ELSEVIER

Journal of Clinical Epidemiology 56 (2003) 148–154

Journal of Clinical Epidemiology

Evaluating the quality of self-reports of hypertension and diabetes

Noreen Goldman[a,*], I-Fen Lin[b], Maxine Weinstein[c], Yu-Hsuan Lin[d]

[a]Office of Population Research, Princeton University, 243 Wallace Hall, Princeton, NJ 08544-2091, USA
[b]Department of Sociology, Bowling Green State University, 217 Williams Hall, Bowling Green, OH 43403, USA
[c]Center for Population and Health, Graduate School of Arts and Sciences, Georgetown University, 312 Healy Hall, Box 571197, Washington, DC 20057-1197, USA
[d]Center for Population and Health Survey Research, Bureau of Health Promotion, Department of Health, 5th Floor 503, Section 2, Li-Ming Road, Taichung 408, Taiwan, Republic of China

Received 15 January 2002; received in revised form 23 May 2002; accepted 11 October 2002

Abstract

Increasingly, researchers and health specialists are obtaining information on chronic illnesses from self-reports. This study validates self-reports of two major health conditions, hypertension and diabetes, based on a recent survey in Taiwan (SEBAS 2000). These data, based on a large, nationally representative sample of respondents aged 54 and older, include both self-reported health information and a physical examination. Average blood pressure readings, laboratory measures of glycosylated hemoglobin, and information on whether the respondent was taking medication for hypertension or diabetes are used to validate respondents' reports of high blood pressure and diabetes. The resulting comparisons reveal that self-reports vastly underestimate the prevalence of hypertension, but yield a reasonably accurate estimate of the prevalence of diabetes. Significant correlates of the accuracy of the self-reports include age, education, time of the most recent health exam, and cognitive function.

Keywords: Diabetes; Hypertension; Self-reports; Validity; Accuracy

1. Introduction

Increasingly, researchers and health specialists obtain information about chronic illnesses, conditions, and risk factors for disease from self-reports of the target population. The data, often collected via face-to-face interviews, mail-back questionnaires, or telephone interviews, have an obvious advantage over clinical records, namely that they can be obtained readily for a large and representative sample of the population without great expense. Unfortunately, they have a potential drawback as well: the accuracy of self-reports depends on the respondents' knowledge of the relevant information, ability to recall it, and willingness to report it.

In an effort to assess the accuracy of such data, researchers have compared them with "gold standards" such as medical or administrative records, medical provider surveys, or measurements obtained from physical examinations. Several notable problems characterize the majority of these comparisons. First, as Vargas et al. [1] lament, few validation studies have been based on a nationally representative sample. (One notable exception is the NHANES [1].) Most studies have been restricted geographically [2–7]; relied on volunteers [4,8]; or included only persons in good health [8], participants in a particular organization (such as an HMO) or screening program [2,9], or hospital patients [6]. Second, many validation studies have been based on small samples [4,5,8] that limit the ability of the analyst to identify characteristics of the respondents that are associated with inaccurate reporting. Finally, almost all studies have been carried out in wealthy Western populations, even though health interview surveys have become critical sources of data in developing and newly industrialized countries.

This study uses a recent survey in Taiwan to validate respondents' reports of two major health conditions: hypertension and diabetes. Taiwan provides an interesting case study for comparison with Western European and North American countries in that rapid industrialization and economic growth have produced a society with a modern health care system and high life expectancy, yet a substantial fraction of the older population is illiterate or lacks formal education. The survey, which includes both self-reported health information and a physical examination, is based on a large nationally representative sample. The resulting data permit not only unbiased estimation of the accuracy of self-reports, but also an examination of how demographic and social characteristics of participants affect the accuracy. The analysis also includes an assessment of the extent to which participants in the physical examination represent a select subsample of those interviewed.

* Corresponding author. Tel.: 609-258-5724; fax: 609-258-1039.
E-mail address: ngoldman@lotka.princeton.edu (N. Goldman).

