BMJ Open

Association between an individual housing-based socioeconomic index and inconsistent self-reporting of health conditions: a prospective cohort study in the Mayo Clinic Biobank

Euijung Ryu,[1] Janet E Olson,[1] Young J Juhn,[2] Matthew A Hathcock,[1] Chung-Il Wi,[2] James R Cerhan,[1] Kathleen J Yost,[1] Paul Y Takahashi[3]

[1]Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA
[2]Department of Pediatric and Adolescent Medicine, Mayo Clinic, Rochester, Minnesota, USA
[3]Department of Internal Medicine, Mayo Clinic, Rochester, Minnesota, USA

Correspondence to
Dr Paul Y Takahashi;
takahashi.paul@mayo.edu

ABSTRACT

Objective Using surveys to collect self-reported information on health and disease is commonly used in clinical practice and epidemiological research. However, the inconsistency of self-reported information collected longitudinally in repeated surveys is not well investigated. We aimed to investigate whether a socioeconomic status based on current housing characteristics, HOUsing-based SocioEconomic Status (HOUSES) index linking current address information to real estate property data, is associated with inconsistent self-reporting.

Study setting and participants We performed a prospective cohort study using the Mayo Clinic Biobank (MCB) participants who resided in Olmsted County, Minnesota, USA, at the time of enrolment between 2009 and 2013, and were invited for a 4-year follow-up survey (n=11 717).

Primary and secondary outcome measures Using repeated survey data collected at the baseline and 4 years later, the primary outcome was the inconsistency in survey results when reporting prevalent diseases, defined by reporting to have 'ever' been diagnosed with a given disease in the baseline survey but reported 'never' in the follow-up survey. Secondary outcome was the response rate for the 4-year follow-up survey.

Results Among the MCB participants invited for the 4-year follow-up survey, 8508/11 717 (73%) responded to the survey. Forty-three per cent had at least one inconsistent self-reported disease. Lower HOUSES was associated with higher inconsistency rates, and the association remained significant after pertinent characteristics such as age and perceived general health (OR=1.46; 95% CI 1.17 to 1.84 for the lowest compared with the highest HOUSES decile). HOUSES was also associated with lower response rate for the follow-up survey (56% vs 77% for the lowest vs the highest HOUSES decile).

Conclusion This study demonstrates the importance of using the HOUSES index that reflects current SES when using self-reporting through repeated surveys, as the HOUSES index at baseline survey was inversely associated with inconsistent self-report and the response rate for the follow-up survey.

Strengths and limitations of this study

► The HOUsing-based SocioEconomic Status index is a well-established socioeconomic status measure reflecting socioeconomic situation around the time of surveys which serves as an excellent measure for studying the association between current socioeconomic status and inconsistent self-reporting in repeated surveys.
► Loss of follow-up, an important aspect when repeating surveys, in the cohort (the Mayo Clinic Biobank participants residing in a local community) was negligible.
► This study is limited by characterising the likelihood of inconsistent self-reports, without knowing the true medical status.

INTRODUCTION

In many epidemiological studies, self-report is widely used to collect health information including subjects' disease status, typically collected through survey.[1] These data are used to calculate the incidence and prevalence of various diseases. Self-administered surveys are popular and have several advantages compared with other approaches such as manual medical chart reviews or computational phenotyping approaches using electronic health records (eg, using natural language processing for case identification[2–4]) because they are easy to implement, less labour intensive, less expensive and may produce more complete data collection.[5 6] In addition, surveys may be the only approach to obtain health information if the research team lacks access to the medical records. Despite the popularity of surveys as data collection tools, the limitations are also well documented, including selection bias due to non-participation to the surveys and high

misclassification (either false-positives or false-negatives) driven by many factors such as recall bias and health literacy.[7–9] The accuracy of self-reports varies by disease type (ie, conditions with easy diagnosis compared with those presenting non-specific symptoms), and may also reflect the challenges that many patients face when trying to understand whether they suffer or do not suffer from an illness.[10–13] While extensive investigations have been performed to identify risk factors associated with the accuracy of self-reports for a single survey, studies assessing inconsistency of self-reports from repeated surveys are less common.

