

Sheldon Haber, George Washington University
The George Washington University, Department of Economics, Washington D.C. 20052

KEY WORDS: Employer health insurance, cost survey; survey design

In August 1987 the U.S. Bureau of the Census conducted a special survey to determine the feasibility of collecting information from employers about their contributions for health insurance, life insurance, and private pension plans on behalf of specified individuals in their employ. This special survey, known as the SIPP Employer Fringe Benefits Survey (which for simplicity is referred to below as "the survey"), was conducted at the end of the last wave, i.e., the 8th wave, of the 1985 panel of the Survey of Income and Program Participation (SIPP). An objective of the survey, besides the primary one of determining the response rate, and the one we report on in this paper was to ascertain the characteristics of individuals who were least likely to participate in the field test. Such information is useful in determining the extent of selectivity bias present in the survey and in deciding how much time to allocate to different individuals so as to increase the response rate.¹

Since household survey respondents typically are unable to provide information about employer fringe benefit contributions, it was necessary to obtain a signed waiver from respondents permitting the Census Bureau to obtain the desired information from their employers. Thus the overall survey response rate depends on the response rate of respondents in signing the waiver and the response rate of employers in returning the survey questionnaire to the local Census Bureau field office.

While the survey was primarily undertaken to determine whether response rates would be sufficiently high to warrant the collection of fringe benefit data from a larger sample, the goal of this paper is to gain some insight into the characteristics of workers who are least likely to sign a waiver.

The major conclusion of the study is that there was only a weak relationship between the propensity to sign the waiver and variables related directly or indirectly to socio-economic status. Other SIPP data not contained in the fringe benefit survey on the other hand indicate that there is a strong positive relationship between socio-economic status, e.g., as measured by education, and the decision to refer to W-2 forms in providing the Census Bureau with accurate earnings information. This latter finding, that the better educated are more likely to refer to their W-2 forms, is somewhat surprising in light of the reluctance of better educated, and presumably wealthier, individuals to provide wealth data.²

1. Survey Response Rates

The survey consisted of a random sample of one-half of the respondents in Wave 8 of the 1985 SIPP panel. Only employed wage and salary workers age 18 and older who completed an interview questionnaire were included in the test; 1,352 respondents met this criterion.

Employers were asked if they provided medical insurance, life insurance, or pension plan benefits and whether particular employees identified by name and social security number who were still in their employ at the time they received the questionnaire were eligible for benefits as of July 1, 1987, and if eligible, whether they were covered under a provided plan as of that date. For employees covered by a fringe benefit plan, employers were asked the amount they, the employers, paid per week, month, year, or some other period. For employers who did not return the questionnaire there was a telephone follow-up.

Respondents were asked to sign the waiver whether or not they were covered by an employer provided benefit. Signed waivers were collected by the local field office and sent to employers at the address provided by the respondent. This procedure was followed by respondents who were self-interviewed and those who were mailed or left a form to sign because they could not be interviewed in person; the latter are referred to below as proxy interviewees. No follow-up was made of respondents for whom a signed waiver was not obtained.

Of the 1,352 respondents in the test, 569 or 42 percent signed the waiver. Of the latter, 503 or 88 percent were returned by employers after the first mailing. Another 45 questionnaires, i.e., an additional 9 percent, were received after the follow-up. In total, the employer response rate was thus 96 percent. Taking account of the questionnaires that were not returned by employers, returns were received for 41 percent of the individuals participating in the survey.³

It is to be noted that the computer tape file upon which our study is based contained information for a smaller number of respondents, 1,308, than the number reported by the field offices. Likewise the number of returned questionnaires in the file was found to be smaller, 527. The overall response rate however, 40 percent, was almost the same as that noted above.⁴ As might be expected the response rate for self-interviewees, 57 percent, was higher than that for proxy interviewees, 19 percent.⁵ Presumably the response rate for proxy interviewees would have been higher had a follow-up procedure been used to remind respondents to return their signed waivers if they had not already done so.

2. Survey Employer Cost Data

Preliminary examination of the questionnaires that were returned indicated that the employer cost data were most complete for the health insurance benefit and least complete for the pension benefit. At this point a decision was made to limit further analysis of the data to the health insurance benefit. The rationale for this decision was that the Census Bureau already collects information on private pension benefits in SIPP on an as-received basis, so that employer pension plan contributions, while of

interest for studying worker compensation (as a complement to worker income), are of lesser importance than information about employer health insurance contributions. Because employer contributions for life insurance are relatively small (being one-tenth of the contributions for health insurance) it was decided that further analysis of these data should likewise be abandoned.

