

## Article

# Respondent differences and length of data collection in the Behavioral Risk Factor Surveillance System

by Mohamed G. Qayad, Pranesh Chowdhury,  
Shaohua Hu and Lina Balluz



December 2010

# Respondent differences and length of data collection in the Behavioral Risk Factor Surveillance System

Mohamed G. Qayad, Pranesh Chowdhury, Shaohua Hu and Lina Balluz<sup>1</sup>

## Abstract

The current economic downturn in the US could challenge costly strategies in survey operations. In the Behavioral Risk Factor Surveillance System (BRFSS), ending the monthly data collection at 31 days could be a less costly alternative. However, this could potentially exclude a portion of interviews completed after 31 days (late responders) whose respondent characteristics could be different in many respects from those who completed the survey within 31 days (early responders). We examined whether there are differences between the early and late responders in demographics, health-care coverage, general health status, health risk behaviors, and chronic disease conditions or illnesses. We used 2007 BRFSS data, where a representative sample of the noninstitutionalized adult U.S. population was selected using a random digit dialing method. Late responders were significantly more likely to be male; to report race/ethnicity as Hispanic; to have annual income higher than \$50,000; to be younger than 45 years of age; to have less than high school education; to have health-care coverage; to be significantly more likely to report good health; and to be significantly less likely to report hypertension, diabetes, or being obese. The observed differences between early and late responders on survey estimates may hardly influence national and state-level estimates. As the proportion of late responders may increase in the future, its impact on surveillance estimates should be examined before excluding from the analysis. Analysis on late responders only should combine several years of data to produce reliable estimates.

Key Words: BRFSS; Responders; Differences; Length of data collection.

## 1. Introduction

The Behavioral Risk Factor Surveillance System (BRFSS) is a state-based household telephone survey in the United States (U.S.) and its territories which monitors health risk behaviors and chronic disease conditions for the adult noninstitutionalized population (Centers for Disease Control and Prevention [CDC] 2009a, BRFSS Turning Information into Public Health, <http://www.cdc.gov/brfss/about.htm>). It is the largest telephone survey in the world and is implemented by the 50 states, the District of Columbia, and U.S. territories, in collaboration with the CDC. The survey is conducted continuously throughout the year.

CDC dispenses the samples (phone numbers) to states quarterly. At the state level, the samples are divided into 12 monthly lists for operational purposes. Trained interviewers call each sampled telephone number. After each call to a sampled telephone number, a disposition code is assigned. States and their contractors are required to give final dispositions to their monthly released samples within that month. Over 90% of the monthly samples and completed interviews receive final dispositions within 31 days. States continue to complete their remaining samples afterwards (Qayad, Balluz and Garvin 2009).

Because of economic downturns, states and survey organizations may face budget cuts that could adversely affect their survey operations. Such unforeseen circumstances warrant searching for alternative operational strategies. A

cost-effective alternative could be to end data collection at the end of each month. However, ending data collection within one month excludes interviews completed after 31 days. Such exclusion could influence the variability of the respondents, surveillance estimates and the size of completed interviews, which could affect other operational decisions. Currently, the size of late responders is small and may not influence surveillance estimates. However, the current trend in survey responses heralds a continuous decline in survey responders, which could prolong the duration to reach respondent and the eventual increase in the proportion of late responders. Such circumstances require thorough examination of the influence of late responders on surveillance estimates in the future. This study examines whether respondents who completed the interviews within 31 days and those who completed after 31 days are different in demographics, risk behaviours, and chronic disease conditions.

## 2. Methods

We used the 2007 BRFSS data, which is an ongoing state-based random digit dialing (RDD) telephone survey among the non-institutionalized civilian population in the US. We divided the duration of the interview into two periods, 0-31 days and >31 days. Respondents who completed the interviews within 31 days (referred as early responders) and those completed after 31 days (referred as late responders).

1. Mohamed G. Qayad, Pranesh Chowdhury, Shaohua Hu and Lina Balluz, Division of Adult and Community Health, Behavioral Surveillance Branch, Centers for Disease Control and Prevention, Atlanta, GA 30341, U.S.A. E-mail: [maq3@cdc.gov](mailto:maq3@cdc.gov).

Demographic factors included were - gender, race, income and age. Race had four groups - white non-Hispanic, Black non-Hispanic, Hispanic and other race. Education had three levels: not a high school graduate, high school graduate, and more than high school education. Income categories were <\$15,000, \$15,000 - \$34,999, \$35,000 - \$49,999 and \$50,000 or more. Age had the following categories: 18 - 24 years, 25 - 44 years, 45 - 64 years, and 65 or more years. Respondents <65 years old who did not have any health plan (including health insurance, prepaid plans such as HMOs, or government plans such as Medicare) were considered not to have health plan. General health was dichotomized into good health (excellent, very good, or good health) and fair or poor health.

