

A comparison of depression prevalence estimates measured by the Patient Health Questionnaire with two administration modes: computer-assisted telephone interviewing versus computer-assisted personal interviewing

Chaoyang Li · Earl S. Ford · Guixiang Zhao · James Tsai · Lina S. Balluz

Received: 8 November 2010 / Revised: 29 March 2011 / Accepted: 3 April 2011 / Published online: 27 April 2011
© Swiss School of Public Health (outside the USA) 2011

Abstract

Objective To compare depression prevalence estimates measured by the 8-item Patient Health Questionnaire (PHQ-8) with two administration modes in two national surveys.

Methods Data on adults aged 18 years and older who participated in the 2006 Behavioral Risk Factor Surveillance System (BRFSS) ($n = 198,678$) and those who participated in the 2005–2006 National Health and Nutrition Examination Survey (NHANES) ($n = 4,800$) were analyzed.

Results The crude PHQ-8 depression prevalence estimate using the diagnostic algorithm was higher in BRFSS with computer-assisted telephone interviewing (CATI) (9.16%, SE 0.15) than in NHANES with computer-assisted personal interviewing (CAPI) (6.28%, SE 0.59) ($P < 0.001$). After

adjustment for demographic characteristics, the difference in the prevalence estimates remained (9.68% in BRFSS vs. 6.13% in NHANES, $P < 0.001$). Similar differences in the depression prevalence estimates using the PHQ-8 cutoff score ≥ 10 were detected (9.22% in BRFSS vs. 5.15% in NHANES, $P < 0.001$). Significant differences in the depression prevalence persisted in subgroups stratified by demographic characteristics and major health risk factors and outcomes.

Conclusions The PHQ-8 administered by CATI yielded about 3.5% higher depression prevalence estimate than that by CAPI.

Keywords Depression · Patient Health Questionnaire · Mode of administration · Computer-assisted telephone interviewing · Computer-assisted personal interviewing

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

C. Li (✉) · L. S. Balluz
Division of Behavioral Surveillance, Office of Surveillance, Epidemiology, and Laboratory Services, Centers for Disease Control and Prevention, 1600 Clifton Road, NE, MS E-97, Atlanta, GA 30333, USA
e-mail: cli@cdc.gov

E. S. Ford · G. Zhao
Division of Adult and Community Health,
National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA, USA

J. Tsai
Division of Blood Disorders, National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, GA, USA

Introduction

Prevalence estimates of depression in the US have varied widely, ranging from 3.8 to 27.3% in studies based on structured patient interviews, and from 11.5 to 60.7% in studies based on self-administered questionnaires (Anderson et al. 2001). A recent report, based on 2006 data from Behavioral Risk Factor Surveillance System (BRFSS), estimated the prevalence of the depressive disorder assessed by the 8-item Patient Health Questionnaire (PHQ-8) to be 9.1% in the general population and 16.6% among persons with diabetes (Kroenke et al. 2009; Li et al. 2008).

The identification of depression reflects the sensitivity and specificity of the diagnostic instrument. The Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders (SCID) is currently the gold standard but

must be administered by an experienced clinician or trained mental health professional and is not suitable for population surveys (First et al. 2002). The PHQ-8 used in BRFSS was derived from the 9-item Patient Health Questionnaire (PHQ-9) that queries the nine criteria for depressive disorders defined in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) (American Psychiatric Association 1994). The validity of the PHQ-9 to establish provisional diagnoses of major and minor depression and evaluate the severity of depressive symptoms (Kroenke and Spitzer 2002) has been established in the general populations (Martin et al. 2006) and in diverse ethnic groups (Huang et al. 2006). The PHQ-9 diagnoses of major depression have 73% sensitivity and 98% specificity compared with independent diagnoses made by mental health professionals using the SCID (Spitzer et al. 1999). The PHQ-8 does not include queries as to suicidality and has similar validity for major depression as the PHQ-9 (Kroenke and Spitzer 2002). The PHQ-8 internal consistency reliability coefficient from previous studies is 0.82 (Pressler et al. 2010).

In addition to the diagnostic instrument, modes of questionnaire administration in surveys impact both data quality and prevalence estimates, particularly for sensitive topics and stigmatized behaviors (Tourangeau and Smith 1996; Bowling 2005; Hanmer et al. 2007; Salomon et al. 2009; Smith et al. 2009). A validation study conducted among 346 participants showed that the PHQ-9 scores were highly correlated ($r = 0.92$) between the self-administered and telephone-administered questionnaire (Pinto-Meza et al. 2005). Good agreement also exists between face-to-face interview and telephone interview scores for mental health assessed by the General Health Questionnaire (intra-class correlation coefficient $r = 0.83$) and Revised Clinical Interview Schedule (intra-class correlation coefficient $r = 0.92$) (Evans et al. 2004). We are unaware of studies comparing identification of depression by the PHQ-8 administered by computer-assisted telephone interviewing (CATI) and computer-assisted personal interviewing (CAPI). Thus, we conducted this study to compare the prevalence estimates of the PHQ-8 depression between the CATI, as administered in the 2006 BRFSS, and CAPI, as administered in the 2005–2006 National Health and Nutrition Examination Survey (NHANES).