0895-4356/03/$ – see front matter
doi:10.1016/S0895-4356(02)00580-2

2. Materials and methods

2.1. Data

The data come from a follow-up of the Survey of Health and Living Status of the Near Elderly and Elderly in Taiwan. This longitudinal survey began in 1989 with a national sample (including the institutionalized population) of 4,049 persons aged 60 and over and was extended in 1996 to include a national sample of 2,462 near-elderly persons, aged 50 through 66 in 1996 [10]. Since the initial interview, respondents have been reinterviewed at 2- to 3-year intervals. The survey contains extensive information on household composition and exchanges, emotional and instrumental support, physical and mental health, and health-related behaviors.

In 2000, a national subsample of respondents was randomly selected for the Social Environment and Biomarkers of Aging Study (SEBAS 2000). Elderly respondents (71 and older in 2000) and urban residents were oversampled. Respondents were interviewed in their homes between July and December, 2000. Several weeks after the interview, respondents who were willing and able collected 12-hr (overnight) urine samples and received physical examinations in nearby hospitals the following morning.

Among the 1,713 respondents selected for this study, a total of 1,497 provided interviews (92% of survivors), and 1,023 supplied biomarkers (68% of those interviewed). Biomarkers include measures obtained from 12-hr urine specimens and blood specimens, blood pressure readings, and anthropometric measures such as height, weight, and waist and hip circumference.

This study uses information from both the home-based interviews and the physical examinations. During the interviews, respondents were asked whether they ever had, and whether they currently have, high blood pressure and diabetes. The questionnaire employed nontechnical terms (in Chinese) for these conditions, and did not provide any description about the nature of the conditions. As part of the physical examination, registered nurses took two blood pressure measurements (about 1 min apart), using a mercury sphygmomanometer with the respondent in a seated position, at least 20 min after the respondent arrived at the hospital. A third blood pressure reading was taken during the physician's examination. Blood samples, collected during the hospital examination, provided estimates of glycosylated hemoglobin (HbA_{1c})—a measure of the percentage of hemoglobin molecules in the blood that are bound to glucose, obtained by HPLC assay (by Union Clinical Laboratory in Taipei). This test provides an integrated measure of glucose levels in the blood for the 2- to 3-month period prior to the test and is currently being used in the United States to monitor the effectiveness of diabetes treatment [11].

During the hospital visit, respondents also filled out questionnaires regarding their medical history, including all current medications. These data permit us to identify participants taking medication for hypertension and diabetes.

2.2. Statistical methods

The estimates presented in this analysis are based on one or both of the following samples: (1) persons interviewed in their homes (1,497); and (2) the subset of participants in the physical exam (1,023). The actual sample sizes used in the analysis are slightly smaller because of the exclusion of those interviewed by proxy (who, by definition, could not provide *self-reports* of illness) and those with missing values on relevant variables. Proxy respondents supplied interview information for 5.4% of the interviewed sample and 1.7% of participants in the physical exam.

2.2.1. Accuracy of self-reports

To assess the accuracy of the self-reports, biomedical criteria for hypertension and diabetes are compared with the proportion of respondents who report that they *currently* have high blood pressure or diabetes. Because more than 90% of respondents who report that they ever had one of these two conditions also report that they currently have the condition, the estimates would change little if we were to consider respondents' reports of ever having had the condition. Almost all respondents who report currently having one of these conditions indicate that a physician diagnosed it (97.8% for hypertension and 97.3% for diabetes).

Hypertension was identified according to the definition given in the Sixth Report of the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure [12]: systolic blood pressure equal to or greater than 140 mmHg and/or diastolic blood pressure equal to or greater than 90 mmHg and/or taking antihypertensive medication. The blood pressure readings were based on an average of the first two readings; however, we also explored the robustness of findings to the inclusion of the third reading. Respondents were considered to have diabetes if HbA_{1c} values were greater than or equal to 7.0%. Although HbA_{1c} is not routinely used as a screening procedure for diabetes, its high correlation with fasting plasma glucose levels suggests that HbA_{1c} values of about 6.5–7.0% indicate a need for diabetic treatment [11,13].

