The relationship between linkage refusal and selected health conditions of survey respondents

Judith Weissman[*], **Jennifer D Parker**[†], **Donna M Miller**[‡], **Eric A Miller**[**], and **Renee M Gindi**[††]

[*]**Institution:** formerly of National Center for Health Statistics

[†]**Institution:** National Center for Health Statistics

[‡]**Institution:** National Center for Health Statistics

[**]**Institution:** former at National Center for Health Statistics

[††]**Institution:** National Center for Health Statistics

Abstract

To maximize limited resources and reduce respondent burden, there is an increased interest in linking population health surveys with other sources of data, such as administrative records. Health differences between adults who consent to and refuse linkage could bias study results with linked data. National Health Interview Survey (NHIS) data are routinely linked to administrative records from the Social Security Administration and the Centers for Medicare and Medicaid Services. Using the NHIS 2010-2013, we examined the association between selected health conditions and respondents' linkage refusal. Linkage refusal was significantly lower for adults with serious psychological distress, chronic obstructive pulmonary disease, diabetes, heart disease, stroke, hypertension, and cancer compared to those without these conditions. Linkage refusal decreased as the number of conditions increased and health status decreased. Our finding that linkage consent was associated with respondents' health characteristics suggests that researchers should try to address potential linkage bias in their analyses.

Keywords

national health interview survey; health survey; consent bias; record linkage

BACKGROUND

Linking data collected in surveys with other sources of data, such as administrative data, can maximize limited resources and reduce respondent burden. Similar to nonresponse bias, differential linkage refusal increases the possibility of 'consent' or 'linkage' bias (Hill et al. 2002; Jenkins et al. 2006; Sala et al. 2012). As for nonresponse bias (Groves 2006), linkage bias can lead to inaccurate estimates of population characteristics (Sakshaug and Kreuter 2012; Sakshaug et al. 2012). Previous studies have examined health differences between those who consent to link and those who do not. Some studies reported poorer health among those who consented, other studies found no systematic health differences by consent status, and some studies found differences by consent status that varied among health endpoints

(Bohensky et al. 2010; Carter et al. 2010; Cruise et al. 2015; Dunn et al. 2004; Huang et al. 2007; Kho et al. 2009; Knies et al. 2012; Knies and Burton 2014; Young et al. 2001).

The National Health Interview Survey (NHIS), conducted by the Centers for Disease Control and Prevention's National Center for Health Statistics, is a principal source of information on the health of the civilian noninstitutionalized U.S. population. The cross-sectional NHIS data are routinely linked to administrative records from the Social Security Administration and the Centers for Medicare and Medicaid Services to maximize the scientific value of the survey without increasing respondents' reporting burden (Golden et al. 2015). Respondents' refusal to provide a Social Security Number (SSN) or Medicare Health Insurance Claim Number (HICNUM), used to facilitate linkage, may affect the composition of the linked sample. Our objective was to examine the role of health status on the propensity to refuse linkage in a large nationally representative U.S. sample. We estimated associations of health conditions, general health status, and the total number of health conditions with linkage refusal.

METHODS

DATA SOURCE

Data came from the 2010–2013 NHIS. Details about the sampling frame and sample design have been published previously (Parsons et al. 2014). Interviews were conducted in respondents' homes, with some telephone follow-ups. One person from each family answered questions about family members' age, health insurance coverage, and general health status. One randomly selected "sample adult" aged 18 years and older from each family answered questions about himself or herself. In each year, the final response rate among sample adults was over 60 percent and data were obtained from approximately 30,000 sample adults. Details of the sample sizes and response rates are available in the survey documentation for each relevant year (e.g. NCHS 2013).

LINKAGE AND LINKAGE REFUSAL

At the conclusion of the sample adult interview, respondents were asked: "To help us link your survey data with vital statistics and health-related records of other government agencies, we would like the last four digits of your Social Security Number. The National Center for Health Statistics uses this information for research purposes only. Providing this information is voluntary. Federal laws authorize us to ask for this information and require us to keep it strictly private. There will be no effect on your benefits if you do not provide this information. What are the last four digits of your Social Security Number?" Respondents were similarly asked for the last four digits of the HICNUM if they were Medicare eligible. Respondents who answered "no," "refused," "don't know," or "don't have a SSN (or Medicare number)," were asked: "May we try to link your survey data without [your SSN/HICNUM]?"