Due to the significant impact on health, socioeconomic status (SES) is often investigated in studies for assessing inconsistency of self-reported diseases from repeated surveys.[14–18] For example, a study of mammography history from a repeated survey done over a 2-year time span reported that women from low income households had higher inconsistent responses (ie, ever user at baseline but never user at follow-up).[19] However, the inconsistency was not significantly associated with education levels, a commonly used SES measure.[19] Another study using seven chronic conditions showed that low wealth (net worth) was one of the key predictors for inconsistent self-reports, but was not associated with education levels and income.[20] These inconclusive findings are partially due to the type of SES measure (ie, education level, income, wealth or other SES measures) used in studies. Specifically, some SES measures (eg, education level) remain largely static over time which may not be sufficient to capture important changes over the life course, especially due to ageing or disability. Although other SES measures such as wealth and income can be collected close to the time of the baseline surveys, this information is mostly unavailable in patients' medical records, and the self-reported data suffer from similar biases that most survey-based data do, including response bias and inaccurate reporting.[21 22]

To overcome these problems, we developed and validated an objective (not relying on self-reports) SES index, termed the HOUSES (HOUsing-based SocioEconomic Status) index, by linking current address information to publicly available real estate property data.[23] The validity of the HOUSES as an SES measure and its utility in studying health outcomes have been well established.[23–26] In addition, it is suggested that the HOUSES index may integrate multidimensional factors beyond what commonly used SES measures provide which includes ageing, social isolation and functional declines.[27]

In this study, we sought to characterise the degree of inconsistent self-reports among Mayo Clinic Biobank (MCB) participants having a wide range of health conditions, by using baseline and follow-up questionnaires administered 4 years apart. Our primary aim was to test whether SES measured by the HOUSES index is associated with inconsistency in self-reported illness. Since the HOUSES index is based on housing characteristics close to the time of enrolment into the MCB, it will likely reflect

SES around the time of surveys that may capture individual characteristics such as health literacy or cognitive ability.[28] Health literacy is often defined as the degree to which individuals have the capacity to obtain, process and understand basic health information, and the services needed to make appropriate health decisions.[29] A commonly used screening tool for assessing health literacy measures the extent to which the person is confident in his/her ability to complete medical forms,[30] a skill very similar to filling out a research survey about their medical history. Furthermore, poor health literacy has been associated with poor illness understanding in diverse chronic conditions.[31 32] Therefore, we hypothesise that subjects with low HOUSES index will have less reliable self-reported information (ie, higher inconsistency when asked repeatedly), potentially due to difficulty completing the survey and poor illness understanding driven by low health literacy. As a secondary aim, we sought to assess the response rate of the follow-up survey in relation to the HOUSES index.

MATERIALS AND METHODS
Study design and subjects
This cohort study used data from the MCB participants who were Olmsted County, Minnesota, residents at the biobank enrolment and were invited for a 4-year follow-up questionnaire as of December 2016. The MCB started its enrolment in April 2009, mostly based on adult Mayo Clinic primary care patients via mailed invitation while allowing volunteers. These participants are predominantly whites (95%), more females (58%) and relatively old (median age at the enrolment was 62 years).[33] The MCB participants completed self-reported health-related questionnaires at enrolment, provided blood samples and consented to link their electronic medical records.[33–35] Approximately 4 years after the initial enrolment, a follow-up questionnaire was mailed to the participants to update their health information. For investigating whether socioeconomic disparity exists for reporting inconsistent health conditions between repeated questionnaires, we included participants with a HOUSES index score.

Primary outcomes
The status of the follow-up survey response was recorded among the MCB participants invited as of December 2016.[33] Questions on a total of 63 health conditions from 11 disease categories were included in both the baseline and the 4-year follow-up surveys (online supplementary table 1). In both surveys, the participants were asked to indicate the age when they were first diagnosed with each health condition (≤19, 20–49, 50–64, 65–79 and ≥80 years), while marking 'None' if they have not been diagnosed with the condition. The participants who indicated any of the age group described above were classified as ever been diagnosed with a given disease. Those indicated 'None' were classified as never been diagnosed. For each disease category, a subject was defined to have

Table 1 Baseline sociodemographic characteristics of the study cohort, stratified by the response status of the follow-up survey

Characteristics	Survey responders (n=8508)	Survey non-responders (n=3209)	P values
Age (years) at enrolment, n (%)			<0.01
18–44	1807 (21.2)	1418 (44.2)	
45–54	1775 (20.9)	673 (21.0)	
55–64	1875 (22.0)	484 (15.1)	
≥65	3051 (35.9)	634 (19.8)	
Female sex, n (%)	5426 (63.8)	1969 (61.4)	0.02
Race, n (%)			<0.01
Whites	8014 (94.2)	2726 (84.9)	
Asians	135 (1.6)	112 (3.5)	
Black	38 (0.4)	51 (1.6)	
Others	321 (3.8)	320 (9.9)	
Perceived general health, n (%)			<0.01
Poor/fair	373 (4.4)	234 (7.7)	
Education, n (%)			<0.01
High school or less	1137 (13.4)	489 (15.2)	
Some college	2730 (32.1)	1110 (34.6)	
College graduate or higher	4545 (54.4)	1437 (44.8)	
Not reported	96 (1.1)	173 (5.4)	
HOUSES, n (%)			<0.01
H1 (HOUSES<10%)	636 (7.5)	512 (16.0)	
H2 (≤10% HOUSES<20%)	826 (9.7)	344 (10.7)	
H3 (≤20% HOUSES<30%)	856 (10.1)	316 (9.8)	
H4 (≤30% HOUSES<40%)	865 (10.2)	299 (9.3)	
H5 (≤40% HOUSES<50%)	861 (10.1)	290 (9.0)	
H6 (≤50% HOUSES<60%)	841 (9.9)	331 (10.3)	
H7 (≤60% HOUSES<70%)	888 (10.4)	292 (9.1)	
H8 (≤70% HOUSES<80%)	906 (10.6)	276 (8.6)	
H9 (≤80% HOUSES<90%)	918 (10.8)	279 (8.7)	
H10 (HOUSES≥90%)	911 (10.7)	270 (8.4)	