The prevalence rates of hypertension and diabetes are estimated separately from self-reports and from the medical criteria. Four measures of agreement are calculated: (1) sensitivity (the percentage fitting the medical criteria for hypertension/diabetes who report that they have the condition); (2) specificity (the percentage not fitting the medical criteria who report that they do not have the condition); (3) overall agreement (the percentage for whom the medical criteria and the self-reports are in agreement); and (4) the kappa statistic. All measures are unweighted; calculations not shown here indicate that incorporation of weights has little effect on the resulting estimates.

2.2.2. Multivariate models

Multivariate models are used in two stages of the analysis. First, logistic models are used to explore the factors associated with the probability that a respondent participates in the physical examination. Second, logistic models are used to identify characteristics associated with the likelihood that exam participants provide accurate reports of hypertension and diabetes. Separate models are estimated for respondents who fit the medical criteria for the condition and for those who do not. The rationale underlying the use of separate models is that different factors may be associated with a failure to report an existing condition and with a claim to have a condition that is not apparent.

The choice of explanatory variables for inclusion in the model of participation derives partly from the survey protocol and partly from the reasons provided by the respondent for declining to participate. Survey protocol prohibited the inclusion of respondents with severe health limitations. Reasons for declining to participate included: the examination was too much trouble; the respondent was healthy, and hence, had no need for the exam; the respondent just had a physical exam or, in general, had regular checkups; and the respondent's family felt that there was no need for the respondent to participate.

Variables in the model of participation in the physical exam include: demographic characteristics (age, sex), socioeconomic factors (level of education, employment status), coresidence pattern (whether the respondent lives alone, with a spouse only, with children or grandchildren, or with others), timing of the most recent health examination (within the most recent 3 months, between 3 months and 1 year ago, or no visits during the past year), the number of social activities (out of eight) in which the respondent participates, and three measures of self-reported health status. These three health measures comprise: (1) the respondent's assessment of his or her current health in terms of five categories ranging from poor to excellent; (2) the number of activities of daily living (out of six) for which the respondent reports having at least some difficulty performing; and (3) a score derived from 10 components of the CES-D scale, with higher values reflecting both more frequent and a greater number of depressive symptoms.

Because of the limited sample sizes and the expectation that some of the variables included in the participation model are unlikely to affect the accuracy of reports (e.g., coresidence patterns), fewer variables are incorporated into the accuracy models. Four of the same variables are included: age, sex, level of education, and recency of health exam. An additional variable is an index of cognitive function, calculated as the number of items that a respondent answers correctly. These items are derived from three tests: the modified Short Portable Mental Status Questionnaire [14], the modified Rey Auditory Verbal Learning Test [15], and a modification of the Digits Backward test [16]. Similar variables have been included in earlier validation studies [2,4,6,7].

Due to the sampling design of this survey, controls for age and urban–rural residence were introduced in preliminary models. However, because urban–rural residence showed little association with the outcomes, this variable was subsequently dropped from the models.

3. Results

3.1. Participation in the physical examination

Table 1 provides the distribution of explanatory variables used in the statistical models by participation status in the physical examination. The overall distributions reveal an excess of males over females (because of the selective migration of males after World War II when the Nationalist army came to Taiwan), the generally low levels of formal education among older Taiwanese, and the large proportion of Taiwanese that live in multigenerational households. Chi-square statistics indicate whether participants differ from nonparticipants with respect to each of these variables.

The estimated odds ratios in Table 2 show the effects of these factors on the likelihood that respondents participated in the physical exam based on a multivariate logistic model. Even in the presence of controls for health status, persons older than 70 were much less likely to participate than persons aged 54 to 59 (the omitted category). Sex, level of education, and working status were not significantly related to participation. Persons living with someone other than a spouse, child, or grandchild were less likely to take part in the exam than those living with children or grandchildren. Surprisingly, the recency of the last health exam was not significantly related to participation. Finally, the model indicates that persons with higher levels of social activity were more likely than their counterparts to participate in the exam.

Consistent with the survey protocol, limitations in activities of daily living (ADLs) are negatively related to participation in the physical exam. In contrast, respondents reporting themselves to be in excellent health—not those in fair or poor health—were much less likely to participate than those reporting average health. The net effect of these two counteracting selection processes is that participants in the health exam have the same average level of self-rated health status as nonparticipants (2.94 for nonparticipants and 2.93 for participants, coded on a five-point scale; data not shown).

