BMJ Open

Utilization of a modified housing-based socioeconomic status measure in health disparities research

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-011564
Article Type:	Research
Date Submitted by the Author:	17-Feb-2016
Complete List of Authors:	Ryu, Euijung; Mayo Clinic, Health Sciences Research Wi, Chung-Il ; Mayo Clinic, Pediatric and Adolescent Medicine Crow, Sheri; Mayo Clinic, Pediatric and Adolescent Medicine Armasu, Sebastian; Mayo Clinic, Health Sciences Research Wheeler, Philip; Mayo Clinic, Pediatric and Adolescent Medicine Sloan, Jeff; MAYO clinic Yawn, Barbara; Olmsted Medical Center, Department of Research Beebe, Timothy ; Mayo Clinic Williams, Arthur; James A. Haley Veterans' Hospital Juhn, Young; Mayo Clinic, Community Pediatrics and Adolescent Medicine
Primary Subject Heading:	Paediatrics
Secondary Subject Heading:	Public health
Keywords:	socioeconomic status, HOUSES, health disparity

Utilization of a modified housing-based socioeconomic status measure in health disparities research

Euijung Ryu [1,*], Chung-Il Wi [2,*], Sheri S. Crow [1,2], Sebastian M. Armasu [1], Philip H. Wheeler [2], Jeff A. Sloan [1], Barbara P. Yawn [3], Timothy J. Beebe [1], Arthur R. Williams[4], Young J. Juhn [1,2]

[1] Department of Health Sciences Research, Mayo Clinic, Rochester, MN;[2] Department of Pediatric and Adolescent Medicine, Mayo Clinic, Rochester, MN; [3] Olmsted Medical Center, Rochester, MN; [4]Center of Innovation on Disability and Rehabilitation Research (CINDRR), James A. Haley Veterans Affairs Hospital, Tampa, FL; [5] Department of Health Administration and Policy, George Mason University, Fairfax, VA

* These authors contributed equally

Correspondence to: Young J. Juhn, MD, MPH
Professor of Pediatrics
Department of Pediatric and Adolescent Medicine, Mayo Clinic
200 First Street SW,
Rochester MN, 55905
E-mail: Juhn.Young@mayo.edu
Phone: 507-538-1642
Fax: 507-284-9744

Keywords: Socioeconomic status, HOUSES, health disparity

Word count: 2962

ABSTRACT

Objective: Socioeconomic status (SES) is a well-established risk factor for many health outcomes. Recently, we developed an SES measure based on four housing-related characteristics (termed HOUSES) and demonstrated its ability to assess health disparities. In this study, we aimed to evaluate whether fewer housing-related characteristics could be utilized to provide a similar representation of SES status.

Method: We performed a cross-sectional study using parents/guardians of children aged 1-17 years from two US Midwestern counties (n=728 and n=701). For each subject, housing-related characteristics used in the formulation of HOUSES (assessed housing value, square footage, number of bedrooms, and number of bathrooms) were obtained from the local government assessor's offices, and additional SES measures and health outcomes with known associations to SES (obesity, low birth weight, and smoking exposure) were collected from a telephone survey. Housing characteristics with the greatest contribution for predicting the health outcomes were added to formulate a modified HOUSES index.

Results: Among the four housing characteristics used in the original HOUSES, the strongest contributions for predicting health outcomes were observed from assessed housing value and square footage (combined contribution ranged between 89% and 96%). Based on this observation, these two were used to calculate a modified HOUSES index. Correlation between modified HOUSES and other SES measures was comparable to the original HOUSES for both locations (e.g., 0.43 [original HOUSES] vs 0.49 [modified HOUSES] with family annual income). Consistent with the original HOUSES formula, the strongest association with modified HOUSES was observed with smoking exposure (OR=0.24 with 95% CI 0.11–0.49 for comparing subjects in highest HOUSES [Q4] vs lowest [Q1] group; overall $p < 0.001$).

Conclusion: The modified HOUSES requires only two readily available housing characteristics thereby improving the feasibility of using this index as a proxy for SES in multiple communities, especially in the US Midwestern region.

Strengths and limitations of this study:

- Robust performance of the modified HOUSES demonstrates its application to different geographic regions with minimal property data.
- The modified HOUSES index does not rely on the quality of imputation of missing housing characteristics.
- This study is limited by self-reported health outcomes.
- The study is limited by testing the modified index in only two locations.
- This study may not work well in other countries where housing data is not routinely collected or not made publically available.

INTRODUCTION

The impact of socioeconomic status (SES) on health outcomes has been well documented in the US and elsewhere and included in assessments of frameworks for health disparities.[1-6] Overall, these frameworks for health disparities suggest that distal factors such as individual SES impact human health independently, jointly, and interactively with proximal factors (e.g., genetic predisposition or biological responses). Thus, SES and its definition and method of calculation can have important consequences on clinical practice, research, and health policy concerning health disparities.