HOUSES, HOUsing-based SocioEconomic Status.

inconsistent self-report if one reported to have ever been diagnosed with a disease(s) within the category in the baseline survey but reported never having been diagnosed with the same disease(s) in the follow-up survey. A subject was defined to have an overall inconsistent self-reporting if he or she had at least one inconsistent disease category. The primary outcome was the overall inconsistency ('1' if reporting to have 'ever' been diagnosed in the baseline survey but reported 'never' in the follow-up survey for at least one disease category; '0' if consistent self-reports for all disease categories).

Predictors
The main predictor of interest is the HOUSES index, an individual-level housing-based SES measure. Briefly, the HOUSES is a composite index formulated from objective (not relying on self-reporting) housing characteristics (assessed housing value, square footage of the unit, number of bedrooms and number of bathrooms) associated with an individual's address information provided at the MCB enrolment.[23 24 27] While the HOUSES index is commonly analysed as quartiles, this study used deciles (from the lowest percentile (HOUSES≤10%) to the highest percentile (HOUSES>90%)) deciles, to see any granular effect of the HOUSES index. The higher the decile, the higher the SES.[23]

Other factors considered for an association with inconsistency were age at enrolment by category (18–44, 45–54, 55–64 and 65 years or older), sex, race (whites, Asians, blacks and others), education attainment (high school or less, some college, college graduate or higher)
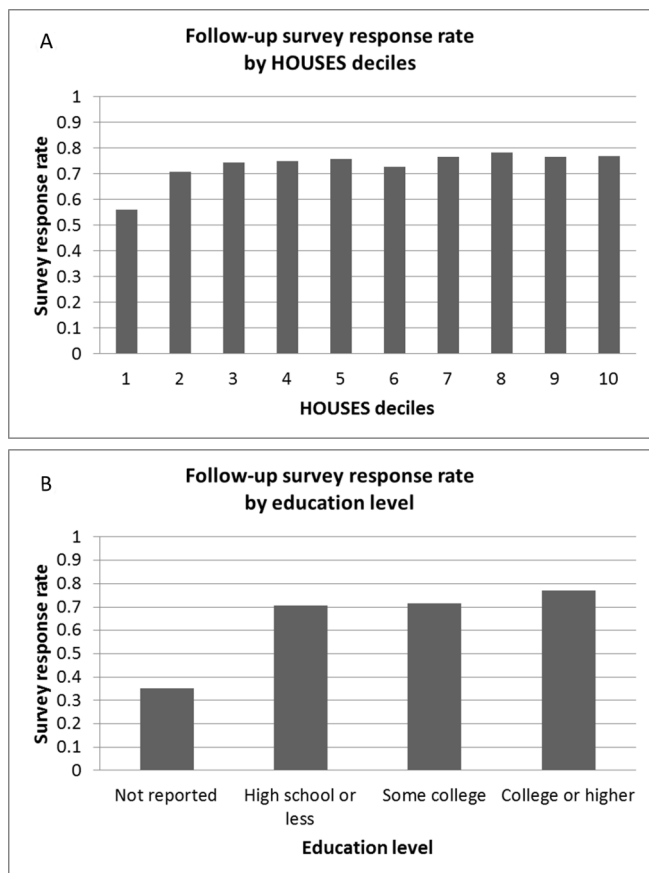
Figure 1 The proportion of the study subjects who answered the 4-year follow-up survey by HOUSES deciles (A) and education level (B). HOUSES, HOUsing-based SocioEconomic Status.