3.2. Accuracy of self-reports of hypertension and diabetes

Table 3 presents the proportions of respondents reporting that they currently have hypertension and diabetes, as well as the proportions who satisfy the medical criteria for these two conditions. These estimates indicate that self-reports vastly underestimate the prevalence of hypertension, but fairly accurately measure the prevalence of diabetes.

The measures in Table 3 reveal how the huge discrepancy for hypertension arises. Only half of those who fit the medical criteria for hypertension report having high blood pressure (i.e., the sensitivity equals 49.4%), whereas few respondents (4.7%) who do not satisfy the medical criteria self-report the condition. The specificity is also extremely high (98.3%) for diabetes, but the sensitivity is much higher than for hypertension (85.2% vs. 49.4%). Not surprisingly, the levels of overall agreement and the kappa statistic are considerably higher for diabetes than for hypertension.

The estimates for hypertension are robust to the use of the average of three rather than two blood pressure readings. Spe-

Table 1
Distribution and means of explanatory variables (unweighted) for participants and nonparticipants in the physical examination, SEBAS (2000)

	Percentage or Mean				
Variables	Total	Participants	Nonparticipants	χ^2 test	*P*-value
Age					
54–59	19.6	22.0	13.8		
60–69	29.8	31.7	25.3		
70–79	40.8	38.4	46.9		
80+	9.7	8.0	14.0	30.08	.00
Sex					
Male	56.5	58.3	52.1		
Female	43.5	41.7	47.9	4.50	.03
Education[a]					
No formal education	34.9	32.9	40.1		
Primary education	39.6	40.6	37.1		
Secondary education or higher	25.4	26.5	22.9	6.70	.04
Employment status[a]					
Working	20.4	21.9	16.7		
Not working	79.6	78.1	83.3	4.83	.03
Coresidence					
Living alone	8.0	7.5	9.3		
Living with spouse only	18.1	19.1	15.5		
Living with child or grandchild	71.5	71.5	71.5		
Living with parent or relative or friend	2.4	1.9	3.7	7.25	.06
Recency of last health exam					
In the past 3 months	20.3	19.5	22.1		
In the past 4 to 12 months	13.8	14.7	11.3		
Not in the past year	64.6	64.6	64.6		
Missing	1.4	1.1	2.0	5.08	.17
Number of social activities[b]	0.6	0.7	0.5	30.63	.00
Number of ADL difficulties[c]	0.2	0.1	0.5	55.11	.00
Self-assessed health					
Excellent	13.4	12.1	16.7		
Good	12.5	13.0	11.6		
Average	45.6	48.0	39.8		
Fair or poor	28.4	27.0	31.9	11.91	.01
Depression score (CES-D)[d]	5.9	5.5	6.8	49.52	.01
Number of correct memory items[e]	16.1	16.6	15.1	63.20	.00
Number of observations	1411	1004	407		

[a] These variables contain fewer than 10 missing values; missing values were replaced by the modal response according to whether the respondent was a participant or nonparticipant.

[b] The eight social activities include: neighborhood, religious, farmer, political, social service, and village associations, elderly clubs, and elderly education.

[c] The six ADL activities include: bathing, dressing, eating, getting out of bed, moving about the house, and using the toilet.

[d] The 10 CES-D items include: having a poor appetite, feeling that doing anything was exhausting, sleeping poorly, being in a terrible mood, feeling lonely, feeling that people weren't friendly, feeling anguished, unable to gather energy, feeling joyful, and feeling that life was going well. Respondents indicated whether these experiences occurred during the past week, and if so, whether they happened rarely (1 day), sometimes (2-3 days), or often (4 or more days). The two positive items were reverse coded so that a high score indicates more depressive symptoms. The maximum possible score is 30 points.