Methods

Study design and population

The BRFSS uses a random-digit dialing (RDD) method to obtain a state-representative sample of civilian, non-institutionalized population aged ≥ 18 years in all 50 states, District of Columbia, and territories. The BRFSS uses

disproportionate stratified random sampling (DSS) method to obtain a probability sample within each state. Information obtained from the sample can be used to generalize results to the total population in the state as well as to the nation as a whole. With DSS, telephone numbers are drawn from two strata that are based on the presumed density of known household telephone numbers. In this design, telephone numbers are classified into strata that are either high density or medium density to yield residential telephone numbers. Telephone numbers in the high density stratum are sampled at the highest rate. The 2006 BRFSS used CATI to assess key behavioral risk factors and chronic conditions among adults [Centers for Disease Control and Prevention (CDC) 2007].

The NHANES 2005–2006 is a complex, multistage probability sample of the civilian non-institutionalized population of the US. The NHANES 2005–2006 uses a four-stage sample design: (1) primary sampling units (PSUs) comprising mostly single counties, (2) area segments within PSUs, (3) households within area segments, and (4) persons within households (Plan and operation of the Third National Health and Nutrition Examination Survey 1994; Centers for Disease Control and Prevention 2007). NHANES combines in-home face-to-face CAPI and physical examinations at the mobile examination center to collect data on individuals. In this study, we restricted our analyses in the NHANES 2005–2006 sample to adults aged ≥ 18 years.

Measurement of depression

The same PHQ-8 was administered to participants aged ≥ 18 years in the 2006 BRFSS by CATI and to participants aged ≥ 18 years in the 2005–2006 NHANES by CAPI conducted by trained lay interviewers. Major depression was defined as having at least five of eight PHQ-8 criteria, one of which must be depressed mood or loss of interest or pleasure, for ≥ 7 days in the past 2 weeks. Minor depression was defined as having two to four of the eight PHQ-8 criteria, one of which must be depressed mood or loss of interest or pleasure, for ≥ 7 days in the past 2 weeks (American Psychiatric Association 1994; Kroenke and Spitzer 2002). The PHQ-8 depression included both major and minor depression. Alternatively, a severity score of 0–3 was assigned to each item (0 = 0–1 day, 1 = 2–6 days, 2 = 7–11 days, and 3 = 12–14 days), yielding a total score between 0 and 24 points. A total score of 0–4, 5–9, 10–14, 15–19, and 20–24 represents no, mild, moderate, moderately severe, and severe depressive symptoms, respectively. A total score ≥ 10 has been recommended as a cutoff point for screening depression (Kroenke and Spitzer 2002). Because using either a combination of major and minor depression or a total of PHQ-8 score ≥ 10 to define depression has yielded similar prevalence estimates (Li

et al. 2008), we used the former definition in our current analysis.

Assessment of demographic characteristics, lifestyle habits, and chronic illnesses

In both 2006 BRFSS and 2005–2006 NHANES, all covariates were assessed by participants' self-report using similar instruments. Demographic characteristics included sex (men vs. women), age (18–24, 25–44, 45–64, ≥ 65 years), race/ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, other), education attainment (<high school, high school/some college, \geq college), and marital status (married, previously married, never married). Smoking status was determined as current smokers (have smoked at least 100 cigarettes during their entire life and smoked at the interview), former smokers (have smoked at least 100 cigarettes during their entire life but did not smoke at the interview), and persons who never smoked. Body mass index [BMI = weight (kg)/height² (m)] was calculated by using self-reported weight in kilograms and height in meters. Diabetes, myocardial infarction, coronary heart disease (CHD), stroke, and asthma were determined by participants' self-report. Participants' general health status was categorized as being fair or poor versus excellent, very good, or good. Disability was determined by asking participants, "Do you now have any health problem that requires you to use special equipment, such as a cane, a wheelchair, a special bed, or a special telephone?" in both surveys. Persons who answered this question affirmatively were considered to have a disability.

Statistical analysis

We compared sample characteristics between the 2006 BRFSS and the 2005–2006 NHANES. We estimated standardized internal consistency reliability coefficient of the PHQ-8 in the two surveys separately and estimated the crude and age-standardized prevalence of depression by different diagnostic criteria, severity category, and each single item. We also assessed the differences in the prevalence of the PHQ-8 depression in each subgroup stratified by demographic characteristics, lifestyle habits, and chronic illnesses with adjustment for sex, age, race/ethnicity, education attainment, and marital status.