and perceived health in two categories (poor/fair or at least good (excellent/very good/good)) at each questionnaire. Changes in perceived health between the two surveys were also considered and defined as worse ('at least good' at baseline and 'poor/fair' follow-up survey), consistently poor ('poor/fair' at both surveys), improved ('poor/fair' at baseline and 'at least good' at follow-up
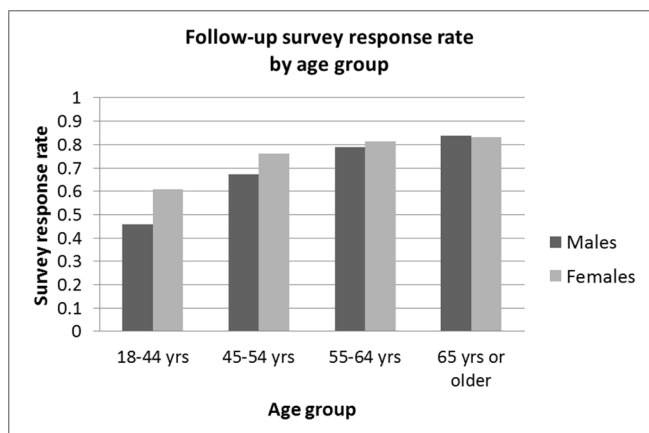

Figure 2 The proportion of the study subjects who answered the 4-year follow-up survey for each age group (18–44, 45–54, 55–64 and 65+ years), stratified by sex.

survey) and consistently good ('at least good' at both surveys).

Statistical analysis

Basic sociodemographic characteristics of the study cohort were summarised using percentages for categorical variables (age category, sex, race, education, HOUSES deciles and perceived health) and compared for the response status of the follow-up survey using Kruskal-Wallis test. The degree of inconsistent self-reporting was summarised descriptively for both overall and disease category specific. For the association between each sociodemographic characteristic, including the HOUSES decile, and the risk of inconsistent overall self-reporting, logistic regression models were used univariately, adjusting only for age. Multivariable models were also used to assess whether HOUSES scores were independently associated with the inconsistency rates, adjusting for age, sex, race, education level and changes in perceived general health between two surveys.

RESULTS

Study cohort and factors associated with survey response status

We observed that the subjects with low SES measured by the HOUSES index were less likely to respond to the follow-up survey when invited. Among 11 717 MCB participants invited for a 4-year follow-up survey, 73% (n=8508) responded to the survey with the median age at the MCB enrolment of 55 years (25th–75th percentiles: 43–67 years), 63% females, 92% whites. Participants in the lowest HOSUES deciles were less likely to respond to the survey when invited (56% vs 77% responded among the lowest vs the highest HOUSES deciles; table 1 and figure 1A: p<0.001). Subjects who did not report education attainment (n=264; 2.3% of the total invited) were less likely to respond to the follow-up survey (35% response rate; figure 1B). The response rate tended to be higher among those with older age (36% vs 20% for 65 years or older, among responders vs non-responders), white race (94% vs 85%), better perceived general health at baseline (96% vs 92%) and higher education level (54% vs 45% for college graduate or higher). Overall, the survey response rate was similar between males and females (64% vs 61%). However, the rate was higher in females in younger age groups (eg, 61% vs 46% in those aged between 18 and 44 years), while similar in the oldest age group (83% each; figure 2).

Degree of inconsistency in self-reporting of health conditions

Reporting health conditions inconsistently in a repeated survey is fairly common in our study cohort. A total of 63 illnesses were checked for inconsistent self-reporting between the baseline and the 4-year follow-up survey. About 90% of the participants reported having ever been diagnosed with at least one disease at the baseline (table 2). Among the participants who reported to have

Table 2 Degree of inconsistent self-reports in a repeated survey based on two time periods (baseline and the 4-year follow-up), sorted by disease prevalence at the baseline survey

	Disease prevalence at baseline	Inconsistent self-reported diseases between two time periods			
		Number of subjects*	One disease	Two diseases	Three or more diseases
Overall, n (%)	7396 (89.8)	3193 (43.2)	2086 (28.2)	748 (10.1)	359 (4.9)
Disease categories (number of sub-conditions), n (%)					
Cardiovascular (8)	4304 (52.3)	879 (20.6)	743 (17.4)	118 (2.8)	18 (0.4)
Rheumatological (4)	2610 (31.8)	761 (29.4)	727 (28.1)	32 (1.2)	2 (0.1)
Gastrointestinal (5)	2352 (28.6)	512 (22.1)	494 (21.3)	17 (0.7)	1 (0.0)
Cancer (22)	1916 (23.3)	455 (23.9)	404 (21.3)	43 (2.3)	8 (0.4)
Eye (4)	3358 (40.9)	321 (9.7)	300 (9.1)	21 (0.6)	0 (0.0)
Respiratory (4)	1846 (22.5)	320 (17.5)	310 (17.0)	10 (0.6)	310 (17.0)
Mental health (6)	2382 (29.0)	640 (27.3)	544 (23.2)	89 (3.8)	7 (0.3)
Neurological (6)	1839 (22.4)	382 (21.1)	368 (20.3)	11 (0.6)	3 (0.2)
Gynaecological (1)	488 (9.2)	90 (18.4)	90 (18.4)	NA	NA
Liver (2)	255 (3.1)	61 (24.6)	58 (23.4)	3 (1.2)	NA
Haematological (1)	30 (0.003)	15 (50.0)	15 (50.0)	NA	NA

