a priori grounds. For some people a gap of two months may be a true interruption, but for others a gap of three months might be required.

C. A Mixing Model

In this section we investigate a "mixing" model in which one definition of a welfare spell is correct for a certain fraction of the population and another definition is correct for the remaining fraction. For simplicity and concreteness, we will assume that there are only two correct definitions of spell lengths, either one based upon a two month interruption or one based upon a three month interruption. The number of months is not essential to our derivation, but the number of distinct possibilities (two in this case) is.

We assume that there is a Bernoulli distributed random variable D indicating whether a person's true spell length is based upon a two month interruption (in which case D is one) or upon a three month interruption (in which case D is zero). Then we can write the observed spell lengths based upon the two different definitions as follows:

$$
\begin{aligned}
X_2 &= D\,X + (1-D)\,(X - \eta) \\
X_3 &= (1-D)\,X + D\,(X + \eta)
\end{aligned}
\qquad (13)
$$

where η is a non-negative random variable. On the one hand, if the person's true spell length is defined by a two month interruption, then D is one and X_2 equals X, the true spell length. In this case, X_3 equals $(X + \eta)$ or the true spell length plus some non-negative quantity because the three month definition has to be equal to or longer than the two month definition. On the other hand, if the person's true spell length is defined by a three month interruption, then D is zero and X_3 equals X and X_2 equals $(X - \eta)$ or the true spell length minus some non-negative quantity because the two month definition has to be less than or equal to the three month definition. Note that the expectation or mean of D will be the fraction of the population whose true welfare spell length is based upon a two month interruption. We will denote this fraction by π.

So far our discussion of equation (13) is just a matter of notation to get the bookkeeping right for X_2 and X_3. We now assume two things. As before, we posit that η is independent of X.[17] We further assume that η is independent of the person's type as

[17] Actually this is a stronger assumption than before where we only assumed that X was uncorrelated with η.

Table 6

Trivariate Regressions of Self-Reported Spell Length on Administrative Definitions Outliers Excluded

Definition	A	B	C	D	E	F	G	H	I
1	.1841**	.2148***	.3080***	.3279***	.3456***				
2	.6005***					.3454*	.5806***	.5207***	.5068***
3		.5804***				.4285**			
4			.4824***				.1905		
5				.4722***				.2592**	
6					.4628***				.2812***
R^2	.3894	.3939	.3848	.3896	.3922	.3899	.3869	.3910	.3943

Definition	J	K	L	M	N	O	Correlation with Survey Response
1							.5829
2							.6202
3	.7702***	.5825***	.5471***				.6217
4	-.0032			.3870**	.4102***		.6057
5		.1922*		.3617**		.4274**	.6051
6			.2346		.3463***	.3169*	.6031
R^2	.3865	.3891	.3918	.3718	.3749	.3571	

* $p < .05$
** $p < .01$
*** $p < .001$

N= 1006

indicated by D. This amounts to assuming that the average overestimate of spell length resulting from using X_3 instead of X_2 for those whose spells are truly defined by two month interruptions is equal to the average underestimate of spell length resulting from using X_2 instead of X_3 for those whose spells are truly defined by three month interruptions. This assumption turns out to be very convenient, and it seems like a good starting point. Note that in any interesting case, η must have a positive mean which we will denote by η^*. (Without loss of generality we assume that X has a zero mean. One way to think about this is to simply subtract the non-zero mean of X from both X_2 and X_3 and to note that this will not change the coefficients for these two variables in a linear regression equation. It will, of course, change the intercept but we are not interested in the intercept in these derivations.)

Our goal, as before, is to see what happens when we estimate a trivariate regression equation using OLS. This requires obtaining the expectations and second central moments (variances and covariances) for all of the observed variables. The results are as follows:[18]

$$
\begin{aligned}
\mathrm{Var}(X_2) &= \mathrm{Var}(X) + (1-\pi)[\mathrm{Var}(\eta) + \pi\,\eta^{*2}] \\
\mathrm{Var}(X_3) &= \mathrm{Var}(X) + \pi\,[\mathrm{Var}(\eta) + (1-\pi)\,\eta^{*2}] \\
\mathrm{Cov}(X_2,Y) &= \mathrm{Cov}[X + (1\text{-}D)\eta, Y] = \mathrm{Cov}(X,Y) \\
\mathrm{Cov}(X_3,Y) &= \mathrm{Cov}(X + D\delta, Y) = \mathrm{Cov}(X,Y) \\
\mathrm{Cov}(X_2,X_3) &= \mathrm{Var}(X) + \pi\,(1-\pi)\,\eta^{*2}.
\end{aligned}
\tag{14}
$$

