Measurement of Volunteering: A Methodological Study Using Indiana as a Test Case

Kathryn S. Steinberg, Patrick M. Rooney and William Chin

Nonprofit and Voluntary Sector Quarterly 2002; 31; 484
DOI: 10.1177/0899764002238097

Measurement of Volunteering: A Methodological Study Using Indiana as a Test Case

Kathryn S. Steinberg
Patrick M. Rooney
Center on Philanthropy at Indiana University
William Chin
Indiana University–Purdue University at Indianapolis

How much does survey methodology matter when measuring volunteering? Every 4 years, the Center on Philanthropy at Indiana University conducts a telephone survey (called "Indiana Gives") of the giving and volunteering behaviors of Indiana citizens. In this most recent wave of the Indiana survey, conducted in October and November 2000, we asked eight groups of approximately 100 randomly selected Indiana residents to complete one of eight surveys related to giving and volunteering. We found that the longer and more detailed the module, the more likely an individual was to have provided volunteer service in the past year, and the higher the average number of hours provided, even after controlling for differences in age, income, household status, itemization status, and race. Further research is needed to ascertain whether there may be a "point of diminishing returns" in terms of optimal survey length and detail and whether recall is inhibited for high-end volunteers only.

How do you measure volunteering? It seems to be a straightforward question. But in the United States at least, the most frequently used methods of measuring giving and volunteering—most of which are based on surveys—produce surprisingly disparate results. Most researchers agree that more precise ways of estimating how much time and money people give are needed. This issue is important because many fund-raisers, nonprofit leaders, and public policy makers often use such estimates to understand how much and why people do volunteer work and how and why they donate to charity, as well as to see how

Note: The authors would like to thank Paul Schervish, who helped design the overall study, and Brian Vargus and the Public Opinion Laboratory at IUPUI for helping with the data collection. We also would like to acknowledge Richard Steinberg and Mark Wilhelm for their invaluable input. An earlier version of this article was presented at ARNOVA, and we would like to thank the session members for several good suggestions.

Nonprofit and Voluntary Sector Quarterly, vol. 31, no. 4, December 2002 484-501
DOI: 10.1177/0899764002238097

their organizations are doing compared with overall patterns in the country and to plan new initiatives.

Because estimates of the amount of philanthropic behavior appear to rely in such large part on the methods and measures of each survey, this has been a topic of enduring debate within the academic and practitioner communities. For example, in 1993, an entire issue of *Voluntas* (Vol. 4, Issue 2) was devoted to this topic, and in 2001, *Nonprofit and Voluntary Sector Quarterly* (Vol. 30, Issue 3) published a symposium of articles discussing methodology in surveying giving and volunteering behavior.

The Center on Philanthropy has set out to examine these issues on two fronts. Following discussions with the Association for Research on Nonprofit Organizations and Voluntary Action (ARNOVA), in November 2000, the center convened virtually all of the leading researchers across the country involved in estimating giving and volunteering for a conference to discuss the pros and cons of various survey methods, their results, and how to develop measures that potentially could improve research results in the field of philanthropy overall.

We also began to test different survey research methods, teaming up with Paul Schervish, 1999-2000 Distinguished Visiting Professor with the center and director of the Boston College Social Welfare Research Institute. The research team sought to assess whether and to what extent different survey methods—such as length, wording, order, and number of prompts—produce different findings. For instance, to what extent do the following factors forge estimates of individual philanthropy: the number of prompts by the method of volunteering (e.g., serving food, fund-raising, coaching, etc.); the number of prompts by the subsector for volunteering (e.g., education, environment, health, etc.); prompts for formal as well as informal volunteering (i.e., volunteering through an organization versus on one's own); prompts for informal volunteering with relatives versus nonrelatives; inducements to participate in the survey; and so forth. Or, would a simple and very short survey generate comparable results, which would produce considerable cost savings in studying the sector—if it yielded comparable results to those from the longer surveys?

Our research was significantly informed and enhanced by the seminal survey work done by our predecessors, such as Independent Sector (1999, 2001); Hall, McKeown, and Roberts (2001); and O'Neill and Roberts (2000), as well as by the innovative diary study conducted by Havens and Schervish (1997). As such, our current research, rather than implying a criticism of this previous work, is in fact a continuation and logical extension of many earlier efforts.

The goals of this project were both empirical and methodological: (a) to update measures of giving and volunteering in Indiana; (b) to compare systematically various methodological survey techniques for estimating giving and volunteering; (c) to assess whether differences in survey methodology—such as in the wording, order, and number of prompts, and in levels of respondent inducements—produce different findings; and (d) to discern whether we

can formulate some practical "rules of thumb" for making sense of different survey results and for improving future surveys.