To control for possible differences in the distribution of demographic characteristics between the two surveys, we estimated adjusted prevalence of PHQ depression using a predicted marginal approach in log-linear regression analyses. Predictive marginals have been proposed as a direct standardization method controlling for differences in the distribution of covariates (Graubard and Korn 1999). PHQ depression was dichotomized as 1 = yes and 0 = no and

modeled as a dependent variable in the log-linear regression. Depression prevalence between the two surveys was estimated by controlling possible variations in sex, age, race/ethnicity, education attainment, and marital status. The difference in the prevalence between the two surveys was tested with a two-sample *t* test. We performed data management and variable recoding using SAS (version 9.2) and conducted all analyses using SUDAAN software (Release 9.0; Research Triangle Institute, Research Triangle Park, NC). In all analyses, sample weights were used to account for the varying probabilities of complex sampling design and non-response. We considered results with a two-tailed *P* value <0.05 to be statistically significant.

Results

Sample characteristics 2005 BRFSS and 2005–2006 NHANES surveys

A total of 198,678 participants in the 2006 BRFSS and 4,800 participants in the 2005–2006 NHANES provided PHQ-8 data. Sex and age distributions were comparable in 2006 BRFSS and the 2005–2006 NHANES surveys (Table 1). However, distributions of race/ethnicity, education, marital status differed significantly between the two samples (all $P < 0.001$). The prevalence of current smoking and self-reported total obesity (BMI ≥ 30 kg/m²) were lower in the 2006 BRFSS than in the 2005–2006 NHANES (both $P < 0.001$), whereas the prevalence of CHD and asthma were higher in the 2006 BRFSS than in the 2005–2006 NHANES (both $P < 0.01$). There were no significant differences in the prevalence of diabetes, MI, stroke, disability, and poor or fair health between the two surveys.

Prevalence of depression

The standardized internal consistency reliability coefficient of PHQ-8 was 0.85 in BRFSS and 0.83 in NHANES. The crude prevalence of the PHQ-8 score ≥ 10 ($P < 0.001$), depression ($P < 0.001$), major depression ($P < 0.001$), and minor depression ($P = 0.04$) was higher in BRFSS than NHANES (Fig. 1). We also observed a significant difference in the crude proportion of depression severity between the two surveys ($P < 0.001$) (Fig. 2). The crude prevalence of each PHQ-8 single item was higher in BRFSS than NHANES, differences ranging from 1.4 to 5.7% (Table 2). The age-standardized prevalence estimates of PHQ depression and each PHQ-8 single item were nearly identical to their crude estimates (data not shown).

After adjustment for sex, age, race/ethnicity, education attainment, and marital status, the prevalence of PHQ-8 depression was about 3.5% higher in absolute value in

Table 1 Sample characteristics between BRFSS (2006) and NHANES (2005–2006), US

Characteristics	BRFSS (<i>N</i> = 198,678)			NHANES (<i>N</i> = 4,800)			<i>P</i> value*
	<i>N</i>	%	SE	<i>N</i>	%	SE	
Sex							0.73
Men	76,288	48.7	0.3	2,312	48.5	0.6	
Women	122,390	51.3	0.3	2,488	51.5	0.6	
Age (years)							0.78
18–24	9,186	13.7	0.3	967	12.9	0.9	
25–44	59,403	38.4	0.2	1,576	37.3	1.8	
45–64	81,556	32.6	0.2	1,284	33.8	1.2	
≥65	47,161	15.4	0.1	973	16.0	1.4	
Race/ethnicity							<0.001
Non-Hispanic white	153,642	68.5	0.3	2,311	72.8	2.7	
Non-Hispanic black	15,819	8.8	0.1	1,143	11.1	1.9	
Hispanic	15,602	16.3	0.2	1,162	11.1	1.4	
Other	11,955	6.4	0.1	184	5.0	0.6	
Education							<0.001
<High school	18,776	11.8	0.2	1,149	16.9	1.5	
High school/some college	112,185	55.0	0.3	2,260	56.6	1.5	
≥College	67,447	33.1	0.2	874	26.5	2.2	
Marital status							<0.001
Married	112,779	60.7	0.3	2,384	56.2	1.5	
Previously married	54,975	16.9	0.2	914	18.1	0.8	
Never married	30,282	22.4	0.3	1,500	25.7	1.0	
Smoking status							<0.001
Current smoker	36,807	19.3	0.2	941	24.1	1.3	
Former smoker	55,293	24.0	0.2	1,114	25.3	1.1	
Never smoked	105,927	56.7	0.2	2,229	50.7	1.4	
BMI (kg/m ²)							<0.001
<25	71,450	38.5	0.3	1,732	36.5	1.3	
25–29	69,836	36.2	0.2	1,526	32.5	0.8	
≥30	49,620	25.3	0.2	1,425	31.0	1.6	
Diabetes							0.32
Yes	19,107	7.8	0.1	431	7.3	0.5	
No	179,409	92.2	0.1	4,364	92.7	0.5	
MI							0.05
Yes	10,464	4.2	0.1	172	3.3	0.4	
No	187,506	95.8	0.1	4,107	96.7	0.4	
CHD							0.006
Yes	11,288	4.3	0.1	167	3.3	0.4	
No	186,021	95.7	0.1	4,100	96.7	0.4	
Stroke							0.65
Yes	6,697	2.6	0.1	156	2.7	0.3	
No	191,552	97.4	0.1	4,124	97.3	0.3	
Asthma							<0.001
Yes	25,719	12.8	0.2	377	8.7	0.6	
No	172,520	87.2	0.2	4,152	91.3	0.6	
Disability ^a							0.55
Yes	16,611	6.6	0.1	347	6.3	0.5	
No	181,976	93.4	0.1	3,938	93.7	0.5	