[e] The 12 memory items include address, year, month, day, day of the week, age, mother's maiden name, current president, former president, subtraction by threes, number of items correctly recalled, and repetition of numbers in reverse order. The maximum possible score is 24 points.

cifically, based on estimates for the 998 respondents with three blood pressure readings, a change from two to three readings leads to a slight increase in sensitivity (from 49.2 to 52.0%), an even smaller increase in specificity (from 95.3 to 96.0%), and a modest decrease in the proportion satisfying the medical criteria for hypertension (from 57.4 to 54.7%; estimates not shown).

3.3. Factors associated with accuracy of self-reports

Table 4 presents the estimates from logistic models of the accuracy of the self-reports, separately among those who satisfy and those who do not satisfy the medical criteria. Age is significantly associated with the accuracy of reports of hypertension: older hypertensive respondents provide more accurate self-reports than persons below age 60, whereas older normotensive respondents provide less accurate reports than their younger counterparts. That is, older respondents are more likely than those aged 54 to 59 to report hypertension, whether or not they actually have the condition.

Among hypertensive respondents, those with formal education (especially secondary schooling) are more likely than uneducated respondents to acknowledge the condition. Hypertensive respondents who had a health exam in the past 3 months are more likely to self-report high blood pressure than those with no exams in the past year. Neither sex nor cognitive function is significantly related to the accuracy of reports of hypertension.

Table 2
Estimated odds ratios and 95% confidence intervals (CI) for logistic model of the probability that a respondent participates in the physical examination, SEBAS (2000)

Variables	Odds Ratios	95% CI
Age		
54–59	1.00	
60–69	0.81	0.55, 1.20
70–79	0.51[a]	0.34, 0.75
80+	0.43[a]	0.26, 0.70
Sex		
Male	1.00	
Female	0.81	0.62, 1.05
Education		
No formal education	1.00	
Primary education	1.07	0.80, 1.44
Secondary education or higher	1.19	0.84, 1.69
Employment status		
Working	0.93	0.66, 1.32
Not working	1.00	
Coresidence		
Living alone	0.89	0.58, 1.38
Living with spouse only	1.17	0.84, 1.63
Living with child or grandchild	1.00	
Living with parent or relative or friend	0.47[a]	0.23, 0.97
Recency of last health exam[b]		
In the past 3 months	0.79	0.58, 1.06
In the past 4 to 12 months	1.18	0.81, 1.71
Not in the past year	1.00	
Number of social activities	1.16[a]	1.01, 1.32
Number of ADL difficulties	0.72[a]	0.62, 0.82
Self-assessed health		
Excellent	0.52[a]	0.36, 0.75
Good	0.87	0.59, 1.28
Average	1.00	
Fair or poor	1.05	0.77, 1.44
Depression score (CES-D)	0.99	0.96, 1.01
Number of observations	1,411	

[a] $P < .05$, two-tailed tests.
[b] A dummy variable indicating that the respondent has missing information on the time of the last health exam was also included in the model.

Only one of the odds ratios for diabetes is significant. Unlike hypertension, the results indicate that better cognitive function is associated with greater accuracy of reporting among nondiabetics. The lack of significance of other factors, even when the odds ratios differ substantially from unity, arises partly because so few respondents inaccurately report diabetic status. Nevertheless, the odds ratios for age and education are generally in a similar direction to those for hypertension.

4. Discussion

Although Taiwan shares many characteristics with industrialized Western societies, such as high life expectancy, a similar cause-of-death structure, and a modern health care system, the social and cultural contrasts with Western populations provide an opportunity to gain insights from a comparison of findings. Overall, the results obtained from this study support those from earlier ones carried out primarily in Western Europe and North America. For example, the high accuracy of self-reports of diabetes in Taiwan is generally consistent with several studies that find self-reports of diabetes to be more accurate than those for other chronic diseases [1,7,17], and the contrast with hypertension confirms an important result noted in previous research, namely that the accuracy of self-reports varies a great deal among chronic conditions. In addition, the covariates that emerge as significant correlates of the accuracy of self-reports in Taiwan—age, education, recency of the last health exam, and cognitive function—are largely the same variables identified in other studies.