Remembering that $E(\eta^2) = \mathrm{Var}(\eta) + \eta^{*2}$ and denoting $\mathrm{Cov}(X,Y)/\mathrm{Var}(X)$ by β, we can write the estimated regression coefficients as follows after substituting (14) into (5) and performing some manipulations:

$$
\begin{aligned}
\beta_2^* &= \pi\,\beta\,\,[1 + \pi\,(1-\pi)\,E(\eta^2)/\mathrm{Var}(X)]^{-1} \\
\beta_3^* &= (1-\pi)\,\beta\,\,[1 + \pi\,(1-\pi)\,E(\eta^2)/\mathrm{Var}(X)]^{-1}.
\end{aligned}
\tag{15}
$$

The quantity in brackets is the same in both cases, and it will always be greater than or equal to one. The quantity $E(\eta^2)/\mathrm{Var}(X)$ is a nuisance parameter which we can denote by γ. Then the coefficients are as follows:

---

[18] To summarize our assumptions: these results assume that D is independent of X, η, and ϵ from the regression of Y on X_2 and X_3, that E(X) is zero, that $E(\eta)$ is η^*, and that η is independent of X and ϵ.

$$\text{(16)}\qquad \begin{aligned} \beta_2^* &= \pi\,\beta\ [1 + \pi\,(1-\pi)\,\gamma]^{-1} \\ \beta_3^* &= (1-\pi)\,\beta\ [1 + \pi\,(1-\pi)\,\gamma]^{-1}. \end{aligned}$$

There are then three unknowns (β,γ, and π) and two equations.

One way to think about these coefficients is that they are comprised of three parts. First there is the coefficient for telescoping or social desirability (β). If this is one, then there is no telescoping or social desirability (or possibly the two cancel out). Second, there is a parameter (π or [1-π])---equal to the fraction of the population for whom a particular administrative measure is the correct one---that indicates how much that measure contributes to the survey measure. If π is not zero or one, then both coefficients (β_2* and β_3*) will be non-zero. The relative sizes of the coefficients will depend solely upon the size of size of π. The regression coefficient for the spell length definition that is correct for the larger fraction of the respondents will be the larger one, and the regression coefficients will be equal if there are equal fractions of people in each group (if π is one-half). Finally, there is an attenuation factor $[1 + \pi\,(1-\pi)\,\gamma]^{-1}$ that arises from the econometric fact that when π is neither zero nor one, each of the administrative definitions is an error laden measure of the true definition of spell length. Hence, this error attenuates the coefficients because the factor $[1 + \pi\,(1-\pi)\,\gamma]^{-1}$ will always be less than one when π is neither zero nor one.

The sum of the coefficients will be:

$$\text{(17)}\quad S = \beta_2^* + \beta_3^* = \beta\ [1 + \pi\,(1-\pi)\,\gamma]^{-1},$$

and this will always be less than or equal to β. Hence, even if β is truly one and there is no telescoping or social desirability bias, the sum of the coefficients in a trivariate regression can be less than one in a mixing model because of the attenuation factor. This suggests that there are two different ways that the sum of the coefficients can be less than one. One way is that there may telescoping or social desirability bias in which case β is less than one. Another way is that more than one measure might be the correct one. This amounts to a very fundamental identification problem stemming from having three unknowns (β,γ, and π) in (16) but only two pieces of information (β_2* and β_3*).

Somewhat surprisingly, however, although we cannot estimate β without knowing γ, we can estimate π. Consider the ratio of β_2* over $\beta_2^* + \beta_3^*$ and the ratio of β_3* over $\beta_2^* + \beta_3^*$:

$$\text{(18)}\qquad \begin{aligned} \beta_2^*/[\beta_2^* + \beta_3^*] &= \pi\,\beta\ [1 + \pi\,(1-\pi)\,\gamma]^{-1} / \{\beta[1 + \pi\,(1-\pi)\,\gamma]^{-1}\} = \pi \\ \beta_3^*/[\beta_2^* + \beta_3^*] &= (1-\pi)\,\beta\ [1 + \pi\,(1-\pi)\,\gamma]^{-1} / \{\beta[1 + \pi\,(1-\pi)\,\gamma]^{-1}\} = 1-\pi. \end{aligned}$$

Thus we can recover the fraction of people for whom a two month interruption (or a three month interruption) is the proper definition of the end of a spell.