METHOD

The measurement of volunteering is a separate methodological question from the measurement of giving. Some researchers (e.g., Independent Sector, 1999) have prompted volunteering by subsector; others (e.g., Hall, 2001; Hall et al., 2001) have prompted by method of volunteering—cleaning, serving food, reading, and so forth. The California survey (O'Neill & Roberts, 2000) queried volunteering by subsector, then by method. An additional question concerns whether measurement of volunteering should focus only on formal volunteering (through organizations) or include informal volunteering (on one's own). Most surveys have looked only at formal volunteering in detail, sometimes addressing informal volunteering with one general question. O'Neill (2001) has issued a call for more attention to the measurement of informal volunteering, as well as to the interactions between informal and formal volunteering.

To investigate these issues, we constructed a multipronged research design whereby we compared and contrasted the findings from different combinations of survey techniques related to giving and volunteering behaviors. The survey components related to giving have been described elsewhere (Rooney, Steinberg, & Schervish, 2001) and will not be covered in this article. Survey components related to volunteering behaviors were included in seven different survey modules, which are summarized in Table 1. Among these, we replicated central design elements of surveys done by Independent Sector (1999) (Area module) and Hall et al. (2001) (Method module). We used an expansion of the Hall methodology in three modules (IU-Method-Area, IU-Area-Method, and Volunteer Only). Another unit replicated an early draft of a survey module designed by the University of Michigan and the Center on Philanthropy for a longitudinal study, the Panel Study of Income Dynamics (PSID module). In addition, we asked a Very Short module (Did you volunteer last year? If so, how much?). For the two longest modules (IU-Method-Area and IU-Area-Method, which could be as long as 90 minutes each), the telephone researchers offered varying levels of inducements randomly ($0, $10, $25, and $50) to ascertain whether this affected the response rates or the levels of volunteering reported. Table 1 summarizes the modules by the types of volunteering prompts, the number of questions, and inducement levels.

To collect our data, the Indiana University Public Opinion Lab at Indiana University–Purdue University, Indianapolis (IUPUI) used random digit dialing of households to obtain samples of at least 100 respondents using the eight surveys to measure personal philanthropy in Indiana. Each respondent participated in only one of the modules. The actual number of subjects surveyed

Table 1. Modules by Types of Volunteering Prompts and Levels of Inducements

	Very Short Module	*IU-PSID Module*	*Area Module*	*Method Module*	*IU-Area-Method Module*	*IU-Method-Area Module*	*Volunteer Only Module*
Number of questions	16	137	337	368	676	680	223
Number of volunteering variables	4	9	170	204	215	215	215
Number of questions measuring volunteering	3	9	104	136	215	215	215
Type of volunteering prompts	General prompt (formal only)	General prompt (formal only), then by two subsectors (human services, community)	Prompts formal volunteering by subsector, general prompt for informal volunteering	Prompts formal and informal volunteering by method	Prompts formal and informal volunteering by method; queries informal separately for relatives and nonrelatives	Prompts formal and informal volunteering by method; queries informal separately for relatives and nonrelatives	Prompts formal and informal volunteering by method; queries informal separately for relatives and nonrelatives
Sample size	110	113	106	103	124	124	104
Inducements							
None	110	113	106	103	28	29	104
$10					29	30	
$25					33	30	
$50					34	34	

Note: IU = Indiana University; PSID = Panel Study of Income Dynamics.

varied somewhat from module to module, ranging from 101 to 124. Table 1 indicates sample sizes for each module.

It should be noted that in each module, we asked respondents about volunteering done by the respondent himself or herself. Previous studies have found that the reliability for reporting on someone else's volunteering ("proxy reporting") is quite low (Hall, 2001; Independent Sector, 1999). Therefore Independent Sector and most U.S. researchers have looked at individual volunteering as the unit of analysis when collecting data on volunteering, and we have followed that convention as well.

One difficulty that we encountered was that previous researchers have used different units in calculating volunteer service. For example, Independent Sector reports volunteer service in terms of hours per week and hours per year. However, the Independent Sector survey questions themselves are expressed in hours per week and hours per month. (The monthly data are used to calculate yearly figures.) The Canadian survey uses a combination of service hours per week, month, and year. Our longest surveys (Area-Method, Method-Area, and Volunteer Only) use a combination of questions about service hours per month and year, whereas the shortest modules (PSID and Very Short) ask about volunteer hours per year only. In calculating our main independent variable, we annualized all figures to create a common unit (total volunteering hours per year).

The main emphasis in this article is on measurement of total formal volunteering—that is, volunteering through organizations. We also measure differences across demographic groups: age, income, household status (couples vs. singles and/or marital status), race, and educational attainment. Three of our modules (Volunteer Only, IU-Area-Method, and IU-Method-Area) include questions that directly measure informal volunteering—that is, volunteering done on one's own, rather than through a particular organization or group. The Method module also includes questions on informal giving, but these questions only assess types of service and do not provide a measure of the number of hours volunteered. In addition, the Area module includes one general question on informal volunteering.