Nevertheless, there are important differences. The accuracy of self-reports of hypertension in this study is lower than reported elsewhere—for example, estimates of sensitivity range between 64 and 91% across several countries in Western Europe and North America in contrast to 49% for Taiwan [2,4,6,8,9,17]. Theoretically, such low levels of accuracy in Taiwan could arise via several interrelated mechanisms: (1) persistence of traditional Chinese beliefs about hypertension; (2) low levels of blood pressure screening; (3) failure of health-care personnel to communicate diagnoses with patients; and (4) poor recall (or unwillingness to report information) by respondents.

Traditional Chinese medicine does not share the Western medical perspective that high blood pressure is a chronic disease associated with physiologic dysfunction, nor Western beliefs regarding the origins of the condition and recommended treatment protocol [18]. Thus, it is plausible that the persistence of such traditional beliefs has led to a lower recognition of hypertension among older Taiwanese. This is likely to be the case even though traditional practitioners are sought less frequently than Western ones in modern Taiwanese society. For example, data from the 1999 wave of the survey indicate that whereas only about one-fifth of the elderly reported visiting a Chinese medical facility in the past year, about 87% had visited a Western one in the past year—a fre-

Table 3
Comparison between self-reports and medical criteria for hypertension and diabetes, SEBAS (2000)

	Hypertension			Diabetes		
	N[a]	Measure	95% CI	N[a]	Measure	95% CI
Self-reporting condition	1003	30.3%	27.5%, 33.2%	1003	14.6%	12.4%, 16.7%
Medical criteria for condition	1003	57.3%	54.3%, 60.4%	1003	15.5%	13.2%, 17.7%
Sensitivity of self-report	575	49.4%	45.3%, 53.5%	155	85.2%	79.5%, 90.8%
Specificity of self-report	428	95.3%	93.3%, 97.3%	848	98.3%	97.5%, 99.2%
Overall agreement	1003	69.0%	66.1%, 71.9%	1003	96.3%	95.1%, 97.5%
Kappa	1003	0.41	0.36, 0.47	1003	0.86	0.79, 0.92

[a] One respondent did not provide a self-report for hypertension and one respondent was missing a measure of glycosylated hemoglobin.

Table 4
Estimated odds ratios (OR) and 95% confidence intervals (CI) for logistic models of the probability that a respondent provides accurate self-reports of hypertension and diabetes, by whether or not the respondent has the condition (based on medical criteria), SEBAS (2000)

	Hypertension				Diabetes			
	With condition		Without condition		With condition		Without condition	
Variables	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Age								
54–59	1.00		1.00		1.00		1.00	
60–69	2.58[a]	1.49, 4.45	0.35	0.07, 1.78	2.63	0.69, 9.93	4.03	0.41, 39.61
70–79	2.17[a]	1.28, 3.68	0.20[a]	0.04, 0.97	2.89	0.82, 10.22	0.75	0.19, 3.03
80+	1.22	0.58, 2.60	0.53	0.04, 6.53	4.29	0.41, 44.82	1.26	0.16, 9.82
Sex								
Male	1.00		1.00		1.00		1.00	
Female	1.12	0.77,1.63	1.02	0.36, 2.88	0.58	0.19, 1.73	0.76	0.22, 2.60
Education								
No formal education	1.00		1.00		1.00		1.00	
Primary education	1.58[a]	1.03, 2.43	0.87	0.25, 3.01	0.66	0.22, 1.97	0.69	0.14, 3.44
Secondary education or higher	2.90[a]	1.69, 4.96	0.56	0.15, 2.13	1.76	0.36, 8.65	0.28	0.05, 1.57
Recency of last health exam[b]								
In the past 3 months	1.80[a]	1.16, 2.79	1.30	0.36, 4.72	6.74	0.83, 55.10	1.29	0.27, 6.12
In the past 4 to 12 months	1.63	0.99, 2.67	1.71	0.37, 7.88	0.46	0.14, 1.44	1.16	0.24, 5.56
Not in the past year	1.00		1.00		1.00		1.00	
Number of correct memory items	0.95	0.90, 1.00	1.07	0.92, 1.24	0.98	0.85, 1.12	1.21[a]	1.06, 1.39
Number of observations	572		428		153		839	

[a] $P < .05$, two-tailed tests.
[b] A dummy variable indicating that the respondent has missing information on the time of the last health exam was also included in the model.

quency of contact with the modern health care system that is only slightly less than that in the United States [19].