The questions were the same whether administered in-person or by telephone. For the analysis, record-linkage consent was defined as permission to use the SSN or HICNUM to link, or permission to link without either an SSN or HICNUM. Record-linkage refusal was

defined as refusal to provide either the SSN or HICNUM combined with refusal to allow linkage without the SSN or HICNUM. For the 2010–2013 NHIS, unweighted linkage refusal ranged from 10.0 percent in 2010 to 11.8 percent in 2013.

HEALTH CHARACTERISTICS

The selected health conditions examined were hypertension, heart disease, diabetes, cancer, chronic obstructive pulmonary disease (COPD), serious psychological distress (SPD), and stroke. Hypertension was defined as having been told by a doctor or other health professional of having hypertension on two or more different visits. Heart disease was based on responses to questions about having ever been told by a doctor or other health professional of having coronary heart disease, angina (angina pectoris), a heart attack (myocardial infarction) or any heart disease or condition. Diabetes (excluding borderline diabetes) and stroke were also based on "yes" or "no" responses to questions about having ever been told by a doctor or other health professional of having these conditions. Cancer was based on responses to having ever been told by a doctor or other health professional of having cancer or a malignancy excluding nonmelanoma skin cancer. COPD was based on positive responses to questions about having ever been told by a doctor or other health professional of having emphysema or having been told in the past 12 months of having had chronic bronchitis (American Thoracic Society 2004). SPD was measured using the K6, a series of six psychological distress questions asking how often a respondent experienced symptoms of psychological distress during the past 30 days (Kessler et al. 2003). Small numbers of respondents were missing data: hypertension (n=253); heart disease (n=54); diabetes (n=91); cancer (n=106); COPD (n=44); SPD (n=2,025); stroke (n=121) respondent-assessed health status (n=66); and total number of conditions (n=2,534). Records with missing data were not included in the health-outcome specific analyses but were retained in the overall study population.

DEMOGRAPHIC CHARACTERISTICS

Demographic characteristics included race and ethnicity, age, and poverty. Race and ethnicity were categorized as Hispanic, non-Hispanic White, non-Hispanic Black and all other races/ethnicities. Age groups included 18–44 years, 45–64 years, and 65 years and over. Annual family income was categorized into a poverty index ratio (PIR) of below 100 percent of the federal poverty level, 100 percent–199 percent of the federal poverty level, 200 percent–399 percent of the federal poverty level, and 400 percent or more of the federal poverty level. Imputation was used to assign a poverty level for records with missing income data (Schenker et al. 2009). Because Miller et al. found a possible association with imputation status and record-linkage refusal (Miller et al. 2011), imputation status was included as an independent variable, with categorical responses: providing all income information, providing income by categories, and providing no income information.

STATISTICAL ANALYSIS

Point estimates, corresponding variances, and 95 percent confidence intervals were calculated using SAS callable SUDAAN software (SUDAAN 2008) to provide weighted estimates and account for the complex sample design. Overall associations for categorical variables were evaluated using the Rao Scott Chi Square statistic for weighted survey data.

We used multivariable logistic regression models that included one health condition or characteristic and controlled for age group, sex, race/ethnicity, PIR, and income imputation status. Trend tests identified linear relationships between linkage refusal, health status, and total number of health conditions.

RESULTS

The study population included 129,253 adults with complete demographic data. Approximately 11 percent of respondents refused record linkage. Close to 50 percent were aged less than 45 years (sample mean age was 47 years) and almost 70 percent were non-Hispanic White (Table 1). The majority of adults were in the 200–399 percent PIR (30 percent) and 400 percent PIR (37 percent) groups. Nearly 39 percent of adults had at least one health condition.

In bivariate analysis, Hispanic adults were the most likely to refuse record-linkage (13 percent) compared to non-Hispanic White (10 percent) and non-Hispanic Black (10 percent) adults (Table 1). Adults in the youngest age group were less likely to refuse linkage than other age groups. Adults in the lowest income groups were less likely to refuse linkage than those in higher income groups (p for trend < 0.01). Adults who did not provide income information were more likely to refuse linkage (29 percent) than adults who reported their income (9 percent). Adults with any one health condition were less likely to refuse linkage than adults without the condition.