D. Final Empirical Analysis

The mixing model helps us get a better grip on the findings reported earlier. It provides us with an explanation of why we obtained so few zero coefficients in the trivariate regressions reported earlier. It seems likely that a mixing model fits these data better than our first model. The only case where this may not be true is in regression J, where the three-month and four-month definitions are the independent variables. In this case, we can conclude that the three month definition is always better than the four month definition. Furthermore, the mixing model suggests that having the regression coefficients always sum to less than .77 does not necessarily imply telescoping or social desirability biases. Instead, the attenuation factor arising in the mixing model may be causing the values of β_2* and β_3* to be diminished in magnitude.

As we have mentioned before, the three-month and the two-month definitions provide the best fit to the survey data. We also believe that for this mixing model these are the definitions we would want to use. Substituting the estimated coefficients into (18), we find that

$$\pi = .3006/(.3006 + .4526) = .3991$$

and

$$1- \pi = .4526/(.3006 + .4526) = .6009$$

Therefore, about 40 percent of the survey sample fits better with the two month definition and about 60 percent of the sample fits better with the three month definition.

It is possible that we need to use more than two administrative definitions to correctly model the data. To explore this, we ran another series of regressions, first using all six of the administrative definitions, and then excluding the longer definitions, which we seem very unlikely to us on a priori grounds (See Table 7). Even when all six definitions are used, the three-month definition works surprisingly well, despite the amount of multicollinearity among the independent variables. The two-month definition is also quite substantial, although the one month definition adds little to the fit. One unexpected result is that the six-month definition also appears to be fairly large. This may be due to an effect of question wording. The portion of respondents who answered that they had been using AFDC since they first started were never given the opportunity to mention breaks in their spells. In these cases, even longer breaks may be forgotten. Nevertheless, at the moment we rely primarily upon our strong prior beliefs as a basis for disregarding this anomalous finding.

If we expand the mixing model to allow for three independent variables, it seems likely

Table 7

Regressions of Self-Reported Spell Length on Administrative Definitions Based on 1 to 6 Month Interruptions (Standard Errors in Parentheses)

Definition	P	Q	R	S
1	.1090	.1093	.1061	.1064
	(.0868)	(.0868)	(.0868)	(.0867)
2	.2069	.1954	.2032	.2027
	(.2145)	(.2143)	(.2145)	(.2027)
3	.4636*	.4762*	.4749*	.4572**
	(.2685)	(.2684)	(.2686)	(.1992)
4	-.2831	-.2862	-.0181	
	(.2493)	(.2493)	(.1846)	
5	.0341	.2787		
	(.2589)	(.1743)		
6	.2488			
	(.1948)			
R^2	.2471	.2459	.2440	.2440

* $p < .1$
** $p < .05$

N=1021

(although we have not yet proved it) that we can derive the fraction of the population for each of the three administrative definitions is correct by extending the results in (18) in an obvious fashion. If we define π as the probability a case will use the two-month definition, ρ as the probability a case will use the three-month definition, and $1-\pi-\rho$ that a case will use the one-month definition, then we hypothesize that we estimate the fractions as follows:

$$1-\pi-\rho = .1064/(.1064 + .2027 + .4572) = .1388,$$
$$\pi = .2027/(.1064 + .2027 + .4572) = .2645,$$
and
$$\rho = .4572/(.1064 + .2027 + .4572) = .5867 .$$

From this, we can deduce that the one-month definition is applicable to less than 15 percent of the cases, while the three-month definition is still applicable to nearly 60 percent of the cases.

VI. Conclusions and Future Work

We started with two concerns: how do we define a spell of welfare and how well do surveys measure welfare spells. We do not have complete answers for these questions, but we now know more than we did before. We know that administrative churning seems to occur and that a break of up to three months is probably necessary in a large fraction of welfare cases before we can say that they have really gone off welfare. Indeed, a break of only one month may define a true break in less than fifteen percent of the cases. We also know that it is very hard to detect social desirability or telescoping effects unless we can determine the proper administrative definition of a welfare spell. We have shown that there is a fundamental identification problem when we use survey data to cross-check administrative data. Finally, the relatively poor fits from regression survey data on administrative data suggests that there is a lot of error in the survey data.