On the other hand, low levels of blood pressure screening are unlikely to provide an adequate explanation. For example, data for early waves of the survey indicate that 77% of the elderly in 1989 (persons aged 60 and over) and 85% of this cohort in 1996 had their blood pressure checked within the previous year. However, data from the 1989 survey reveal that the vast majority of elderly persons were not aware of the levels of blood pressure associated with a diagnosis of hypertension or of the health risks associated with this condition. Thus, it is possible that, despite high participation in blood pressure screening, many hypertensive respondents never received clear diagnoses or feedback from their physicians.

Explanations pertaining to the frequency of screening, communication of information, and recall of the information are consistent with results from the multivariate analysis. The logistic models indicate that older respondents (with or without hypertension) are more likely than their younger counterparts to self-report hypertension. This finding is unlikely to result solely from memory problems among older respondents; rather, it is probably also related to older individuals having more exposure than their younger counterparts to screening procedures (but not necessarily accurate information about the results). Similar interpretations are consistent with the estimated differentials by education. That is, the much higher acknowledgment of high blood pressure among hypertensive respondents with formal education, particularly those with a secondary education, suggests that the low accuracy of self-reports of hypertension may be partly attributable to the low levels of education among older Taiwanese. Education in Taiwan is positively associated with seeking Western medical care, participating in blood pressure screening, and presumably having a better understanding and recall of the information.

The low agreement between the self-reports and medical criteria for hypertension may result not only from the factors described above, which bias the self-reports, but also from inadequacies of the "gold standard." Our results suggest that the problem is not simply the number of blood pressure measurements: for example, the estimates of sensitivity and specificity are robust to the inclusion of a third blood pressure reading taken at a later time. Rather, the weakness of the medical criteria used to identify hypertension may result from reliance on measurements taken during a single visit to a hospital, instead of multiple readings over longer time intervals in a setting familiar to the respondent. The presence of white coat hypertension (WCH)—i.e., abnormally elevated blood pressure readings in a clinical setting [20]—could readily generate an overestimate of hypertension prevalence and a concomitant underestimate of the sensitivity of the self-reports. (For example, raising the threshold values of systolic and diastolic blood pressure from the current definition of 140/90 mm to 150/100 mm—to compensate for WCH—would increase the sensitivity of the self-reports in this study from 49.4 to 67.1%.) WCH-type effects may partly explain the low levels of accuracy of hypertension reports found in this study. However, they cannot account for the similarity between our estimate of sensitivity (49.4%) and the estimate (49.3%) obtained for three Taiwanese districts in a separate undertaking [4]. The latter study defined hypertension on the basis of a respondent's medical history or elevated blood pressure readings at more than two visits. Thus, these comparisons suggest that inaccuracy of self-reports—via the pathways discussed above—continues to be a central part of the explanation.

An important lesson to be drawn from this study is that it is possible to obtain detailed self-reported information *and* extensive medical data from a nationally representative sample of the population. Although about 30% of persons interviewed did not participate in the medical exam in SEBAS 2000, this rate is considerably lower than corresponding values in other studies (e.g., more than 80% of interviewed persons did not participate in Wu et al.'s [4] clinical study). Given that the vast majority of sampled persons in SEBAS responded to the pre-examination interview, the potential effects of nonparticipation in the medical examination can be readily assessed by a comparison of the characteristics of nonparticipants and participants. These comparisons reveal that respondents who received the medical exam reported the same health status, on average, as those who did not, a result that suggests that, in the aggregate, the resulting estimates from the medical data are not likely to be seriously biased. Thus, data collected in a sample survey of the general population such as SEBAS are much more suitable for an evaluation of the accuracy of self-reports than are the selective samples frequently used in other studies, such as those based on participation in health care or screening programs. These latter samples, which are probably selective of persons who obtain regular medical care or are highly informed about their health, are likely to yield overly optimistic conclusions about the use of self-reports to estimate the prevalence of chronic conditions.