To test the data on formal volunteering, we first use ANOVAs to test whether the samples receiving each module have comparable demographic characteristics and reported volunteering. Then we perform several multivariate analyses (as suggested by O'Neill, 2001) to see whether differences in mean reported volunteering across modules can be explained by variations in sample characteristics or appear to be pure effects of the module administered. To do so, we explain volunteering hours in a regression framework by including a set of dummy variables for six modules, along with age, age squared, income, race (Whites vs. all others), marital status (married/cohabitating vs. single, widowed, and divorced), education (high school or less vs. all others), itemization status, and gender. If there are pure module effects, they will show up as significant coefficients for the module dummy variables.

Our initial regressions use the ordinary least squares (OLS) approach. However, OLS predicts a symmetrical distribution for the error term of the dependent variable. Previous research has shown that giving data has a nonnormal error structure because donations cannot be negative (Bradley, Holden, & McClelland, 1999; Rooney et al., 2001). We looked at the error structure of our data and found that our volunteering data followed a similar pattern—the OLS approach suffers from truncation bias because volunteering hours are never negative. Hence, we also performed tobit analyses, which do not generate negative predicted volunteer hours. Finally, we also performed Heckman two-stage analyses (including a probit analysis) to identify whether various independent variables can predict whether an individual volunteers at all, in addition to predicting the amount of volunteering. Because our data included outliers, we also performed each of these analyses with outliers excluded.

DATA AND BIVARIATE RESULTS

The demographic characteristics of each sample have been published elsewhere (Rooney et al., 2001), except for the sample for the Volunteer-Only module. Table 2 presents demographics for the Volunteer-Only module and a comparison to the total project sample.

In Table 3, we compare the simple means and medians of formal volunteering for each of the modules, and we see some interesting similarities and differences. (Note that Table 3 represents the entire sample, including outliers. We discuss results excluding outliers later in this article.) As shown in Table 3, the Very Short, IU-PSID, and Method module means were significantly lower, and the IU-Area-Method and Volunteer Only module means were significantly higher than the total sample mean. The IU-PSID survey shows a mean of 35.7 volunteer hours per year and a median value of 0 hours. The means for the others range from 50.7 (Very Short) to 462.84 hours (Volunteer Only), and the medians range from 0 (Very Short, Area, and Method modules) to 78.5 hours (Volunteer-Only module). It is worth noting that the highest mean values come from the IU-Area-Method, IU-Method-Area, and Volunteer-Only modules, which include the highest number of questions measuring volunteering. There is an even bigger impact when examining the median values. The median values for the three long modules are the only values greater than 0. In addition, the percentages of respondents reporting that they do volunteer work are much higher for the three long modules than the shorter ones. Collectively, these results offer substantiation to our thesis that the number of prompts related to volunteering do matter in collecting data about individual philanthropy.

In addition to these indicators, we looked at the correlations between module means, the number of volunteering variables, and the number of questions directly measuring volunteering. Some modules included questions related

Table 2. Demographics of Volunteer-Only Module

	Total Sample	*Volunteer Only*
Sample size	781	104
Female (%)	68.0	65.8
Couples (%)	62.5	54.5
White (%)	88.2	81.6
Age		
Mean	46.1	41
Median	43	39.5
Min	18	19
Max	89	82
Education (%)		
High school diploma or less	32.2	26.3
Some college	33.2	35.5
Bachelor's degree	20.6	18.4
Graduate or professional school (joint tests)	14.0	19.7
Income (%)		
$0 to 40,000	47.7	50.0
$40,000 to $80,000	29.3	26.8
$80,000 or more (joint tests)	23.0	23.2

Note: No significant differences between individual modules and total sample means or proportions.

to, but not directly measuring, volunteering (e.g., motivations for volunteering, childhood experiences, etc.). These correlations, shown at the bottom of Table 3, show strong statistically significant links between the number of prompts related to volunteering and reported hours of formal volunteer service.

Table 3 also includes the results for informal volunteering. Informal volunteering was measured in four modules: Area, Area-Method, Method-Area, and Volunteer Only. Because of the wide range between means of these modules, all four are significantly different from the total sample mean for informal volunteering. When looking at volunteers only, the Area module is highly significantly different from the total sample mean, and the Area-Method module is weakly significantly different. We note that for the three modules with many questions measuring informal volunteering (the Volunteer-Only, Area-Method, and Method-Area modules), the mean, median, and percentage volunteering indicators are all higher for informal than formal volunteering. This is consistent with the results of other researchers who used in-depth prompts for informal volunteering (Havens & Schervish, 2001; O'Neill, 2001). On the other hand, the formal volunteering mean, median, and percentage in the Area module (which only includes one question on informal volunteering) are all higher than those for informal volunteering. This is consistent with results obtained by Independent Sector (1999) in their national survey.