Table 1 continued

Characteristics	BRFSS (<i>N</i> = 198,678)			NHANES (<i>N</i> = 4,800)			<i>P</i> value*
	<i>N</i>	%	SE	<i>N</i>	%	SE	
General health							
Fair/poor	34,001	15.5	0.2	981	15.9	0.7	0.58
Excellent/very good/good	164,162	84.5	0.2	3,819	84.1	0.7	

BMI body mass index, *BRFSS* Behavioral Risk Factor Surveillance System, *CHD* coronary heart disease, *MI* myocardial infarction, *NHANES* National Health and Nutrition Examination Survey, *SE* standard error

* A *t* test was used to test the equality of percentages between BRFSS and NHANES

^a Defined as currently having any health problem that requires a person to use special equipment, such as a cane, a wheelchair, a special bed, or a special telephone

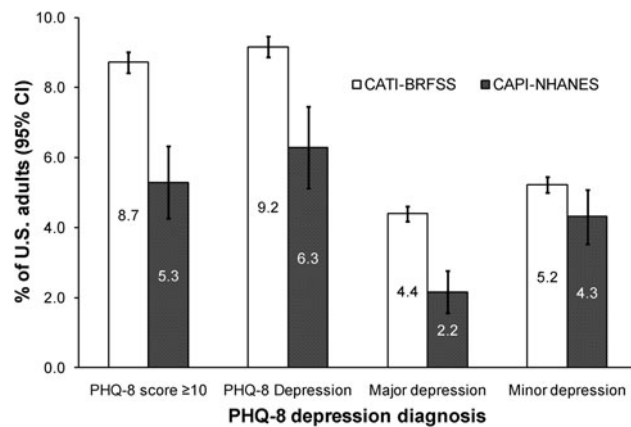


Fig. 1 Crude Prevalence of Depression Diagnosis Assessed by the PHQ-8 with CATI in the 2006 BRFSS and CAPI in the 2005–2006 NHANES. *PHQ* patient health questionnaire, *CATI* computer-assisted telephone interviewing, *CAPI* computer-assisted personal interviewing

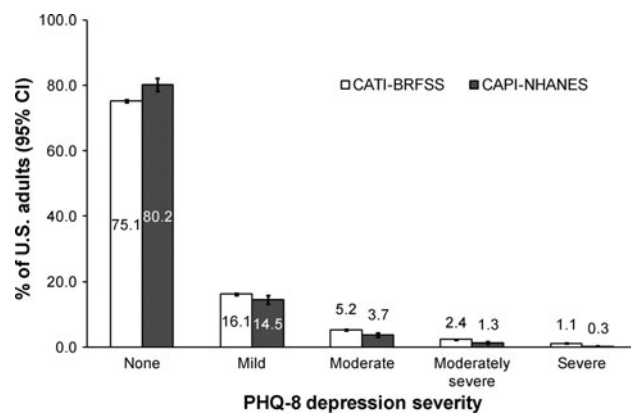


Fig. 2 Crude Proportion of Depression Severity Assessed by the PHQ-8 with CATI in the 2006 BRFSS and CAPI in the 2005–2006 NHANES. A total PHQ-8 score 0–4, 5–9, 10–14, 15–19, and 20–24 represents no, mild, moderate, moderately severe, and severe depressive symptoms, respectively. *PHQ* patient health questionnaire, *CATI* computer-assisted telephone interviewing, *CAPI* computer-assisted personal interviewing

BRFSS (9.7%) than that NHANES (6.1%) ($P < 0.001$) (Table 3). The difference in the prevalence of depression between the two surveys ranged from 1.5 to 7.0% in the subgroups stratified by demographic characteristics, lifestyle habits, and chronic illness status. We observed significant differences in the prevalence in each subgroup, except for participants with stroke or asthma.