There is a substantial agenda for future work. We would like to analyze what happens if we make different assumptions about how error enters. For example, what can we conclude if error enters multiplicatively as in equation (4)? We would also like to deal with the problem of truncation due to the fact that Y and X cannot be less than zero. And we would like to find an explanation for the anomalous empirical results for the administrative measure based upon a six month interruption. Finally, and possibly most importantly, we would like to get more detailed data about what is happening to cases where there is a one, two, or three month break. Do we find that there is any evidence for administrative churning such as a month of no payment followed by a month of double payment? Our preliminary look at our data suggests that this might be happening.

References

Ayasse, Susan. 1995. "California Welfare Recipients and Spells on Aid -- Survival Analysis. Paper presented at the Population Association of America Meetings, San Francisco, April.

Aquilino, W.S. 1994. "Interview Mode Effects in Surveys of Drug and Alcohol Use: A Field Experiment." *Public Opinion Quarterly* 58: 210-240.

Bane, Mary Jo and David Ellwood. 1983. The Dynamics of Dependence: The Routes to Self-Sufficiency." Report prepared for the U.S. Department of Health and Human Services by Urban Systems Research and Engineering, Inc., Harvard University.

Blank, Rebecca. 1989. "Analyzing the Length of Welfare Spells." *Journal of Public Economics* 39: 245-273.

Bradburn, Norman, Lance J. Rips, and Steven K. Shevell. 1987. "Answering Autobiographical Questions: The Impact of Memory and inference in Surveys." *Science* 236: 157-161.

Brown, N.R., L.J. Rips, and S.K. Shevell. 1985. *Cognitive Psychology* 17: .

Cahalan, Don. 1968. "Correlates of Respondent Accuracy in the Denver Validity Survey." *Public Opinion Quarterly* 32: 607-21.

Ellwood, David. 1986. Targeting 'Would Be' Long-Term Recipients of AFDC. Report prepared for U.S. Department of Health and Human Services by by Mathematica Policy Research, Inc., Princeton, N.J.

Gilens, Martin Isaac. 1991. Racial Attitudes and Opposition to the American Welfare State. Dissertation, University of California at Berkeley.

Hadaway, C. Kirk, Penny Long Marler, Mark Chaves. 1993. "What the Polls Don't Show: A Closer Look at U.S. Church Attendance." *American Sociological Review* 58: 741-752.

Hardle, Wolfgang. 1989. *Applied Nonparametric Regression.* Cambridge: Cambridge University Press.

Harris, Kathleen Mullan. 1993. "Work and Welfare Among Single Mothers in Poverty." *American Journal of Sociology* 99: 317-352.

Hill, Kim Quaile and Patricia A. Hurley. 1984. "Nonvoters in Voters' Clothing: The Impact of Voting Behavior Misreporting on Voting Behavior Research." *Social Science Quarterly* 65: 199-206.

Katz, Michael B. 1986. *In the Shadow of the Poorhouse: A Social History of Welfare in America*. New York: Basic Books.

Plotnick, Robert. 1983. "Turnover in the AFDC Population: An Event History Analysis." *Journal of Human Resources* 18: 65-81.

Roos, John and Ken Grant. 1985. "Measuring Participation Using Public Opinion Surveys: Who Lies and Why?" *Political Methodology* 11: 293-298.

Smolensky, Eugene, Eirik Evenhouse, and Siobhan Reilly. 1992. "Welfare Reform in California." Working Paper, Graduate School of Public Policy, UC Berkeley.

Trattner, Walter I. 1974. *From Poor Law to Welfare State: A History of Social Welfare in America.* New York: The Free Press.

Traugott, Michael W. and John P. Katosh. 1979. "Response Validity in Surveys of Voting Behavior." *Public Opinion Quarterly* 43: 359-377.

Volgy, Thomas J. and John E. Schwarz. 1984. "Misreporting and Vicarious Political Participation at the Local Level." *Public Opinion Quarterly* 48: 757-765.

Wagenaar, Willem Albert. 1986. Cognitive Psychology 18:

Weiss, Carol H. 1968-1969. "Validity of Welfare Mothers' Interview Responses." *Public Opinion Quarterly* 32: 622-633.

Williams, B.L. and H. Suen. 1994. "A Methodological Comparison of Survey Techniques in Obtaining Self-Reports of Condom-Related Behaviors." *Psychological Reports*. 75: 1531-1537.