An interesting dilemma with both giving and volunteering data occurs when the sample includes extreme outliers. Generally, surveys of household

Table 3. Comparison of Yearly Volunteering, Including Outliers

	Total Sample[a]	*Very Short*	*PSID*	*Area*	*Method*	*Area-Method*[a]	*Method-Area*[a]	*Volunteer Only*[a]
Total *N*	885	110	113	106	103	124	124	104
N (volunteering)	777	106	111	106	103	124	124	104
Number of questions		16	137	337	368	676	680	223
Number of volunteering variables		4	9	170	204	215	215	215
Number of questions measuring volunteering		3	9	104	136	215	215	215
Formal volunteering (hours)								
All subjects								
Mean	170.39	50.7***	35.7***	143.09	63.04***	236.42*	183.74	462.84***
Median	6	0	0	0	0	48	36	78.5
Standard deviation	4458.81	212.9	115.34	291.74	154.79	432.18	332.33	959.3
Percentage volunteering	52.8	39.6	28.8	40.6	37.1	66.1	70.2	76.0
Volunteers only								
Mean	322.9	126.36***	123.84***	352.74	169.93***	357.51	261.89	609.3**
Median	120	40	65	240.18	115	160	116	144
Standard deviation	591.63	325.89	189.58	370.22	216.61	489.7	370.48	1060.58
Informal volunteering (hours)								
All subjects								
Mean	422.39	NA	NA	17.21***	NA	719.23***	636.48**	758.19***
Median	48.02	NA	NA	0	NA	198	210	348
Standard deviation	952.59	NA	NA	78.47	NA	1297.7	1208.99	1185.4
Percentage volunteering	62.9	NA	NA	12.50	NA	76.6	87.9	88.5
Volunteers only								
Mean	671.11	NA	NA	137.64***	NA	972.59*	724.97	807.95
Median	276.00	NA	NA	96.07	NA	408	252	420
Standard deviation	1129.65	NA	NA	186.8	NA	1508.29	1381.96	1298.02

(continued)

Table 3 (continued)

For all subjects (formal volunteering)
Correlation between number of questions and module means: .24401
Correlation between number of volunteering variables and module means: .618196**
Correlation between number of questions measuring volunteering and module means: .74615***
For volunteers only (formal volunteering)
Correlation between number of questions and module means: .223853
Correlation between number of volunteering variables and module means: .644337**
Correlation between number of questions measuring volunteering and module means: .705423**

Note: PSID = Panel Study of Income Dynamics; NA = Not Applicable. Asterisks indicate significant differences between individual modules and total sample means or proportions. *$p \leq .10$. **$p \leq .05$. ***$p \leq .01$.
a. Sample includes outliers (>2000 hours/yr).

giving behavior do not include many, if any, very wealthy individuals because these individuals are hard to reach unless a large sample is drawn or special techniques are used, such as oversampling (Havens & Schervish, 2001), and so the probability of including them in the sample is very low. Hence, household survey data is unlikely to include donations above several hundred thousand dollars per year, although donations much higher than this actually do occur. However, with volunteering data, the situation can be different. When we examined our data without outliers (see Table 4), we found similar results to those in Table 3. We defined outliers as respondents who reported more than 2,000 hours of formal voluntary service in the past year, which represents 50 weeks of full-time (40 hours) work. By excluding outliers, the mean of the Very Short module becomes the lowest among all modules. The mean of the Volunteer-Only module also drops considerably, although it is still the highest among the modules. In Table 4, as in Table 3, the means, medians, and percentages of volunteering respondents are highest for the three longest modules (Volunteer Only, Area-Method, and Method-Area). We should note that these three modules not only included detailed questions about volunteering but also specifically included questions likely to elicit outlier responses (such as prompting religious mission work, foster care, and Peace Corps service). The sample for the Very Short module was the only other module that included an outlier response. (The Area-Method module included two outliers, Method-Area included one, and the Volunteer-Only module included seven.)

When we looked at the types of volunteer service that our outliers provided, we found an interesting pattern. All of the outliers provided service in multiple ways (teaching, raking leaves, walkathons, etc.). Most of the outlier subjects reported the majority of their service in one or two areas, with a smattering in other areas (high-service areas in our sample included Peace Corps; baby-sitting or foster care; missionary, deacon, or clergy work; choir or

Table 4. Comparison of Yearly Volunteering, No Outliers Greater Than 2000 Hours/Year