In order to increase the number of responses with usable health insurance cost data,⁶ another telephone call follow-up was conducted. Employers were called in rough order of the difficulty of supplying missing information about their health insurance costs. These employers were also asked about their life insurance and pension costs, since once an employer was contacted the marginal cost of collecting the additional information was low. The follow-up was discontinued when the rate at which new information was being collected declined to a level where it was deemed that further calls would only marginally improve the data. Counts made of the number of questionnaires with usable employer cost information before and after the follow-up are shown in Table 1.

As can be seen from the first column of Table 1, after the follow-up the sample contained 339 records with usable employer health insurance cost data.⁷ These usable records comprised 64 percent of the 527 employer questionnaires contained in our data set. It is to be noted however that another 26 percent of the respondents were not covered by a health insurance plan either because their employer had no plan or if there was one the individual chose not to enroll in the plan. The remaining 10 percent of the cases could not be resolved because, e.g., no call back was attempted or the firm could not provide the cost data requested.

The number of usable employer health insurance responses was increased by 14 percent as a result of the follow-up procedure. Most of the initially unusable employer cost data are attributable to the way the questionnaire was worded. As indicated only three periods — weekly, monthly, and yearly — were explicitly provided for. It thus became necessary to obtain the payment period where the periodicity of the payment was other than those enumerated, e.g., bi-weekly.⁸

Another difficulty with the questionnaire leading to the low number of usable pension benefit records is that in numerous cases employers indicated on the form the percentage

of a worker's pay that was contributed rather than the amount contributed. This information was not keyed in when the file was created; had this been done the amount contributed could have been estimated utilizing earnings data from SIPP.

One other aspect of the quality of the data warrants mention and that pertains to the consistency of the employer and respondent answers regarding health insurance coverage. Employers reported that 87 percent of the workers for whom they received questionnaires were covered by their health plan. On the other hand 64 percent of the survey respondents indicated (in the SIPP survey itself) that they were covered under an employer provided plan. The higher figure for employers suggests that persons who signed the waiver were more likely to respond that they were covered by an employer's policy than those who did not. Few disagreements however were found between the employer and respondent answers as to whether an individual was covered by health insurance, given that an employer filled out the coverage questions of the health insurance portion of the questionnaire. In the latter instance the proportion of observations with disagreements was 7 percent.

3. The Empirical Model

As indicated above the primary focus of this paper is the question of whether some workers are less likely than others to sign a waiver that would allow the collection of information that can only be obtained from their employer. Since employers returned 96 percent of the questionnaires, the receipt of a questionnaire by a field office can be used to identify individuals who signed the waiver. In so doing it is assumed that no selectivity bias was exhibited by employers with respect to the questionnaires they returned.

Logistic regression was employed to determine whether persons who signed the waiver were different from those who did not. The empirical model estimated was

$$\log p/(1-p) = a + \sum B_i X_i$$

where p is the probability of a person signing the waiver, $p/(1-p)$ is the odds of signing the waiver, the coefficients B_i show the log of the odds for a unit change in X_i , and the X_i represent variables associated with each respondent. For the logistic regression model $\log p/(1-p)$ varies linearly with the X_i . It is

Table 1

	Number of Questionnaires with Usable Employer Cost Data		
	Health Insurance	Life Insurance	Pension Plan
Before follow-up	297	195	107
After follow-up on health insurance costs	339	206	118

also seen that as $\log p/(1-p)$ increases so does p .

Personal, geographic, and job related variables are used to define a respondent's characteristics. In this paper the first group of variables includes the sex, age, race, marital status, and education of the individual. The second group includes the individual's region of residence and whether he or she resides in a metropolitan area. The last group includes the individual's hourly wage, whether he or she is a full-time worker, a member of a union, employed in a white-collar occupation, and employed in a service industry.