Discussion

Our results indicated that the overall prevalence estimates of the PHQ-8 depression were approximately 3.5% higher with the administration of CATI in BRFSS (9.7%) than that with the administration of CAPI in NHANES (6.1%) after adjustment for demographic characteristics. The difference in the adjusted prevalence estimates persisted in most of the subgroups stratified by demographic characteristics, lifestyle habits, and chronic illness status.

It is surprising to observe the significant differences in the prevalence estimates of the PHQ-8 depression between the two modes administered in these two national surveys. Our results were in contrast to the previous findings that showed similar results of mental health scores between telephone interviews and self-administered assessments or face-to-face interviews in two validation studies (Evans et al. 2004; Pinto-Meza et al. 2005). Our study was not intended to validate the results of the PHQ-8 between CATI and CAPI because the two modes were not applied within the same survey. Rather, we attempted to compare the population prevalence estimates of the PHQ-8 depression in two surveys that targeted the same population with two different modes.

There are several limitations to this study. First, the PHQ-8 assessment in BRFSS was conducted among participants selected in 38 states, whereas the PHQ-8 assessment in NHANES was conducted among participants of a representative sample of the US population. Differences in the geographic coverage between the two surveys

Table 2 Crude prevalence of the PHQ-8 single item in BRFSS (2006) and NHANES (2005–2006), US

PHQ-8 items	CATI-BRFSS		CAPI-NHANES		Diff. (%)	P value*
	%	SE	%	SE		
PHQ 1 (little interest or pleasure in doing things)	9.2	0.2	5.3	0.6	3.9	<0.001
PHQ 2 (feeling down, depressed or hopeless)	6.5	0.1	5.1	0.5	1.4	0.03
PHQ 3 (trouble falling or staying asleep, or sleeping too much)	17.1	0.2	11.7	0.7	5.4	<0.001
PHQ 4 (feeling tired or having little energy)	19.3	0.2	13.6	0.8	5.7	<0.001
PHQ 5 (poor appetite or overeating)	11.3	0.2	6.3	0.6	5.0	<0.001
PHQ 6 (feeling bad about yourself, or that you are a failure or have let yourself or your family down)	5.5	0.1	3.2	0.4	2.3	<0.001
PHQ 7 (trouble concentrating on things, such as reading the newspaper or watching television)	5.9	0.1	4.2	0.5	1.7	0.003
PHQ 8 (moving or speaking so slowly that other people could have noticed, or the opposite, being so fidgety or restless that you have been moving around a lot more than usual)	3.6	0.1	2.2	0.3	1.4	0.001

BRFSS Behavioral Risk Factor Surveillance System, CAPI computer-assisted personal interviewing, CATI computer-assisted telephone interviewing, NHANES National Health and Nutrition Examination Survey, PHQ patient health questionnaire, SE standard error

* A *t* test was used to test the equality of prevalence rates between CATI-BRFSS and CAPI-NHANES

may also lead to the small variations of prevalence estimates. Because the PHQ-8 assessment in the BRFSS was conducted among participants from about 76% of all 50 states in the US, exclusion of participants in the rest of 12 states could have minimally influenced the prevalence estimates.

Second, the PHQ-8 assessment in BRFSS was conducted in 2006, whereas the PHQ-8 assessment in NHANES was conducted during 2005 and 2006; both assessments were done for the first time. No data on PHQ-8 were available in prior BRFSS and NHANES surveys. The slightly different time frame may lead to bias for the comparison of the prevalence estimates between the two surveys. However, the difference in the prevalences between two time periods could be minimal because we found similar prevalence estimates of the PHQ-8 depression in the 2006 BRFSS survey and in the 2008 BRFSS survey (data not shown) in the eight states that had the PHQ-8 assessment in both years.

Third, because the PHQ-8 assessment in BRFSS was conducted among participants who had land phones at home, people with advanced age, disability, or severe chronic illness were likely to be included in the survey, whereas people in households without landline telephones and those who were hospitalized or institutionalized were excluded from the survey. In contrast, the PHQ-8 assessment in NHANES was conducted among participants who attended the Mobile Examination Center, and people with advanced age, disability, or severe chronic illness may have been less likely to have attended. However, their impact on our results could be minimal because our

subgroup analyses stratified by age, disability, and chronic illness showed persistent differences in the prevalence estimates of depression between the two surveys.

Fourth, variations in the response rates between BRFSS and NHANES may have possibly contributed to the differences in the PHQ-8 depression prevalence estimates. The median response rate was 51.4%, ranging from 35.1 to 66.0% at the state level in the 2006 BRFSS survey, whereas the response rate was 77.4% in the 2005–2006 NHANES survey. A low response rate in the BRFSS may have increased non-response bias in the PHQ-8 depression prevalence estimates. However, recent research has shown that the standard RDD telephone survey with a response rate of 36.0% and the rigorous RDD survey with a response rate of 60.6% yielded similar results (Keeter et al. 2000).