Health risk behaviors included were - binge drinking, current smoking, (lack of) physical activity, and (insufficient) fruit and vegetable consumption. Binge drinking was defined as having five or more drinks for men and four or more drinks for women on at least one occasion during the preceding month. Respondents who smoked  $\geq 100$  cigarettes in their lifetime and smoked every day or some days were classified as current smokers. Physical activity had following categories - meet recommendations for physical activity, insufficient physical activity, and do not participate in physical activity. Respondents who consumed 5 or more servings of fruits and vegetables everyday were classified as meet recommendation for fruit and vegetable consumption.

Chronic conditions or illness included were Cerebro-cardio-vascular disease, hypertension, had high cholesterol, diabetes, asthma, and overweight or obesity. Respondents were considered to have myocardial infarction, or angina, or stroke or high blood pressure if they had ever been told by a doctor, nurse, or other health professional to have myocardial infarction or stroke or high blood pressure respectively. Respondents were classified as having high blood cholesterol if they had checked their blood cholesterol and was told by a health professional that their blood cholesterol was high. Respondents were classified as having diabetes if they had ever been told by a doctor that they had diabetes. Asthma was self reported and physician or health care professional diagnosed; it had three categories - current asthma, former asthma, and never asthma. Self-reported weight and height were used to calculate Body Mass Index (BMI) ( $BMI = \text{weight}[\text{kg}] / (\text{height}[\text{m}])^2$ ). Participants were classified as overweight if their BMI was  $\geq 25 \text{ kg/m}^2$  and were classified as obese if their BMI was  $\geq 30 \text{ kg/m}^2$ .

We estimated the percent differences between early and late responders by demographics, health behaviors and chronic health conditions or illness. We used SUDAAN and

SAS for the analysis (SAS Institute Inc., Cary, NC, USA 2004).

### 3. Results

In the 2007 BRFSS survey, there were 430,912 interviews completed in the U.S. We excluded 14,189 records from two states (Michigan and Louisiana) and 49 cases with missing information. We analyzed the remaining 416,674 respondents of which 394,427 (95%) were early responders, and 22,247 (5%) were late responders. We estimated weighted and unweighted percent differences between early and late responders. The absolute differences between the weighted and unweighted percentages in the variables examined ranged between 0.06% and 2.6%, except white non-Hispanics where the absolute difference was 7%. We presented the unweighted analysis for the purpose of this study.

Significant differences were observed between early and late responders in demographics, access to health-care coverage, and general health status variables (Table 1). Compared to early responders, late responders were significantly more likely to be male, to report race/ethnicity as Hispanic, to have annual income of  $\geq \$50,000$ , to be younger than 45 years of age, to have less than high school education, to have access to health-care coverage, and to report good health. The absolute value of these significant differences in the variables above ranged from 1.3% to 7.6%. The percentage of Unknowns in the health-care coverage variable was 21% for late responders and 30% for early responders. The difference between early and late responders remained significant, even when we assumed the Unknowns to have a similar percentage of access to health-care coverage to those with known status in each respondent group.

A significant difference between early and late responders was also observed in health risk behaviors (Table 2). Compared to early responders, late responders were significantly less likely to meet the recommended guidelines for physical activity and daily consumption of fruits and vegetables. The absolute value of these significant differences ranged from 1.7% to 3.1%. The differences between early and late responders remained significant even when the Unknowns were assumed to have a similar percentage to those of known status for both variables.

Table 3 shows the differences between early and late responders in chronic disease conditions or illnesses. Compared to early responders, late responders were significantly more likely to report high cholesterol, significantly less likely to report hypertension and diabetes, and were significantly less likely to be obese. The absolute value of these significant differences ranged from 1.8% to 5.8%.

**Table 1**  
**Percent differences between early responders and late responders by demographics, health-care coverage and general health, BRFSS 2007**