In separate multivariable logistic regression models, each of the health conditions remained inversely associated with linkage refusal after controlling for age, race/ethnicity, sex, PIR, and income imputation status (Table 2). In addition, the inverse association was stronger as number of conditions increased (p for trend <0.01; Table 2). Similarly, those with better general health status had higher rates of linkage refusal (p for trend <0.01). Sensitivity analyses that excluded the income imputation variable (an indicator of refusal to provide income information) and included the survey year (to allow for changes in record-linkage refusal over time) had similar results to the main findings.

DISCUSSION

Our major finding was that adults with hypertension, heart disease, diabetes, cancer, COPD, SPD, or stroke were less likely to refuse linkage compared to adults without one of these health conditions. The finding was further supported by a dose-response association between linkage refusal and number of health conditions and health status. As the number of health conditions increased, adults were less likely to refuse record-linkage. Similarly, refusal to link decreased with poorer reported health status.

The results are consistent with other studies showing associations between health and linkage refusal (Dunn et al. 2004; Knies et al. 2012; Knies and Burton 2014; Young et al. 2001). Differences in linkage refusal by health characteristics may be explained by leverage-salience theory; faced with a survey request of interest, respondents cooperate at higher rates than those less interested (Groves et al. 2000).

Health differences between adults who consent and refuse linkage could bias study results. Weighting methods to decrease linkage bias, similar to survey nonresponse bias, have been proposed for analyses of NCHS linked data (Judson et al. 2013). However, weighting methods may not fully account for potential biases. The best approach may depend on the study question. It is possible to obtain information for a large number of health factors to inform approaches for analyses using the linked data.

Our report on record-linkage refusal in a national sample found that respondents without selected health conditions were more likely to refuse linkage. Researchers should evaluate potential biases in their analyses due to linkage refusal to determine appropriate adjustment.

REFERENCES

American Thoracic Society. 2004 Standards for the diagnosis and management of patients with COPD. American Thoracic Society New York, NY.

Bohensky MA, Jodlley D, Sundararajan V, Evans S, Pilcher DV, Scott I and Brand CA. 2010 Data linkage: a powerful research tool with potential problems. BMC Health Services Research 10(1): 346. doi: 10.1186/1472-6963-10-346. [PubMed: 21176171]

Carter K, Shaw C, Hayward M and Blakely T. 2010 Understanding the determinants of consent for linkage of administrative health data with a longitudinal survey. K_ò_tuitui: New Zealand Journal of Social Sciences Online 5(2): 53–60.

Cruise SM, Patterson L, Cardwell CR and O'Reilly D. 2015 Large panel-survey data demonstrated country-level and ethnic minority variation in consent for health record linkage. Journal of Clinical Epidemiology 68(6): 684–692. doi: 10.1016/j.jclinepi.2015.01.011. [PubMed: 25753658]

Dunn KM, Jordan K, Lacey RJ, Shapley M and Jinks C. 2004 Patterns of consent in epidemiologic research: evidence from over 25,000 responders. American Journal of Epidemiology 159(11): 1087–1094. doi: 10.1093/aje/kwh141. [PubMed: 15155293]

Golden C, Driscoll AK, Simon AE, Judson DH, Miller EA and Parker JD. 2015 Linkage of NCHS Population Health Surveys to Administrative Records from Social Security Administration and Centers for Medicare Medicaid Services. Vital and Health Statistics 1(58): 1–53.

Groves RM 2006 Nonresponse rates and nonresponse bias in household surveys. Public Opinion Quarterly 70(5): 646–675.

Groves RM, Singer E and Corning A. 2000 Leverage-saliency theory of survey participation: description and an illustration. Public Opinion Quarterly 64(3): 299–308. [PubMed: 11114270]

Hill S, Atkinson J and Blakely T. 2002 Anonymous record linkage of census and mortality records: 1981, 1986, 1991, 1996 Census cohorts Department of Public Health, Wellington School of Medicine and Health Sciences. University of Otago, Wellington, New Zealand.

Huang N, Shih SF, Chang HY and Chou YJ. 2007 Record linkage research and informed consent: who consents? BMC Health Services Research 7(18): 1–5. doi: 10.1186/1472-6963-7-18. [PubMed: 17199886]

Jenkins SP, Cappellari L, Lynn P, Jäckle A and Sala E. 2006 Patterns of consent: evidence from a general household survey. Journal of the Royal Statistical Society: Series A (Statistics in Society) 169(4): 701–722.

Judson DH, Parker JD and Larsen MD. 2013 Adjusting sample weights for linkage-eligibility using SUDAAN. National Center for Health Statistics, Hyattsville Maryland. Available at http://www.cdc.gov/nchs/data/datalinkage/adjusting_sample_weights_for_linkage_eligibility_using_sudaan.pdf.