Fifth, different sampling methods used in BRFSS and NHANES may be partially responsible for the observed differences in the prevalence estimates of depression between the two surveys. Indeed, we observed some differences in the proportions of race/ethnicity, education attainment, and marital status and the prevalences of several health risk factors and chronic illnesses between the two surveys. However, the impact of sampling methods on our results could be minimal for the following two reasons: first, sample weights were used in all analyses to account for the varying probabilities of complex sampling design and non-response to generate unbiased estimates to the US population in both surveys. Second, we adjusted for the possible variations in demographic characteristics between the two surveys for the comparisons in the prevalence estimates of depression between the two surveys.

Table 3 Adjusted prevalence of the PHQ-8 depression in BRFSS (2006) and NHANES (2005–2006), US

Stratification variable	CATI-BRFSS (<i>N</i> = 198,678)			CAPI-NHANES (<i>N</i> = 4,800)			Diff. (%)	<i>P</i> value [†]
	<i>N</i>	% ^a	SE	<i>N</i>	%	SE		
Total	198,678	9.7	0.3	4,800	6.1	0.5	3.5	<0.001
Sex								
Men	76,288	8.4	0.3	2,312	4.9	0.5	3.5	<0.001
Women	122,390	10.9	0.3	2,488	7.3	0.7	3.6	<0.001
Age (years)								
18–24	9,186	10.8	0.7	967	5.5	0.9	5.3	<0.001
25–44	59,403	9.7	0.4	1,576	5.3	0.9	4.4	<0.001
45–64	81,556	10.2	0.3	1,284	7.8	0.9	2.5	<0.001
≥65	47,161	7.3	0.3	973	5.0	0.8	2.3	<0.001
Race/ethnicity								
Non-Hispanic white	153,642	8.5	0.2	2,311	5.3	0.6	3.3	<0.001
Non-Hispanic black	15,819	13.4	0.7	1,143	10.5	1.3	3.0	0.02
Hispanic	15,602	11.9	0.6	1,162	7.3	1.3	4.6	0.001
Other	11,955	11.9	0.8	184	6.2	1.8	5.8	0.003
Education								
<High school	18,776	17.2	0.8	1,149	11.0	1.1	6.2	<0.001
High school/some college	112,185	10.5	0.2	2,260	6.5	0.8	4.0	<0.001
≥College	67,447	4.4	0.2	874	2.9	0.5	1.5	0.003
Marital status								
Married	112,779	7.2	0.2	2,384	4.6	0.6	2.6	<0.001
Previously married	54,975	14.7	0.4	914	10.5	1.0	4.2	<0.001
Never married	30,282	12.4	0.6	1,500	6.7	0.9	5.7	<0.001
Smoking status								
Current smoker	36,807	17.8	0.7	941	10.8	1.0	7.0	<0.001
Former smoker	55,293	8.3	0.3	1,114	5.0	0.9	3.3	<0.001
Never smoked	105,927	7.2	0.2	2,229	4.7	0.6	2.5	<0.001
BMI (kg/m ²)								
<25	71,450	8.4	0.3	1,732	5.1	0.7	3.2	<0.001
25–29	69,836	8.5	0.3	1,526	5.5	0.7	2.9	<0.001
≥30	49,620	13.4	0.4	1,425	7.9	0.9	5.5	<0.001
Diabetes								
Yes	19,107	16.2	0.7	431	10.2	1.8	6.0	0.001
No	179,409	9.1	0.3	4,364	5.8	0.6	3.3	<0.001
MI								
Yes	10,464	17.4	1.1	172	10.8	2.3	6.6	0.009
No	187,506	9.2	0.3	4,107	6.0	0.6	3.3	<0.001
CHD								
Yes	11,288	16.3	1.1	167	10.2	2.5	6.2	0.02
No	186,021	9.2	0.3	4,100	6.0	0.5	3.3	<0.001
Stroke								
Yes	6,697	19.9	1.3	156	18.3	2.8	1.6	0.58
No	191,552	9.4	0.3	4,124	5.8	0.5	3.6	<0.001
Asthma								
Yes	25,719	16.2	0.6	377	11.1	3.1	5.1	0.10
No	172,520	8.7	0.2	4,152	5.5	0.5	3.1	<0.001

Table 3 continued

Stratification variable	CATI-BRFSS (<i>N</i> = 198,678)			CAPI-NHANES (<i>N</i> = 4,800)			Diff. (%)	<i>P</i> value [†]
	<i>N</i>	% ^a	SE	<i>N</i>	%	SE		
Disability ^b								
Yes	16,611	24.3	1.2	347	19.9	2.1	4.4	0.05
No	181,976	8.5	0.2	3,938	5.3	0.5	3.2	<0.001
General health								
Fair/poor	34,001	26.0	0.7	981	21.8	1.6	4.2	0.02
Excellent/very good/good	164,162	6.5	0.2	3,819	3.2	0.4	3.2	<0.001