A positive B_i , i.e., log of the odds, indicates that the probability of signing the waiver increases as the value of X_i increases; conversely a negative B_i indicates that as X_i increases in value the likelihood of signing the

waiver diminishes.⁹ For the problem at hand the expected signs of the coefficients cannot be ascertained a priori. For example, highly educated persons may have signed the waiver because they felt it was their civic duty to do so. On the other hand, if they were more likely than less educated persons to be covered by a health plan they may have declined signing the waiver because they felt that the disclosure of their employer's contribution to that plan might rebound to their disadvantage at a future date. Thus the sign of the coefficient of the education variable is uncertain. The same ambiguous outcome holds for each of the other variables.

4. Empirical Results: The Propensity To Sign The Waiver
The B_i for three versions of the model are

Table 2
Logistic Regression: Characteristics of Individuals who Signed the Survey Waiver and Who Provided an Employer Identification Number (EIN)

Variables ^a	Signed the Waiver			Provided an EIN
	All Respondents	Self-Interviewees	Proxy Interviewees	
Intercept	-2.833*	-.540	-3.949	-3.927*
Self-interview	1.815*	—	—	—
Female	-.145	-.078	-.152	.209
Age	.002	-.012***	.028*	.002
White	.855*	.974*	.689***	.946*
Married	.113	.004	.228	.540*
Region	-.054	-.029	-.113	-.023
Metro	-.212	-.151	-.385	-.352*
Full-time	.137	.149	.197	.468**
Education	.032	.031	.040	.107*
Hourly Wage	.031**	.035***	.030	.032**
Union	.012	-.042	.117	.204
White-collar	.189	.336***	-.113	-.050
Service Ind	-.210	-.379***	.115	-.199
Health Ins	-.608*	-.812*	-.045	—
2 log likelihood	1,429.8	908.8	496.6	1,459.6
Chi-square	277.8	87.4	27.2	97.2

* Significant at .01 level.
** Significant at .05 level.
*** Significant at .10 level.

Sample sizes: 1,284 (survey, all respondents); 720 (survey, self-interviewees); 564 (survey, proxy-interviewees); 1,284 (survey, all respondents).

^a Self-interview = 1 for self-interviewee, Female = 1 for women, White = 1 for white, Married = 1 for married, Region = 1 for South, Metro = 1 for metropolitan area, Full-time = 1 for full-time worker, Union = 1 for union member, White-collar = 1 for white-collar worker, Service Ind = 1 for service industry worker, and Health Ins = 1 for persons reporting they are covered by an employer health plan. Age, Education, and Hourly Wage are continuous variables.

shown in the first three columns of Table 2. The first is for all respondents, the second and third for self-interviewees and proxy respondents, respectively. In each instance the dependent variable equals 1 if the respondent signed the waiver and 0 if not. All of the independent variables mentioned above are included in the models. Additionally two other independent variables are delineated. One, denoted as Self-interview, identifies individuals who were self-interviewees and is entered only in the first regression. The expected sign of this variable will be positive if signing the waiver depends on how much information one has regarding the purpose of the survey. The other, denoted as Health Ins, identifies individuals who reported that they participated in an employer provided health plan (as distinct from their employer reporting that they were covered in such a plan) and is entered in all three regressions. The expected sign of this variable will be negative if persons who had no employer provided health plan declined to sign the waiver because they thought that to do so would serve no useful purpose.

From the first regression in Table 2 it is noticed that being a proxy interviewee decreases the log of the odds and hence the probability of signing the waiver, everything else the same. The same is true for persons who report they had no employer health insurance, i.e., they too, were less likely to sign the waiver. It is also noticed from the second regression that for self-interviewees there is only a moderate relationship between signing of the waiver and the independent variables. In this regression four of the five significant demographic/economic variables are significant at the .10 level. Not surprisingly, from the third regression for proxy interviewees it is found that the event of signing the waiver is subject to greater randomness than a similar event being observed for a self-interviewee. The only significant variable with the same sign common to the second and third regressions is White, indicating that minority members are less likely than others to sign the waiver.

5. Empirical Results: The Propensity To Provide An Employer Identification Number (EIN)

The above results indicate that the propensity to release information via an employer is only weakly related to socio-economic class. To check on the generality of this finding we looked at the characteristics of respondents who referred to their W-2 forms, to provide information about their earnings and also their employer's identification number, i.e., EIN. Indeed the presence of an EIN number is the best way to determine whether individuals looked at their W-2 form, since in answering questions about annual earnings and taxes paid, respondents were asked to refer to a W-2 form or a completed worksheet; for this reason it is not possible to ascertain those instances when the W-2 form was utilized without reference to the EIN. Besides being of interest in and of itself, the relationship between the use of the W-2 form and socio-economic class also provides

a baseline for assessing the strength of the similar relationship between the signing of the waiver and socio-economic class. To assess the former relationship we again use logistic regression.