Kessler RC, Barker PR, Colpe LJ, Epstein JF, Gfroerer JC, Hiripi E, Howes MJ, Normand SLT, Manderscheid RW, Walters EE and Zaslavsky AM. 2003 Screening for serious mental illness in the general population. Archives of General Psychiatry 60(2): 184–189. [PubMed: 12578436]

Kho ME, Duffett M, Willison DJ, Cook DJ and Brouwers MC. 2009 Written informed consent and selection bias in observational studies using medical records: systematic review. British Medical Journal 338: b866. [PubMed: 19282440]

Knies G and Burton J. 2014 Analysis of four studies in a comparative framework reveals: health linkage consent rates on British cohort studies higher than on UK household panel surveys. BMC Medical Research Methodology 14: 125. doi: 10.1186/1471-2288-14-125. [PubMed: 25430545]

Knies G, Burton J and Sala E. 2012 Consenting to health record linkage: evidence from a multi-purpose longitudinal survey of a general population. BMC Health Services Research 12: 52. doi: 10.1186/1472-6963-12-52. [PubMed: 22390416]

Miller D, Gindi R and Parker J. 2011 Trends in Record Linkage Refusal Rates: Characteristics of National Health Interview Survey Participants Who Refuse Record Linkage Paper presented at the JSM Proceedings, Government Statistics Section. Miami Beach, Florida.

National Center for Health Statistics. 2013 National Health Interview Survey, Public-use data file and documentation. http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm.

Parsons VL, Moriarity C, Jonas K, Moore TF, Davis KE and Tompkins L. 2014 Design and estimation for the National Health Interview Survey, 2006–2015. National Center for Health Statistics. Vital and Health Statistics 2(165): 1–53.

Sakshaug JW and Kreuter F. 2012 Assessing the magnitude of non-consent biases in linked survey and administrative data. Survey Research Methods 6(2): 113–122.

Sakshaug JW, Couper MP, Ofstedal MB and Weir DR. 2012 Linking survey and administrative records mechanisms of consent. Sociological Methods & Research 41(4): 535–569. [PubMed: 27375305]

Sala E, Burton J and Knies G. 2012 Correlates of obtaining informed consent to data Linkage respondent, interview, and interviewer characteristics. Sociological Methods & Research 41(3): 414–439.

Schenker N, Raghunathan TE, Chiu P, Makuc DM and Cohen AJ. 2009 Multiple imputation of family income and personal earnings in the National Health Interview Survey: methods and examples. Journal of the American Statistical Association 101: 924–933.

SUDAAN Release 10.0 [computer software]. 2008 Research Triangle Park, NC.

Young AF, Dobson AJ and Byles JE. 2001 Health services research using linked records: who consents and what is the gain? Australian and New Zealand Journal of Public Health 25(5): 417–420. [PubMed: 11688619]

Table 1

Percent distribution of demographic and health characteristics and percent refusing linkage among adults aged 18 years and over; NHIS, 2010–2013.

	Sample Size unweighted	Percent distribution		Refused linkage		p-Value
	No.	Weighted %	95% CI	Weighted %	95% CI	
Overall	129,253	100		10.7	10.3, 11.0	
Sex						
Male	57,510	48.3	47.9, 48.6	10.4	10.0, 10.9	0.04
Female	71,743	51.7	51.4, 52.1	10.9	10.5, 11.3	
Age group (years)						
18–44	58,170	47.6	47.1, 48.1	9.9	9.5, 10.3	<0.01 [c]
45–64	43,617	34.9	34.5, 35.3	11.1	10.6, 11.6	
65 and over	27,466	17.5	17.2, 17.9	12.0	11.4, 12.6	
Race/ethnicity						
Hispanic	22,827	14.5	14.0, 15.0	13.4	12.8, 14.1	<0.01
Non-Hispanic White	76,944	67.7	67.0, 68.3	10.0	9.6, 10.4	
Non-Hispanic Black	20,129	11.9	11.5, 12.4	9.9	9.2, 10.6	
All other races and ethnicities	9,353	5.9	5.6, 6.2	13.0	12.0, 14.1	
Poverty index ratio [a]						
Less than 100%	23,458	13.7	13.3, 14.2	9.9	9.3, 10.5	<0.01 [c]
100–199%	27,519	19.2	18.8, 19.6	10.6	10.0, 11.2	
200–399%	37,790	30.2	29.7, 30.6	10.8	10.2, 11.3	
400% or more	40,486	37.0	36.2, 37.7	10.9	10.5, 11.4	
Income imputation status						
Provided income	106,645	81.2	80.8, 81.6	8.6	8.2, 8.9	<0.01
Income reported in categories	17,873	15.0	14.7, 15.3	17.5	16.7, 18.4	
Did not provide income	4,735	3.8	3.6, 4.0	29.0	27.3, 30.8	
Serious psychological distress [b]						
Yes	4,862	3.4	3.3, 3.5	7.1	6.2, 8.1	<0.01
No	122,366	96.6	96.5, 96.8	10.8	10.4, 11.1	