	Total Sample	*Very Short*	*PSID*	*Area*	*Method*	*Area-Method*	*Method-Area*	*Volunteer Only*
Total *N*	885	110	113	106	103	124	124	104
N (volunteering)	766	105	111	106	103	122	123	97
N outliers	11	1	0	0	0	2	1	7
Number of questions		16	137	337	368	676	680	223
Number of volunteering variables		4	9	170	204	215	215	215
Number of questions measuring volunteering		3	9	104	136	215	215	215
Formal volunteering (hours)								
All subjects								
Mean	131.59	30.73***	35.7***	143.09	63.04***	203.62**	168.39	259.47***
Median	5	0	0	0	0	39.5	36	62
Standard deviation	287.52	75.91	115.34	291.74	154.79	349.22	286.16	454.62
Percentage volunteering	52.1	39.0	28.8	40.6	37.1	65.6	69.9	74.2
Volunteers only								
Mean	252.63	78.71***	123.84***	352.74*	169.93**	310.52	240.84	349.57*
Median	114	40	65	240.18	115	155.5	115	119.5
Standard deviation	358.11	105.41	189.58	370.22	216.61	391.38	316.03	497.51
Informal volunteering (hours)								
All subjects								
Mean	393.37	NA	NA	17.21***	NA	719.23***	636.48**	758.19***
Median	48.00	NA	NA	0	NA	198	210	348
Standard deviation	843.49	NA	NA	78.47	NA	1297.7	1208.99	1185.4
Percentage volunteering	62.1	NA	NA	12.50	NA	76.6	87.9	88.5
Volunteers only								
Mean	633.36	NA	NA	137.64***	NA	972.59*	724.97	807.95
Median	264.00	NA	NA	96.07	NA	408	252	420
Standard deviation	997.28	NA	NA	186.8	NA	1508.29	1381.96	1298.02

(continued)

Table 4 (continued)

For all subjects (formal volunteering)
Correlation between number of questions and module means: .532996**
Correlation between number of volunteering variables and module means: .773665***
Correlation between number of questions measuring volunteering and module means: .875793***
For volunteers only (formal volunteering)
Correlation between number of questions and module means: .503038*
Correlation between number of volunteering variables and module means: .770821***
Correlation between number of questions measuring volunteering and module means: .742585***

Note: PSID = Panel Study of Income Dynamics; NA = Not Applicable. Asterisks indicate significant differences between individual modules and total sample means or proportions. *$p \leq .10$. **$p \leq .05$. ***$p \leq .01$.

orchestra work; emotional care; animal care; health care; and search and rescue operations). We conclude that the likelihood of eliciting outlier responses may be increased by asking a large number of detailed questions, including questions about providing foster care, service in the Peace Corps, and missionary work. In other words, if a researcher wants to study high-end donations, he or she must draw a very large sample or oversample the very wealthy; to study high-end volunteering behavior, the researcher needs to include survey questions likely to elicit such responses. There also may be a respondent self-selection problem with the Volunteer-Only module. People who volunteer a great deal may be more likely to agree to answer a series of questions in a survey specifically about volunteer behaviors.

Based on these three important indicators of volunteering (mean and median volunteer hours per year and the percentage who volunteered), it does seem that a greater number of detailed prompts do stimulate greater recall. Of course, there is a danger that respondents report volunteer service that they did not actually perform in an effort to conform to a perceived set of social expectations or to please or impress the interviewer. Although this is a problem in any type of survey research, it may be exacerbated by repeated questions about volunteering, which may convey the message to the respondent that it is expected or normal to provide volunteer service. Similarly, respondents may be embarrassed or bored if they repeatedly report no volunteer work in surveys that ask about service in many different areas or methods. On the other hand, Hall (2001) noted that respondents are likely to forget brief or infrequent volunteering, which may cause them to underreport their service. Furthermore, Havens and Schervish's (1997) diary study, which used daily prompting, found that 100% of their sample spent time volunteering or providing unpaid assistance during the course of a year. This strongly suggests that fewer prompts lead to an understatement of volunteer service, rather than that more prompts create a false or exaggerated picture.

In addition to looking at volunteering itself, we also examined the relationship between giving and volunteering. Other researchers have reported that volunteers are more likely than nonvolunteers to make charitable contributions (Hall et. al., 2001; Independent Sector, 1999; Jalandoni & Hume, 2001). Likewise, for most of our modules we found that more volunteers made contributions than nonvolunteers (see Table 5). We found the same proportion of donors from volunteers versus nonvolunteers in the Area-Method module, and a slightly higher proportion of donors from nonvolunteers in the Method module. In all of the other modules, volunteers were more likely to be donors than were nonvolunteers. Chi-square tests for differences in the proportion between donors and nondonors among volunteers and nonvolunteers found that volunteers were much more likely to be donors than nonvolunteers in the total sample, as well as the PSID module, the Area module, and the Method-Area module. However, when we looked at the correlations between total dollars given and the number of hours volunteered, we found that the correlations for most modules were not particularly high and also not statistically significant. (The correlations for the Method and PSID modules, and for the total sample, were all statistically significant at the .05 level or less.) On the other hand, the number of questions per module is highly correlated with percentage of volunteers who make donations. This is yet another validation of our main theme that the length of the survey, and the detail of the questions, do matter in the measurement of volunteering.