Demographics	Length of data collection		Difference (Early-late) %	P-Value
	Early responders* (N = 394,427) %	Late responders** (N = 22,247) %		
Gender				
Female	62.8	60.2	2.5	0.000
Male	37.3	39.8	-2.5	
Race				
White non-Hispanic	79.1	71.5	7.6	0.000
Black non-Hispanic	7.3	8.2	-0.9	0.168
Hispanic	7.1	13.5	-6.4	0.000
Others	5.5	5.8	-0.3	0.635
Unknown	1.0	1.0	0.0	0.977
Income				
<15,000	9.7	8.7	1.0	0.146
15-34,999	26.1	24.3	1.8	0.004
35-49,999	14.1	13.4	0.8	0.252
50,000+	36.6	39.7	-3.1	0.000
Unknown	13.5	14.0	-0.4	0.496
Age				
18-24	3.6	4.9	-1.3	0.025
25-44	25.7	33.3	-7.6	0.000
45-64	40.9	40.6	0.3	0.612
65+	29.0	20.2	8.8	0.000
Unknown	0.8	1.0	-0.1	0.827
Education Level				
<High School	10.3	12.3	-2.0	0.001
High School Graduate	30.6	28.7	1.9	0.001
> High School	58.8	58.2	0.6	0.177
Unknown	0.3	0.8	-0.5	0.264
Health care coverage (<65 years)				
Yes	59.3	65.4	-6.2	0.000
No	10.8	13.2	-2.5	
Unknown	30.0	21.4	8.6	
Health Status				
Good health	80.1	81.8	-1.7	0.000
Fair or poor health	19.4	17.6	1.8	
Unknown	0.5	0.6	-0.1	

\*Completed the survey within 31 days.

\*\*Completed the survey after 31 days.

**Table 2**  
**Percent differences between early responders and late responders by health risk behaviors, BRFSS 2007**

Risk factors	Length of data collection		Difference (Early-late) %	P-Value
	Early responders* (N = 394,427) %	Late responders** (N = 22,247) %		
Binge drinking				
Yes	11.1	11.8	-0.7	0.261
No	86.9	82.8	4.1	
Unknown	1.9	5.4	-3.4	
Smoking cigarettes				
Current smokers	18.3	17.5	0.9	0.182
Not a smoker	81.3	82.1	-0.8	
Unknown	0.4	0.5	0.0	
Physical activity recommendations				
Met recommended moderate/vigorous activity	43.4	41.8	1.7	0.000
Insufficient physical activity	35.4	31.8	3.6	
No physical activity	14.3	11.3	3.0	
Unknown	6.9	15.2	-8.3	
Fruit & vegetable consumption				
Consumed $\geq$ 5 times/day	25.0	21.9	3.1	0.000
Consumed < 5 times/day	73.0	69.7	3.3	
Unknown	2.0	8.5	-6.4	

\*Completed the survey within 31 days.

\*\*Completed the survey after 31 days.

**Table 3**  
**Percent differences between early responders and late responders by chronic conditions and illnesses, BRFSS 2007**

Diseases/chronic conditions	Length of data collection		Difference (Early-late) %	P-Value
	Early responders* (N = 394,427) %	Late responders** (N = 22,247) %		
Cerebral and CVD:				
Myocardial Infarction				
Yes	5.9	4.9	1.0	0.177
No	93.6	94.7	-1.1	
Unknown	0.5	0.4	0.1	
Angina				0.053
Yes	6.0	4.5	1.5	
No	93.1	94.7	-1.6	
Unknown	0.9	0.8	0.1	
Stroke				0.183
Yes	3.8	2.8	1.0	
No	95.9	97.0	-1.1	
Unknown	0.3	0.2	0.1	
Other illnesses/conditions:				
High cholesterol				0.000
Yes	57.0	60.8	-3.8	
No	42.3	38.4	3.8	
Unknown	0.8	0.8	0.0	
Hypertension				0.000
Yes	35.8	30.1	5.8	
No	64.0	69.8	-5.8	
Unknown	0.2	0.2	0.0	
Diabetes				0.010
Yes	11.2	9.4	1.8	
Yes-Pregnancy	0.9	1.2	-0.2	
No	86.4	88.2	-1.9	
Borderline	1.4	1.2	0.2	
Unknown	0.1	0.1	0.0	
Asthma				0.158
Current	8.7	7.7	1.0	
Former	3.8	4.0	-0.2	
Never	86.9	87.8	-0.8	
Unknown	0.6	0.6	0.1	
Overweight or Obese				0.000
Normal weight	34.5	35.5	-1.1	
Over weight	35.0	34.7	0.4	
Obese	26.0	23.6	2.4	
Unknown	4.5	6.2	-1.7	

\*Completed the survey within 31 days.

\*\*Completed the survey after 31 days.

#### 4. Discussion

Our study found significant differences between early and late responders in demographic factors, and in some of the health risk behaviors and chronic disease conditions or illnesses. This shows that the composition of the two groups of responders is different with respect to these attributes. The differences observed could be due to difficulty in reaching persons working long hours and being away from their residences.