BMI body mass index, *BRFSS* Behavioral Risk Factor Surveillance System, *CAPI* computer-assisted personal interviewing, *CATI* computer-assisted telephone interviewing, *CHD* coronary heart disease, *MI* myocardial infarction, *NHANES* National Health and Nutrition Examination Survey, *PHQ* patient health questionnaire, *SE* standard error

[†] A *t* test was used to test the equality of prevalence rates between CATI-BRFSS and CAPI-NHANES

^a Data were adjusted for sex, age, race/ethnicity, education attainment, and marital status

^b Defined as currently having any health problem that requires a person to use special equipment, such as a cane, a wheelchair, a special bed, or a special telephone

Moreover, as shown in a previous study (Nelson et al. 2003), BRFSS provided national estimates of major health measures comparable to those of the National Health Interview Survey, which uses similar sampling methods as the NHANES does.

Despite these limitations, the mode of administration (CATI vs. CAPI) for the PHQ-8 questionnaire may play a major role in the differences in the prevalence estimates of the PHQ-8 depression because BRFSS data have consistently been found to provide national estimates comparable to results from national household surveys with a face-to-face interview in health risk factors, chronic conditions, and health care access (Mokdad et al. 2003; Nelson et al. 2003). Numerous studies have shown that survey administration mode may have significant impact on the data quality and results (Bowling 2005; Newman et al. 2002), particularly for sensitive issues and stigmatized behaviors (Durant and Carey 2000; Smith et al. 2009; Tourangeau and Smith 1996). Studies with either CATI or self-administered surveys have been found to report a higher rate of stigmatized behaviors including sexual behaviors and drug use than those using face-to-face interviewing (Durant and Carey 2000; Smith et al. 2009; Tourangeau and Smith 1996). Similarly, studies have suggested that telephone administration has yielded more positive health-related quality of life scores than either the self-administered survey (Hanmer et al. 2007) or mail survey (Hays et al. 2009). Moreover, a recent study compared the four national surveys including BRFSS (with CATI), NHANES (with CAPI), National Health Interview Survey (NHIS with CAPI), and Current Population Survey (CPS with CATI/CAPI) and found that BRFSS yielded a higher prevalence of fair or poor self-rated general health than either

NHANES or NHIS, but a lower prevalence than CPS (Salomon et al. 2009).

Consistent with those findings, CATI as administered in the 2006 BRFSS yielded a higher prevalence estimate of the PHQ-8 depression than CAPI as administered in the 2005–2006 NHANES. Telephone interviewing may be more successful in eliciting frank responses to sensitive issues than face-to-face interviewing (Sykes and Collins 1988). Throughout history, stigmatization of people with mental illness has long been recognized as a public health issue (US Department of Health and Human Services 1999). Stigmatizing attitudes about mental illness in the general public and self-stigma and fear of rejection by others have a negative impact on people's lives (Corrigan and Watson 2002; Link et al. 1997). The negative impact may also be reflected in their responses to the PHQ-8 items in the surveys, particularly in the face-to-face interviewing mode. Alternatively, differences in sampling frames (i.e., telephone sampling frame in BRFSS versus household sampling frame in NHANES) between the two studies may have contributed to the difference.

In sum, by comparing the prevalence estimates of the PHQ-8 depression between the BRFSS and NHANES surveys that used two different modes of questionnaire administration, we identified significant variations in the prevalence estimates of depression between the two modes. Although our results are unable to provide reasons for the actual causes of the variations in the PHQ-8 depression prevalence estimates in the US adult population between CATI and CAPI, the notable differences in the prevalence estimates between the two administration modes warrant cautions when one compares and interprets the results from these two national surveys. In recognition of stigmatization

for mental illness in the general public, appropriate selection of administration modes for the PHQ-8 questionnaire needs to be scrutinized to elicit accurate information regarding depression.

Acknowledgments The authors would like to acknowledge Dr. William C Reeves of the CDC for his critical reviews and contribution to this manuscript.

Conflict of interest The authors declare that they have no competing interests.