*The number of subjects having at least one prevalent disease with inconsistent reporting (ie, self-reported 'ever' having the disease at the baseline but 'never' had it at the follow-up survey).
NA, not applicable.

ever been diagnosed with a disease, 43% reported at least one condition inconsistently (ie, reported 'ever' been diagnosed at the baseline, but reported 'never' at the follow-up survey; table 2). Twenty-eight per cent reported one condition inconsistently, while 5% of the participants had three or more inconsistent self-reported conditions (table 2). Among 10 disease categories excluding haematological conditions that were rare in the study cohort, rheumatological conditions had the highest likelihood of inconsistency (29% of the subjects reported inconsistently between two surveys), and the lowest likelihood of inconsistency was for eye diseases (10%; table 2). Among individual conditions with at least 5% participants self-reported at the baseline survey, rheumatoid arthritis had the highest inconsistency rate (40%; online supplementary table 1).

Demographic characteristics for inconsistent self-reporting

We observed that some demographic characteristics such as older age and poorer perceived general health were associated with a higher proportion of subjects with inconsistent self-reporting. Univariately, a higher inconsistency rate was associated with older age (OR=2.7 (95% CI 2.4 to 3.1) for 65 years or older vs the youngest group (18–44 years) and lower education (OR=1.2, 95% CI 1.0 to 1.4)). In terms of perceived general health, participants reported poor/fair health in each survey (OR=1.8 (95% CI 1.5 to 2.1) in baseline and OR=2.0 (95% CI 1.6 to 2.5) in the follow-up survey, compared with at least good health). In addition, compared with those with consistently good general health, the participants reporting poor/fair general health in at least one survey

had a higher inconsistent self-reporting (eg, OR=2.3 (95% CI 1.6 to 3.3) for comparing those with worse in the follow-up survey vs consistently good general health; table 3).

HOUSES index for inconsistent self-reporting

Low SES measured by the HOUSES index was significantly associated with a higher likelihood of inconsistent self-reports for prevalent health conditions. Univariately, lower HOUSES deciles (lower SES) was associated with a higher likelihood of inconsistency (OR=1.6 (95% CI 1.3 to 2.0) for the highest vs the lowest decile; table 3) in a dose–response manner. Adjusting for pertinent variables such as age, perceived general health and education level, the lowest HOUSES index was independently associated with a higher likelihood of inconsistency (OR=1.5 (95% CI 1.2 to 1.8) for the lowest vs the highest HOUSES decile; table 3). In addition to the HOUSES index, age and perceived general health had independent residual effect for inconsistent self-reporting, while education level showed no association.

DISCUSSION

In this cohort study, we found socioeconomic disparities as measured by the HOUSES index exist in reporting health conditions consistently when asked repeatedly, as well as the response rate for repeated surveys. Low HOUSES index (ie, low SES at the time of surveys) based on current housing characteristics is associated with higher likelihood of inconsistent reporting of prevalent diseases. The association remained significant, even after

Table 3 Association between sociodemographic characteristics and inconsistent self-reports