Another issue that we investigated in this study relates to refusal rates for the different modules. We found the expected negative correlation between incentive amounts and refusal rates, and positive correlations between number of prompts and refusal rates. After we removed outliers, the incentive amounts did not make a significant difference in the reported amount of total annual volunteering. Hence, we conclude that higher refusal rates for longer surveys means that they are more costly to conduct, but incentives do not change the total amount of volunteering reported. These results are summarized in Table 6.

MULTIVARIATE RESULTS

Now let's turn to the multivariate results in which we used OLS, tobit, probit, and Heckman two-stage regression models to examine the marginal impacts of the independent variables and to test the impact of differences in the modules more formally. (The probits determine the marginal probability of an individual volunteering at all given various characteristics.) Table 7 presents the main results of our regression analyses. These models assume that the various survey modules do not alter the coefficients of the other variables but that they shift the regression line up or down depending on the impact a survey module has on reported volunteering. (The second stage of the Heckman

Table 5. Relationship Between Giving and Volunteering Across Modules

	Total Sample	*Very Short*	*PSID*	*Area*	*Method*	*Area-Method*	*Method-Area*	*Volunteer Only*
Sample size	781	110	113	106	103	124	124	104
Number of questions		16	137	337	368	676	680	223
Volunteers (%)								
No donations	8.4	25.0	17.2	14.0	8.7	2.4	1.1	NA
Donations	91.6	75.0	82.8	86.0	91.3	97.6	98.9	NA
Nonvolunteers (%)								
No donations	26.6	33.9	37.8	44.4	7.0	2.4	18.9	NA
Donations	73.4	66.1	62.2	55.6	93.0	97.6	81.1	NA
Chi-square	37.544***	0.844	4.059**	10.905***	0.100	0.000	13.581***	NA
Pearson correlations								
Total money given and hours volunteered	.136***	.028	.221**	.148	.468***	.038	.025	NA
Number of questions and percentage of volunteers who donate	.978***							
Number of questions and percentage of nonvolunteers who donate	.663							

Note: PSID = Panel Study of Income Dynamics; NA = Not Applicable.
**$p \leq .05$. ***$p \leq .01$.

Table 6. Refusal Rates

Instrument	*Number of Prompts*	*Number of Volunteering Variables*	*Number of Volunteering Questions*	*Incentive*	*Obtained* N	*Percentage Refusals**
Very Short	16	4	3	0	110	7.90
IU-PSID	137	9	9	0	113	12.10
Volunteer Only	223	215	215	0	104	8.40
Area	337	170	104	0	106	10.70
Method	368	204	136	0	103	11.30
IU-Area-Method	676	215	215	0	28	18.93
	676	215	215	10	29	20.97
	676	215	215	25	33	19.58
	676	215	215	50	34	11.50
IU-Method-Area	680	215	215	0	29	23.00
	680	215	215	10	30	15.80
	680	215	215	25	31	21.40
	680	215	215	50	34	16.80

Correlations	*Method-Area*	*Area-Method*	*Total*
Incentive and percentage refusals	–.45	–.84	–.65*
Number of prompts and percentage refusals (all modules)			.81***
Number of volunteering variables and percentage refusals			.539*
Number of questions measuring volunteering and percentage refusals			.623*

Note: PSID = Panel Study of Income Dynamics.
*$p \leq .10$. ***$p \leq .01$.

two-stage regressions generated results similar to the tobits, so we did not report them. They are available from the authors upon request.)

Looking at the demographic variables in the regression models (i.e., age, race, gender, education, income, household status, and itemization status), the most surprising result is how few of the demographic variables attain statistical significance in the OLS and tobit regressions. None of the following demographic variables are significant at traditional levels: age, age squared, income, race (White), couples, or itemizers. Only education and gender are consistently significant. That is, females were more likely to volunteer than males and to volunteer more hours. Similarly, those with a high school education or less were less likely to volunteer and to volunteer fewer hours than those with more education. This is consistent with the results of other studies, which examine the demographic correlates of volunteering (e.g., Independent Sector, 2001). We also found that Whites are 13% more likely to volunteer than non-Whites but that there is no significant difference in the number of hours volunteered by race after controlling for other socioeconomic demographics. There also is weak evidence (significant at the 10% level) that there is a slight