References

- American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn. American Psychiatric Association, Washington
- Anderson RJ, Freedland KE, Clouse RE, Lustman PJ (2001) The prevalence of comorbid depression in adults with diabetes: a meta-analysis. *Diabetes Care* 24:1069–1078
- Bowling A (2005) Mode of questionnaire administration can have serious effects on data quality. *J Public Health (Oxf)* 27:281–291
- Centers for Disease Control and Prevention (12-18-2007) NHANES 2005–2006. http://www.cdc.gov/nchs/about/major/nhanes/nhanes2005-2006/nhanes05_06.htm
- Centers for Disease Control and Prevention (CDC) (2007) Behavioral Risk Factor Surveillance System technical information and 2006 survey data. http://www.cdc.gov/brfss/technical_infodata/surveydata/2006.htm
- Corrigan PW, Watson AC (2002) Understanding the impact of stigma on people with mental illness. *World Psychiatry* 1:16–20
- Durant LE, Carey MP (2000) Self-administered questionnaires versus face-to-face interviews in assessing sexual behavior in young women. *Arch Sex Behav* 29:309–322
- Evans M, Kessler D, Lewis G, Peters TJ, Sharp D (2004) Assessing mental health in primary care research using standardized scales: can it be carried out over the telephone? *Psychol Med* 34:157–162
- First MB, Spitzer RL, Gibbon M, Williams JBW (2002) Structured clinical interview for DSM-IV-TR axis I disorders, research version. Biometrics Research, New York State Psychiatric Institute, New York
- Graubard BI, Korn EL (1999) Predictive margins with survey data. *Biometrics* 55:652–659
- Hanmer J, Hays RD, Fryback DG (2007) Mode of administration is important in US national estimates of health-related quality of life. *Med Care* 45:1171–1179
- Hays RD, Kim S, Spritzer KL, Kaplan RM, Tally S, Feeny D et al (2009) Effects of mode and order of administration on generic health-related quality of life scores. *Value Health* 12:1035–1039
- Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL (2006) Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. *J Gen Intern Med* 21:547–552
- Keeter S, Miller C, Kohut A, Groves RM, Presser S (2000) Consequences of reducing nonresponse in a national telephone survey. *Public Opin Q* 64:125–148
- Kroenke K, Spitzer RL (2002) The PHQ-9: a new depression diagnostic and severity measure. *Psychiatr Ann* 32:1–7
- Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH (2009) The PHQ-8 as a measure of current depression in the general population. *J Affect Disord* 114:163–173
- Li C, Ford ES, Strine TW, Mokdad AH (2008) Prevalence of depression among U.S. adults with diabetes: findings from the 2006 behavioral risk factor surveillance system. *Diabetes Care* 31:105–107
- Link BG, Struening EL, Rahav M, Phelan JC, Nuttbrock L (1997) On stigma and its consequences: evidence from a longitudinal study of men with dual diagnoses of mental illness and substance abuse. *J Health Soc Behav* 38:177–190
- Martin A, Rief W, Klaiberg A, Braehler E (2006) Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. *Gen Hosp Psychiatry* 28:71–77
- Mokdad AH, Stroup DF, Giles WH (2003) Public health surveillance for behavioral risk factors in a changing environment. Recommendations from the Behavioral Risk Factor Surveillance Team. *MMWR Recomm Rep* 52:1–12
- Nelson DE, Powell-Griner E, Town M, Kovar MG (2003) A comparison of national estimates from the National Health Interview Survey and the Behavioral Risk Factor Surveillance System. *Am J Public Health* 93:1335–1341
- Newman JC, Des J, Turner CF, Gribble J, Cooley P, Paone D (2002) The differential effects of face-to-face and computer interview modes. *Am J Public Health* 92:294–297
- Pinto-Meza A, Serrano-Blanco A, Penarrubia MT, Blanco E, Haro JM (2005) Assessing depression in primary care with the PHQ-9: can it be carried out over the telephone? *J Gen Intern Med* 20:738–742
- Plan and operation of the Third National Health and Nutrition Examination Survey 1988–94 (1994) National Center for Health Statistics, Hyattsville. *Vital Health Stat* 1(32)
- Pressler SJ, Subramanian U, Perkins SM, Gradus-Pizlo I, Kareken D, Kim J et al (2010) Measuring depressive symptoms in heart failure: validity and reliability of the Patient Health Questionnaire-8. *Am J Crit Care*
- Salomon JA, Nordhagen S, Oza S, Murray CJ (2009) Are Americans feeling less healthy? The puzzle of trends in self-rated health. *Am J Epidemiol* 170:343–351
- Smith A, Lyons A, Pitts M, Croy S, Ryall R, Garland S et al (2009) Assessing knowledge of human papillomavirus and collecting data on sexual behavior: computer assisted telephone versus face to face interviews. *BMC Public Health* 9:429
- Spitzer RL, Kroenke K, Williams JB (1999) Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA* 282:1737–1744
- Sykes W, Collins M (1988) Effects of mode of interview: experiments in the UK. In: Groves RM, Biemer PP, Lyberg LE, Massey JT, Nicholls WLI, Waksberg J (eds) *Telephone survey methodology*. Wiley, New York, pp 301–320
- Tourangeau R, Smith TW (1996) Asking sensitive questions: the impact of data collection mode, question format, and question context. *Public Opin Q* 60:275–304
- US Department of Health and Human Services (1999) *Mental health: a report of the surgeon general*. US Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health, Rockville, MD