	At least one inconsistent self-report, n (%)	Univariate analysis*, OR (95% CI)	Multivariate analysis, OR (95% CI)
Age at enrolment			
18–44 years	466 (14.6)	Ref	Ref
45–54 years	542 (17.0)	1.27 (1.10 to 1.47)	1.29 (1.11 to 1.50)
55–64 years	714 (22.4)	1.78 (1.55 to 2.06)	1.81 (1.56 to 2.10)
≥65 years	1471 (46.1)	2.69 (2.37 to 3.06)	2.56 (2.23 to 2.95)
Sex			
Male	1201 (37.6)	Ref	Ref
Female	1992 (62.4)	1.06 (0.96 to 1.16)	1.04 (0.95 to 1.15)
Race			
Whites	3023 (94.7)	Ref	Ref
Asians	37 (1.2)	0.79 (0.53 to 1.15)	0.81 (0.54 to 1.20)
Blacks	17 (0.5)	1.61 (0.82 to 3.11)	1.36 (0.69 to 1.64)
Others	116 (3.6)	1.30 (1.00 to 1.68)	1.25 (0.96 to 1.64)
Education			
High school or less	524 (16.5)	1.20 (1.04 to 1.38)	1.09 (0.94 to 1.26)
Some college	1055 (33.3)	1.08 (0.98 to 1.19)	1.01 (0.91 to 1.13)
College+	1593 (50.2)	Ref	Ref
HOUSES			
H1	267 (8.4)	1.64 (1.32 to 2,04)	1.46 (1.17 to 1.84)
H2	330 (10.3)	1.25 (1.02 to 1.54)	1.18 (0.96 to 1.46)
H3	358 (11.2)	1.29 (1.06 to 1.58)	1.24 (1.01 to 1.52)
H4	361 (11.3)	1.35 (1.10 to 1.65)	1.28 (1.04 to 1.57)
H5	354 (11.1)	1.25 (1.03 to 1.54)	1.19 (0.97 to 1.47)
H6	271 (8.5)	0.91 (0.73 to 1.12)	0.87 (0.70 to 1.07)
H7	331 (10.4)	1.09 (0.90 to 1.34)	1.06 (0.87 to 1.31)
H8	321 (10.1)	1.04 (0.85 to 1.28)	1.04 (0.85 to 1.27)
H9	315 (9.9)	1.08 (0.88 to 1.32)	1.08 (0.88 to 1.32)
H10	285 (8.9)	Ref	Ref
Perceived general health changes between surveys			
Consistently poor	124 (3.9)	1.95 (1.48 to 2.56)	1.76 (1.33 to 2.32)
Improved	192 (6.1)	1.72 (1.38 to 2.14)	1.64 (1.32 to 2.05)
Worse	81 (2.6)	2.31 (1.63 to 3.30)	2.17 (1.53 to 3.11)
Consistently good	2760 (87.4)	Ref	Ref
Perceived health at baseline			
Excellent/very good/good	2848 (90.0)	Ref	NA
Poor/fair	318 (10.0)	1.78 (1.50 to 2.12)	
Perceived health at follow-up			
Excellent/very good/good	2985 (93.4)	Ref	NA
Poor/fair	208 (6.5)	2.02 (1.63 to 2.51)	

*All univariate analyses for risk of inconsistent self-reporting were adjusted for age (except for age variable).
HOUSES, HOUsing-based SocioEconomic Status; NA, not applicable.

accounting for education attainment level, a commonly used SES measure, which implies that HOUSES may measure additional socioeconomic characteristics that education level cannot capture, such as functional declines, social isolation and health literacy.[36] In addition, as the HOUSES index uses housing information close to the time of surveys, it will likely reflect current socioeconomic situation that may affect the quality of surveys.

We also found that subjects with low HOUSES index are less likely to respond to the follow-up survey which may further distort the study findings due to participation bias associated with SES.

In addition to relatively well-established risk factors such as older age and poor perceived general health,[20] low HOUSES index was associated with a higher likelihood of having inconsistent self-reported information, independent of educational level. This supports our hypothesis that subjects with low SES at the time of surveys will have less reliable self-reported information, potentially due to low health literacy. Although health literacy and educational attainment are strongly associated, controlling for education does not fully account for the effect of health literacy on outcomes.[37] Therefore, it is not surprising to see a residual effect of the HOUSES index on inconsistency, independent of education level that remains unchanged once acquired. In addition, an accumulated asset-based SES measure (eg, the HOUSES index) is suggested to be a more appropriate measure for assessing socioeconomic position for older people such as our study subjects, compared with education level.[38] This study finding has both research and clinical ramifications. From the research perspective, self-reported illness from certain groups (like older participants, poor general health or those with low current SES) may be more likely to report inconsistent and/or inaccurate health information which could bias results or conclusions from studies. Clinically, self-reporting is commonly used to collect medical history, especially for patients seen in consultation, as health information from other medical facilities are usually unavailable. Therefore, this same group of people may be more likely to provide less meaningful health information when self-reported. This can also affect potential population-based health management to address health disparities derived from self-report.

In terms of predictors associated with the response rate when invited to the follow-up survey, we found that subjects with low HOUSES index were less likely to respond to the survey. Specifically, roughly 50% of the subjects having the lowest HOUSES index did not respond to the follow-up survey, compared with one-third of those with the highest HOUSES index. In addition, those with younger age, lower self-rated health and lower education levels were less likely to respond to the survey. Interestingly, we observed that subjects who did not report education attainment at the baseline survey were far less likely to respond to the follow-up survey. This observation implies that study findings based only on subjects responding to the follow-up survey might be biased by SES (ie, the study findings may not sufficiently reflect subjects with low SES).