Table 7. Volunteering Regression Results

	OLS (full sample) Coefficient[a,c,f,g]	*OLS (no outliers) Coefficient*[a,c,f,g]	*Tobit (full sample) Impact*[a,c,d,e,g]	*Tobit (no outliers) Impact*[a,c,d,e,g]	*Probit (full sample) Impact*[b,c,d,e,g]	*Probit (no outliers) Impact*[b,c,d,e,g]
Constant	27.01	85.13	–137.76	–53.66	0.035	0.035
Age	13.01*	5.69	6.76	2.72	–0.003	–0.003
Age squared	–0.13	–0.06	–0.07	–0.03	0	0
Income	–5.14	6.34	5.65	9.41	0.027*	0.028*
White	–6.71	–25.55	40.35	13.59	0.129**	0.129**
Couples	–89.12*	–24	–59.56*	–19.70	–0.025	–0.019
High school education or less	–104.79**	–52.3*	–115.77***	–69.24***	–0.151***	–0.149***
Female	97.52**	81.84***	94.08***	76.38***	0.133***	0.136***
Itemize	–10.41	25.49	32.92	41.34*	0.115**	0.12**
Method module	–161.08**	–142.02***	–186.11***	–144.43***	–0.268***	–0.268***
Area module	–40.27	–29.93	–97.87*	–65.82*	–0.23***	–0.231***
Method-Area module	–12.27	–12.2	–10.69	–917.35	0.007	0.005
PSID module	–197.34***	–181.12***	–302.43***	–234.20***	–0.436***	–0.438***
Very Short module	–177.34**	–182.94***	–183.91***	–164.18***	–0.232***	–0.238***
Volunteer Only module	247.47***	43.41	120.00**	16.62	0.022	–0.003
N	506	498	506	498	506	498
Adjusted R^2	.083686	.07957	NA	NA	.214	.213
Log-likelihood	–3833.58	–3555.21	–2375.26	–2202.56	–288.435	–285.042

Note: PSID = Panel Study of Income Dynamics; NA = Not Applicable; OLS = Ordinary Least Squares.
a. Estimates the change in volunteering hours due to changes in each variable.
b. Estimates the marginal probability of volunteering due to changes in each variable.
c. All coefficients for categorical variables are relative to the values of the excluded category for that variable (minority, single, more than high school education, male, and Area-Method module.
d. Statistical significance is determined for the coefficients on the latent index for donations.
e. Statistical significance with respect to the latent indicator variable.
f. Statistical significance was determined using White's robust standard errors.
g. Table with standard errors or *t* scores is available on request from the authors.
*$p < .10$. **$p < .05$. ***$p < .01$.

increase in the probability of volunteering (3%) as income increases, but income has no significant effect on the number of hours volunteered.

Of more interest to this methodological study is the analysis of the dummy variables for different modules. A rank ordering of the module coefficients in each regression model presents an interesting picture. In each regression excluding the extreme outliers, the coefficient for the IU-Method-Area module and the Volunteer-Only module are not significantly different from the IU-Area-Method module (which was the excluded category). This makes sense given that these three modules asked identical questions with respect to volunteering behaviors. The strength of the outliers (those who are full-time volunteers) is shown by the difference in the Volunteer-Only module when we include the outliers: It goes from having a small and insignificant coefficient when we exclude the outliers to a large and highly significant coefficient when we include them. The coefficients for all other modules are negative, indicating that these modules are associated with lower probabilities of reporting any volunteer time (in the probits) and fewer hours of volunteer activities by respondents (in the OLS and tobit regressions), after controlling for income, education, race, household status, age, gender, and whether the person itemized their deductions. The exception to this rule is that the Area (Independent Sector) module is less likely to identify a volunteer, but this is only weakly significant and the difference in volunteer hours, although negative, is imprecisely measured.

In addition, the two shortest modules (IU-PSID and Very Short) have the largest negative impact on reported volunteer hours in all regressions, which suggests that the length and number of prompts in the surveys do matter in terms of gathering data on volunteer behavior. This is further reinforced by the fact that the two shortest modules with respect to volunteer behaviors are also among the least likely to identify any volunteer behavior among respondents in the probits. In summary, our research indicates that survey methodology is important in providing optimal estimates of volunteering. The more individuals are prompted in detail, the more likely they are to report volunteer service, and the higher the average number of hours reported.

CONCLUSIONS

Our results have shown that whether using simple means or multivariate analyses, the longer and more detailed the module, the more likely an individual was to report they provided volunteer service in the past year and the higher the average number of hours reported, even after controlling for differences in age, income, gender, household status, itemization status, and race. However, regardless of which module was used, females were more likely to volunteer than males and to volunteer more hours. Similarly, those with a high school education or less were less likely to volunteer and to volunteer fewer hours than those with more education.

Further research is needed to ascertain whether there may be a "point of diminishing returns" in terms of optimal survey length and detail. One interpretation of our results is that the reported level of volunteering decreased when detailed volunteering questions came after a series of detailed giving questions. In a previous study (Rooney et al., 2001), we found that there was no significant difference in reported giving between modules that contained detailed questions on giving followed by detailed questions on volunteering and one that contained only detailed questions on giving. It may be that, in order to optimize recall of giving or volunteering, the researcher needs to include many detailed questions about either giving or volunteering, but not both. This, of course, is complicated further by prior research and confirmed by our research that volunteers are more likely to be donors than are nonvolunteers. On the other hand, it may be that recall is inhibited for high-end volunteers only.