In this case the dependent variable equals 1 if the respondent permitted the interviewer to transcribe the EIN and zero if not. The independent variables are the same as before except that the variables Self-Interview and Health Ins are excluded. As can be seen from the last column of Table 2 there is a strong relationship between the availability of an EIN and the socio-economic characteristics of a respondent.

In contrast to the first regression where only two demographic/economic variables are found to be significant at the .05 level, six are significant at this level in the last regression. Provision of an EIN is related to race, marital status, educational level, full-time work status, a person's hourly wage, and place of residence. Those individuals who are minority members, single, less educated, work part-time, earn a low wage, and live in a metropolitan area, are least likely to refer to their W-2 form. These characteristics, except for the last one, are directly or indirectly related to socio-economic status. The results regarding the provision of an EIN are contrary to what one would expect if those at the top of the socio-economic pyramid perceive they have the most to lose by reporting their correct earnings.

6. Conclusion

With respect to the limited focus of this paper, it appears that it is feasible to collect employer fringe benefit cost data, particularly data relating to health insurance costs, but the initial survey procedures and questionnaire need to be improved. This conclusion however is subject to further analysis of the cost data themselves to determine their reasonableness.

From a broader perspective the findings of this study suggest that the refusal by respondents in household surveys to provide information may be specific to survey contexts and questions, and that additional study of this aspect of survey design and analysis may be warranted.

FOOTNOTES

1. The question of how to improve survey response rates is part of a wide set of issues, including differences in response rates to particular kinds of questions, e.g., questions relating to income and wealth, and procedures for imputing missing information, that are met in designing and analyzing survey data. See for example the three volume study edited by Madow, Nisselson, and Olkin (1983) and the extensive bibliography in Kalton (1983). Little attention however has been given to the characteristics of persons who refuse to participate in a survey or to provide responses to particular questions, particularly when they are beyond

the scope of the survey in which they are participating.

2. Lamas and McNeil (1985) report that response rates to questions about asset ownership are lowest for college graduates and older persons. They also found no difference in the response rate between blacks and whites. In their paper however response rates are based on "don't know" answers as well as refusals to answer any question about asset ownership.
3. Based on figures cited in Chester E. Bowie, Final Progress Report on the SIPP Employer-Provided Benefits Study, memorandum dated November 12, 1987.
4. Unless otherwise indicated the data utilized in the remainder of this study are from the computer tape file constructed for this project by the Bureau of the Census.
5. Self-interviewees comprised 56 percent of the sample, the remaining 44 percent were proxy interviewees.
6. In the context of this study usable data are data that could be utilized in predicting employer costs for individuals not included in the survey sample and/or in studying why employer contributions vary.
7. A slightly larger number, 345, is reported in Judy Eargle, Employer Survey Form Progress, memorandum dated March 14, 1988.
8. In retrospect this problem could have been mitigated if yearly costs were asked for even

though this would have required some additional effort on the part of employers.

9. To transform a coefficient B_i to show the odds for a specified value of X_i , rather than the log of the odds, the antilog of the right-hand side of the logistic regression equation is calculated setting the variables other than the given one to their mean. The given X_i is set equal to 1 or 0 if it is a dummy variable and to a specified value if it is a continuous variable. If the odds for a given variable, say, region (e.g., where a value of 1 for this variable indicates South) is calculated to be 3.00, i.e., 3.00:1, this would mean that a person residing in the South would sign the waiver 3 out of 4 times. The probability of this event occurring is given by the quotient of the odds to the odd+1. In this case the probability of signing the waiver for persons living in the South, all else the same, would equal 0.75.

REFERENCES

- Kalton, Graham, Compensating for Missing Survey Data, The University of Michigan, Survey Research Center, 1983.
- Lamas, Enrique and John McNeil, The Measurement of Household Wealth in the Survey of Income and Program Participation, American Statistical Association, 1984 Proceedings, Social Statistics Section, 1985, pp. 484-489.
- Madow, William, Harold Nisselson and Ingram Olkin, Incomplete Data in Sample Surveys, Academic Press, 1983.