There are several limitations to this study. First, selection bias may exist as this study is based on subjects who originally participated in the MCB. Compared with patients receiving primary care at Mayo Clinic, the MCB participants are older and sicker which may affect the degree of inconsistency rates.[39] Second, HOUSES index is available only for those residing in Olmsted County, Minnesota, and thus the generalisability of the study finding may be limited to geographic areas having similar subject characteristics. Studies have shown that Olmsted County adult residents have similar characteristics to whites residing in Upper Midwest states.[40] Third, this study characterised the likelihood of inconsistent self-reports between two surveys, without known true medical status (ie, no gold standard). That is, improved health literacy may lead to more accurate reporting of one's health history. While true disease status is required to fully understand the nature of inconsistency, it is impractical to obtain true disease status for all 63 conditions in a large-scale study like this. In addition, some conditions may not be ascertained correctly even after reviewing medical records, especially when patients were cured of a particular condition and/or symptom-free for a long period of time (ie, no relevant medical records with the condition from recent years given that the MCB participants are relatively old). Furthermore, disease histories are often collected through self-reporting during the medical visit (ie, even medical records may not provide the true disease status). However, future studies may use more efficient approaches to obtain true disease stratus. Additionally, the likelihood of inconsistent self-reporting may be influenced by severity of individual diseases (ie, severe conditions such as breast cancer may have lower inconsistency rate, compared with more transient conditions such as migraine). Finally, clinical illnesses that do not use tissues and/or laboratory tests for disease diagnosis may have higher likelihood of inconsistency.

CONCLUSION

The current study presented that socioeconomic disparities exist in inconsistent self-reporting in longitudinal studies and response rates. Therefore, studies using self-reports may consider additional effort to account for these biases, and the study results should be interpreted with caution. As the degree of health disparities is frequently assessed by self-reported survey, the influence of current social positions poses important implications on clinical practice and research.