Acknowledgments

This work has been supported by the Demography and Epidemiology Unit of the Behavioral and Social Research Program of the National Institute on Aging, under grant numbers R01AG16790 and R01AG16661, and by the National Institute of Child Health and Human Development under grant number 5P30HD32030. The authors would also like to acknowledge the generous support of the Graduate School of Arts and Sciences, Georgetown University and the assistance of Dr. Jennifer Cornman, Dr. Dana Glei, and Dr. Chris Seplaki.

References

[1] Vargas C, Burt V, Gillum R, et al. Validity of self-reported hypertension in the National Health and Nutrition Examination Survey III, 1988–1991. Prev Med 1997;26:678–85.

[2] Bush T, Miller SR, Golden AL, et al. Self-report and medical record report agreement of selected medical conditions in the elderly. Am J Public Health 1989;79:1554–6.

[3] Haapanen N, Miilunpalo S, Pasanen M, et al. Agreement between questionnaire data and medical records of chronic diseases in middle-aged and elderly Finnish men and women. Am J Epidemiol 1997;145:762–9.

[4] Wu S, Li C, Ke D. The agreement between self-reporting and clinical diagnosis for selected medical conditions among the elderly in Taiwan. Public Health 2000;114:137–42.

[5] Johansson J, Hellenius M, Elofsson S, et al. Self-report as a selection instrument in screening for cardiovascular disease risk. Am J Prev Med 1999;16(4):322–4.

[6] Kehoe R, Wu S, Leske C, et al. Comparing self-reported and physician reported medical history. Am J Epidemiol 1994;139(8):813–8.

[7] Kriegsman D, Penninx B, van Eijk J, et al. Self-reports and general practitioner information on the presence of chronic diseases in community dwelling elderly. A study on the accuracy of patients' self-reports and on determinants of inaccuracy. J Clin Epidemiol 1996;49(12):1407–17.

[8] Tormo M, Navarro C, Chirlaque M, et al. Validation of self diagnosis of high blood pressure in a sample of the Spanish EPIC cohort: overall agreement and predictive values. J Epidemiol Community Health 2000;54:221–6.

[9] Martin LM, Leff M, Calonge N, et al. Validation of self-reported chronic conditions and health services in a managed care population. Am J Prev Med 2000;18:215–8.

[10] Taiwan Provincial Institute of Family Planning. Population Studies Center and Institute of Gerontology, University of Michigan. 1989 Survey of Health and Living Status of the Elderly in Taiwan: Questionnaire and Survey Design. Comparative Study of the Elderly in Four Asian Countries, Research Report no. 1. Ann Arbor, MI: University of Michigan; 1989.

[11] Davidson M, Schriger D, Peters A, et al. Relationship between fasting plasma glucose and glycosylated hemoglobin. Potential for false-positive diagnoses of type 2 diabetes using new diagnostic criteria. JAMA 1999;281:1203–10.

[12] National Institutes on Health. National Heart, Lung and Blood Institute. The sixth report of the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure, 1997. NIH Publication No. 98–4080.

[13] Nordenson N. 2001. Online Bluecross Blueprint for Health. (http://blueprint.bluecrossmn.com/article/gale/100084629).

[14] Pfeiffer EA. A short mental status questionnaire for the assessment of organic brain deficit in elderly patients. J Am Geriatr Soc 1975;23:433–41.

[15] Lezach MD. Neuropsychological Assessment. 2nd ed. Oxford: The Oxford University Press; 1983.

[16] Wechsler D. WAIS-R manual. New York: Psychological Corporation; 1981.

[17] Robinson J, Young T, Roos L, et al. Estimating the burden of disease. Comparing administrative data and self-reports. Med Care 1997;35(9):932–47.

[18] Chen, J. The concept of treating high blood pressure and Chinese medicine treatments. Chin Med J 2000;6(1):41–50.

[19] National Center for Health Statistics. Health, United States, 2000. Hyattsville, MD: Public Health Service; 2000.

[20] Landray MJ, Lip GYH. White coat hypertension: a recognized syndrome with uncertain implications. J Hum Hypertens 1999;13:5–8.