Chronic obstructive pulmonary disease [b]

	Sample Size unweighted	Percent distribution		Refused linkage		p-Value
	No.	Weighted %	95% CI	Weighted %	95% CI	
Yes	7,130	5.1	5.0, 5.3	7.9	7.1, 8.9	<0.01
No	121,973	94.9	94.7, 95.0	10.8	10.4, 11.2	
Diabetes[b]						
Yes	13,131	9.1	8.9, 9.3	9.0	8.3, 9.7	<0.01
No	116,031	90.9	90.7, 91.1	10.8	10.5, 11.2	
Heart disease[b]						
Yes	16,031	11.5	11.3, 11.7	9.0	8.4, 9.7	<0.01
No	113,022	88.5	88.3, 88.7	10.9	10.5, 11.2	
Stroke[b]						
Yes	4,081	2.7	2.6, 2.8	8.3	7.3, 9.5	<0.01
No	125,051	97.3	97.2, 97.4	10.7	10.4, 11.1	
Hypertension[b]						
Yes	36,271	25.6	25.2, 26.0	9.5	9.1, 10.0	<0.01
No	92,729	74.4	74.0, 74.8	11.0	10.7, 11.4	
Cancer[b]						
Yes	8,340	5.9	5.9, 6.3	9.6	8.8, 10.5	<0.01
No	120,807	94.1	93.7, 94.1	10.7	10.4, 11.1	
Respondent-reported health[b]						
Excellent	33,648	28.3	27.9, 28.7	11.2	10.6, 11.7	<0.01[c]
Very good	40,238	32.0	31.7, 32.4	10.5	10.1, 11.0	
Good	35,570	26.6	26.2, 26.9	10.8	10.3, 11.4	
Fair	15,003	9.9	9.7, 10.2	9.9	9.3, 10.6	
Poor	4,728	3.2	3.1, 3.3	8.6	7.6, 9.7	
Total number of health conditions[b]						
0	74,397	61.4	61.0, 61.9	11.3	10.9, 11.8	<0.01[c]
1	29,236	22.4	22.2, 22.7	10.0	9.5, 10.5	
2 or more	23,086	16.2	15.9, 16.5	8.7	8.2, 9.3	

CI, confidence interval.

[a] PIR is defined as ratio of family income to a federal poverty standard.

[b] Unweighted sample sizes for health outcomes do not sum to the total due to missing data.

[c] p-Value for test of trend.

Table 2

Associations between health characteristics and record-linkage refusal from multivariable logistic models: NHIS 2010–2013.

Health characteristic included separately in adjusted models[a]	Adjusted odds ratio[b]	95 percent confidence interval
Serious psychological distress	0.69	0.60, 0.79
Chronic obstructive pulmonary disease	0.73	0.65, 0.83
Diabetes	0.75	0.69, 0.82
Heart disease	0.79	0.73, 0.85
Stroke	0.71	0.61, 0.82
Hypertension	0.79	0.74, 0.83
Cancer	0.84	0.76, 0.93
Respondent-assessed health status		
Excellent	Reference	
Very good	0.94	0.88, 1.00
Good	0.93	0.87, 0.99
Fair	0.83	0.76, 0.90
Poor	0.72	0.63, 0.83
Total number of health conditions		
0	Reference	
1	0.80	0.75, 0.85
2 or more	0.65	0.61, 0.71

[a] The reference level for a health condition was the absence of a self-reported diagnosis for that health condition.

[b] Record-linkage refusal was the binary outcome. Independent variables included age, race/ethnicity, sex, poverty index ratio (PIR), and income imputation status. Chronic health conditions, health group, or total number of chronic health conditions were included as independent variables in separate models.