REFERENCES

1. Short ME, Goetzel RZ, Pei X, *et al*. How accurate are self-reports? Analysis of self-reported health care utilization and absence when compared with administrative data. *J Occup Environ Med* 2009;51:786–96.
2. Bielinski SJ, Pathak J, Carrell DS, *et al*. A robust e-epidemiology tool in phenotyping heart failure with differentiation for preserved and reduced ejection fraction: the electronic medical records and genomics (eMERGE) network. *J Cardiovasc Transl Res* 2015;8:475–83.
3. Wi CI, Sohn S, Rolfes MC, *et al*. Application of a natural language processing algorithm to asthma ascertainment. an automated chart review. *Am J Respir Crit Care Med* 2017;196:430–7.
4. Doan S, Maehara CK, Chaparro JD, *et al*. Building a natural language processing tool to identify patients with high clinical suspicion for kawasaki disease from emergency department notes. *Acad Emerg Med* 2016;23:628–36.
5. Bourgeois FT, Porter SC, Valim C, *et al*. The value of patient self-report for disease surveillance. *J Am Med Inform Assoc* 2007;14:765–71.
6. Rabin LA, Wang C, Katz MJ, *et al*. Predicting Alzheimer's disease: neuropsychological tests, self-reports, and informant reports of cognitive difficulties. *J Am Geriatr Soc* 2012;60:1128–34.
7. Newell SA, Girgis A, Sanson-Fisher RW, *et al*. The accuracy of self-reported health behaviors and risk factors relating to cancer and cardiovascular disease in the general population: a critical review. *Am J Prev Med* 1999;17:211–29.
8. Stewart J, Sanson-Fisher R, Eades S. Aboriginal and Torres Strait Islander health: accuracy of patient self-report of screening for diabetes, high cholesterol and cervical cancer. *Aust N Z J Public Health* 2016;40(Suppl 1):S3–6.
9. Smith B, Chu LK, Smith TC, *et al*. Challenges of self-reported medical conditions and electronic medical records among members of a large military cohort. *BMC Med Res Methodol* 2008;8:37.
10. Okura Y, Urban LH, Mahoney DW, *et al*. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *J Clin Epidemiol* 2004;57:1096–103.
11. Cumberland PM, Chianca A, Rahi JS, *et al*. Accuracy and utility of self-report of refractive error. *JAMA Ophthalmol* 2016;134:794–801.
12. Boissonnault WG, Badke MB. Collecting health history information: the accuracy of a patient self-administered questionnaire in an orthopedic outpatient setting. *Phys Ther* 2005;85:531–43.
13. Leikauf J, Federman AD. Comparisons of self-reported and chart-identified chronic diseases in inner-city seniors. *J Am Geriatr Soc* 2009;57:1219–25.
14. Braveman PA, Cubbin C, Egerter S, *et al*. Socioeconomic disparities in health in the United States: what the patterns tell us. *Am J Public Health* 2010;100 Suppl 1:S186–S196.
15. Vacek PM, Mickey RM, Worden JK. Reliability of self-reported breast screening information in a survey of lower income women. *Prev Med* 1997;26:287–91.
16. Zajacova A, Dowd JB, Schoeni RF, *et al*. Consistency and precision of cancer reporting in a multiwave national panel survey. *Popul Health Metr* 2010;8:20.
17. Poitras ME, Fortin M, Hudon C, *et al*. Validation of the disease burden morbidity assessment by self-report in a French-speaking population. *BMC Health Serv Res* 2012;12:35.
18. Nelson NW, Anderson CR, Thuras P, *et al*. Factors associated with inconsistency in self-reported mild traumatic brain injury over time among military personnel in Iraq. *Br J Psychiatry* 2015;206:237–44.
19. Bancej CM, Maxwell CJ, Snider J. Inconsistent self-reported mammography history: findings from the National Population Health Survey longitudinal cohort. *BMC Health Serv Res* 2004;4:32.
20. Cigolle CT, Nagel CL, Blaum CS, *et al*. Inconsistency in the self-report of chronic diseases in panel surveys: developing an adjudication method for the health and retirement study. *J Gerontol B Psychol Sci Soc Sci* 2016:gbw063.
21. Davern M, Rodin H, Beebe TJ, *et al*. The effect of income question design in health surveys on family income, poverty and eligibility estimates. *Health Serv Res* 2005;40:1534–52.
22. Ryder AB, Wilkinson AV, McHugh MK, *et al*. The advantage of imputation of missing income data to evaluate the association between income and self-reported health status (SRH) in a Mexican American cohort study. *J Immigr Minor Health* 2011;13:1099–109.
23. Juhn YJ, Beebe TJ, Finnie DM, *et al*. Development and initial testing of a new socioeconomic status measure based on housing data. *J Urban Health* 2011;88:933–44.
24. Takahashi PY, Ryu E, Hathcock MA, *et al*. A novel housing-based socioeconomic measure predicts hospitalisation and multiple chronic conditions in a community population. *J Epidemiol Community Health* 2016;70:286–91.
25. Wi CI, St Sauver JL, Jacobson DJ, *et al*. Ethnicity, socioeconomic status, and health disparities in a mixed rural-urban us community-olmsted county, minnesota. *Mayo Clin Proc* 2016;91:612–22.
26. Hu D, Flick RP, Gleich SJ, *et al*. Construction and characterization of a population-based cohort to study the association of anesthesia exposure with neurodevelopmental outcomes. *PLoS One* 2016;11:e0155288.
27. Ryu E, Juhn YJ, Wheeler PH, *et al*. Individual housing-based socioeconomic status predicts risk of accidental falls among adults. *Ann Epidemiol* 2017;27:415–20.
28. Clouston SAP, Manganello JA, Richards M. A life course approach to health literacy: the role of gender, educational attainment and lifetime cognitive capability. *Age Ageing* 2017;46:493–9.
29. In: Nielsen-Bohlman L, Panzer AM, Kindig DA, eds. *Health literacy: a prescription to end confusion*. Washington(DC), 2004.
30. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. *Fam Med* 2004;36:588–94.
31. Williams MV, Baker DW, Parker RM, *et al*. Relationship of functional health literacy to patients' knowledge of their chronic disease. A study of patients with hypertension and diabetes. *Arch Intern Med* 1998;158:166–72.
32. Gazmararian JA, Williams MV, Peel J, *et al*. Health literacy and knowledge of chronic disease. *Patient Educ Couns* 2003;51:267–75.
33. Olson JE, Ryu E, Johnson KJ, *et al*. The Mayo Clinic Biobank: a building block for individualized medicine. *Mayo Clin Proc* 2013;88:952–62.
34. Takahashi PY, Ryu E, Olson JE, *et al*. Health behaviors and quality of life predictors for risk of hospitalization in an electronic health record-linked biobank. *Int J Gen Med* 2015;8:247–54.
35. Ryu E, Takahashi PY, Olson JE, *et al*. Quantifying the importance of disease burden on perceived general health and depressive symptoms in patients within the mayo clinic biobank. *Health Qual Life Outcomes* 2015;13:95.
36. Svedberg P, Nygren JM, Staland-Nyman C, *et al*. The validity of socioeconomic status measures among adolescents based on self-reported information about parents occupations, FAS and perceived SES; implication for health related quality of life studies. *BMC Med Res Methodol* 2016;16:48.
37. Yost KJ, DeWalt DA, Lindquist LA, *et al*. The association between health literacy and indicators of cognitive impairment in a diverse sample of primary care patients. *Patient Educ Couns* 2013;93:319–26.
38. Galobardes B, Lynch J, Smith GD. Measuring socioeconomic position in health research. *Br Med Bull* 2007;81-82:21–37.
39. Takahashi PY, Ryu E, Olson JE, *et al*. Hospitalizations and emergency department use in mayo clinic biobank participants within the employee and community health medical home. *Mayo Clin Proc* 2013;88:963–9.
40. St Sauver JL, Grossardt BR, Leibson CL, *et al*. Generalizability of epidemiological findings and public health decisions: an illustration from the rochester epidemiology project. *Mayo Clin Proc* 2012;87:151–60.