An important methodological question relates to the optimal service unit to query in a survey (i.e., volunteer hours per week, month, or year). Our shortest modules (PSID and Very Short) asked respondents only to report volunteering hours per year, and we obtained the lowest mean levels of service reported on these two modules. Previous research (e.g., Hall, 2001) indicates that subjects may have difficulty recalling minor episodes of volunteer service, especially service performed a year prior, without more specific prompts. This suggests that asking about volunteer service in smaller units, such as hours per week or month, helps to aid recall. It also may be that using multiple units (e.g., hours per week as well as per month) helps the respondent and researcher fully capture the variety of service performed.

There also is need for further research on other important methodological questions. For example, does increasing the number of prompts increase recall or lead to overestimates of volunteering (resulting from the desire to conform to expectations or to please the interviewer)? Longitudinal research could be used to track changes in personal volunteering behavior over time, as related to characteristics such as youth experiences, household status, education level, and socioeconomic status. Another line of research, which we are planning to pursue in the future, will look at whether the understanding of survey phraseology may be different for each gender or for different racial or ethnic groups.

Additional research also is needed to find out whether the PSID or Very Short survey could serve as a useful proxy for some portion of formal volunteering. We would urge caution in using short surveys to estimate volunteering until further research can be conducted because there may be regional, racial, social class, or other differences in how respondents react to these surveys. If we understood in detail what types of underreporting occur in different populations when confronted with a short survey, then we could make educated guesses about the true levels of volunteering.

Our results coalesce to suggest that when measuring whether a given person is likely to report being a volunteer and how many hours they are likely to

report having volunteered, methodology is destiny. The more detailed the prompts, the more likely the person is to recall being a volunteer and the more hours they are likely to recall having volunteered in the prior year.

References

Bradley, R., Holden, S., & McClelland, R. (1999). *A robust estimation of the effects of taxation on charitable contributions* (Working Paper). Washington, DC: Bureau of Labor Statistics.

Hall, M. (2001). Measurement issues in surveys of giving and volunteering and strategies applied in the design of Canada's national survey of giving, volunteering and participating. *Nonprofit and Voluntary Sector Quarterly*, *30*(3), 515-526.

Hall, M., McKeown, L., & Roberts, K. (2001). *Caring Canadians, involved Canadians: Highlights from the 2000 National Survey of Giving, Volunteering and Participating*. Ottawa: Canadian Centre for Philanthropy and Statistics Canada.

Havens, J. J., & Schervish, P. G. (1997). *Our daily bread: Findings from the first diary study on giving and receiving care*. Chestnut Hill, MA: Boston College, Social Welfare Research Institute.

Havens, J. J., & Schervish, P. G. (2001) The methods and metrics of the Boston area diary study. *Nonprofit and Voluntary Sector Quarterly*, *30*(3), 527-550.

Independent Sector. (1999). *Giving and volunteering in the United States, 1999 edition*. Washington, DC: Author.

Independent Sector. (2001). *Giving and volunteering in the United States, 2001 edition*. Washington, DC: Author.

Jalandoni, N., & Hume, K. (2001). *America's family volunteers*. Washington, DC: Independent Sector.

O'Neill, M. (2001). Research on giving and volunteering: Methodological considerations. *Nonprofit and Voluntary Sector Quarterly*, *30(3), 505-514.*

O'Neill, M., & Roberts, W. L. (2000). *Giving and volunteering in California*. San Francisco: Institute For Nonprofit Organization Management, University of San Francisco.

Rooney, P. M., Steinberg, K. S., & Schervish, P. G. (2001). A methodological comparison of giving surveys: Indiana as a test case. *Nonprofit and Voluntary Sector Quarterly*, *30*(3), 551-568.

Kathryn S. Steinberg serves as the assistant director of research at the Indiana University Center on Philanthropy. She holds a doctorate in educational psychology from Purdue University and has taught at the University of Nebraska at Kearney, University of Puerto Rico, and Henderson Community College in Kentucky. At the Center on Philanthropy, she assists with a number of research projects including Giving USA, *Indiana Gives, America Gives, the Philanthropic Giving Index, the Million Dollar List, and several grant programs.*

Patrick M. Rooney is the chief operating officer and the director of research at the Indiana University Center on Philanthropy. He is an associate professor of economics at Indiana University–Purdue University at Indianapolis, and is a member of the Indiana University Philanthropic Studies faculty. He directs the research activities at the Center on Philanthropy, including the Panel Study on Income Dynamics (in collaboration with the University of Michigan), the Administrative Cost Study (in collaboration with the Urban Institute), and research for Giving USA.

William Chin is a visiting lecturer in the department of economics, Indiana University–Purdue University at Indianapolis. He is all but dissertation from Iowa State University where he did the co-major program in economics and statistics.