The greater likelihood of earning high income, being Hispanic, being young (18-44 years), having health-care coverage, having less than high school education, and reporting good general health among late responders fits the described characteristics of working people and healthy

workers (Li and Sung 1999), (O'Neil 1979). This description is supported by their significantly lower likelihood of reporting hypertension, diabetes and obesity. But certain risk behaviors show a different profile among late responders. Late responders are less likely to meet recommended guidelines for moderate or vigorous physical activity and for daily consumption of fruits and vegetables, which may be related to late responders having long working hours and poor access to healthy foods.

The high income earners, who are mostly white non-Hispanics, and low income earners, who are mostly Hispanics and black non-Hispanics, may spend long hours in their working environments and less likely to be in their homes to receive survey calls (Voigt, Koepsell and Daling 2003). In addition, BRFSS data indicate that interviewers

make more calls on late responders, on average almost 3 times more than on early responders, which bears out the difficulty of reaching them during the 31-day survey period. The reasons for working long hours could be different in the two income groups. Hispanics, black non-Hispanics, and young age groups may have low-paying jobs and need to work long hours to make a living, while the high-income individuals may have jobs requiring them to remain at work after regular working hours.

Surveillance and epidemiological estimates based only on early or late responders should be scrutinized for possible biases prior to making any generalizations. The percentage of interviews completed after 31 days is currently small (5%) and excluding them from the analysis may have no influence on national and state level estimates. However, as the proportions of late responders are expected to increase in the future, the influence of late responders on these estimates could not be ignored (Diehr, Cain, Connell and Volinn 1990). In addition, states should examine the consequences of ending data collection at 31 days on their operations, performance indicators, data quality measures, cost-savings and other contractual agreements with their data collection contractors.

Our study has a few limitations. BRFSS uses RDD methodology to select telephone numbers, which is subject to coverage bias (Rao, Link, Battaglia, Frankel, Giambo, and Mokdad 2005; Frankel, Srinath, Hoaglin, Battaglia, Smith, Wright and Khare 2003). Information collected is self-reported and may be subject to recall bias in some risk behaviors and disease estimations (Troiano, Berrigan, Dodd, Masse, Tilert and McDowell 2008; CDC 2004). In addition, we excluded two states from our analysis (Michigan and Louisiana), and extrapolation of the findings to these states should be done cautiously.

Despite these limitations, this study shows that late responders are significantly different in many respects from early responders. As the proportion of late responders may increase in the future, the influence of late responders on surveillance estimates should be examined carefully.

### Acknowledgements

We would like to acknowledge State BRFSS coordinators and their contractors.

### References

- Centers for Disease Control and Prevention (2009a). BRFSS Turning Information into Public Health, URL <http://www.cdc.gov/brfss/about.htm>.
- Centers for Disease Control and Prevention (2004). Access to health-care and preventive services among Hispanics and non-Hispanics – United States, 2001-2002. *Morbidity and Mortality Weekly Report*, 53, 937-941.
- Diehr, P., Cain, K., Connell, F. and Volinn, E. (1990). What is too much variation? The null hypothesis in small-area analysis. *Health Services Research*, 24, 741-771.
- Frankel, M.R., Srinath, K.P., Hoaglin, D.C., Battaglia, M.P., Smith, P.J., Wright, R.A. and Khare, M. (2003). Adjustments for non-telephone bias in random-digit-dialing surveys. *Statistics in Medicine*, 22, 1611-1626.
- Li, C.Y., and Sung, F.C. (1999). A review of the healthy worker effect in occupational epidemiology. *Occupational Medicine*, 49, 225-229.
- O'Neil, M.J. (1979). Estimating the nonresponse bias due to refusals in the telephone surveys. *Public Opinion Quarterly*, 43, 218-232.
- Qayad, M.G., Balluz, L. and Garvin, W. (2009). Does continuing data collection beyond one month improve the completion and response rates in behavioral risk factor surveillance system survey? *Survey Practice Feb 2009*. URL <http://surveypractice.org/2009/02/>.
- Rao, R.S., Link, M.W., Battaglia, M.P., Frankel, M.R., Giambo, P. and Mokdad, A.H. (2005). Assessing representativeness in RDD surveys: coverage and non-response in the Behavioral Risk Factor Surveillance System. Minneapolis: Joint Statistical Meetings. URL <http://www.amstat.org/sections/SRMS/Proceedings/y2005/Files/JSM2005-000190.pdf>.
- SAS Institute Inc. (2004). SAS version 9.1, Cary, NC: SAS Institute, USA.
- Troiano, R.P., Berrigan, D., Dodd, K.W., Masse, L.C., Tilert, T. and McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise*, 40, 181-8.
- Voigt, L.F., Koepsell, T.D. and Daling, J.R. (2003). Characteristics of Telephone Survey Respondents According to Willingness to Participate. *American Journal of Epidemiology*, 157, 66-73.