SES reflects multifaceted assets or capacities of humans including materialistic, human, and social capital, making accurate measurement of SES a potentially formidable task that could be compared to the challenges in discovering biomarkers accurately predicting human diseases. One of the biggest challenges in health disparities research is the complexity of reporting individual's SES using commonly available data such as information found in medical records and administrative databases. To address this important gap in the ability to operationalize an SES definition using data available to health care researchers, our research group developed an individual housing-based SES measure termed HOUSES. HOUSES is a composite index consisting of assessed housing value, square footage, and the numbers of bedrooms and bathrooms available from property data found in the assessor's office of the county government. Using address information documented in medical records or administrative datasets linked to the county assessor's data, we were able to calculate an effective SES proxy without the need for specific educational or income levels which are rarely available in the medical record or administrative data. The HOUSES index predicts health outcomes in both adults and children that have previously been identified to be associated with SES (low birth weight, obesity, smoking exposure, asthma control status, pneumococcal diseases, post-Myocardial infarction mortality, and rheumatoid arthritis development and mortality risk).[7-13]

One of the challenges in calculating the original HOUSES index is the need for complex assessor's real property data generated for US taxation purpose. However, this data often does not include key variables of interest such as the number of bedrooms and bathrooms. For example, the 2013 real property data of Olmsted County, Minnesota, have 3 - 6% missing data on the number of bedrooms and bathrooms of single family housing units while assessed housing value and square footage is almost complete (<1% missing data). The rate of missing information on the number of bedrooms and bathrooms tend to vary depending on the age of the real property data and/or geographic regions of interest.

To address this concern, we explored whether the original HOUSES index could be modified using fewer housing-related characteristics, especially assessed housing value and square footage as these two components are consistently available in most counties and state property databases. Our aim was to evaluate whether a modified HOUSES index would provide an equivalent representation of SES status to the original HOUSES index.

METHODS

Study subjects and design: The original study enrolled parents/guardians of children aged 1 – 17 years living in Olmsted County, MN (n=746) or Jackson County, MO (n = 704) in 2006. The present study included those who had both successful geocoding of address with real property data and formulation of HOUSES index (728 subjects for Olmsted County, MN, and 701 subjects for Jackson County, MO). Detailed description of the study population and methodology for developing and validating the HOUSES index were previously reported by Juhn et al.[8] Briefly, subjects were originally recruited for the HOUSES derivation and validation study.[8] Data collection included sociodemographic characteristics and health-related information obtained through survey research. This information was then linked to the property data associated with each subject's address. Property data were acquired from the county assessor's offices. For comparison, additional SES measures were included in the survey questionnaire,

to describe parental education level, family annual income, Hollingshead index (a family's composite index using education, occupation, sex, and marital status), and Nakao-Treas index (composite index using educational attainment and income of job incumbents corresponding to the 1980 census).[14-16] Correlation of HOUSES with the health-related outcomes of childhood obesity, low birth weight, and smoking exposure were evaluated because the association between these outcomes and SES has been well demonstrated. [17-19] The original HOUSES was formulated using four housing-related characteristics (assessed housing value, square footage, number of bedrooms, and number of bathrooms).

To assess whether fewer housing-related characteristics could be used, the relative influence (RI) of each characteristic for predicting the health outcomes (obesity, low birth weight, and smoking exposure) was estimated using gradient boosting machine (GBM) models under logistic regression model frameworks. The GBM modeling approach is a machine learning technique for building a multivariable prediction model by incorporating all of the variables without variable selection.[20-22] RI is a measure of a given variable's importance, relative to that of other variables, in the model prediction process. The measure is based on the number of times a variable is selected for splitting in a decision tree, weighted by the improvement of the model fitting as a result of the split and further standardized so that the sum of RI from all variables adds up to 100%. The higher the RI value (maximum of 100%) of a characteristic, the more significant its contribution is to the model. Those with the greatest contribution were summed to formulate a modified HOUSES index for each county.

Statistical analysis: Descriptive statistics were used to represent socio-demographic characteristics and health-related outcomes for subjects for each county. Pearson correlation coefficients were calculated for correlations between original HOUSES and the modified HOUSES indices. For further analysis, both original HOUSES and modified HOUSES scores were collapsed into 4 groups using quartiles (Q1 [lowest] – Q4 [highest]). Spearman correlation coefficients were calculated to see whether correlations with other SES measures (parental education, family annual income, Hollingshead index, and Nakao-Treas index) were similar between original HOUSES. In addition, logistic regression models were used

to assess the association of the modified HOUSES and risk of health-related outcomes (obesity, low birth weight, and smoking exposure), using Q1 as a reference category.

RESULTS

Characteristics of subjects: The results are summarized in Table 1. A total of 728 children from Olmsted County, Minnesota, and 701 children from Jackson, Missouri, were included in the study analysis. The median age of children included in the study was 10 years (25^{th} – 75^{th} percentile: 5 - 14) with roughly 50% females in both counties (Table 1). Residents of Olmsted County, MN were more likely to have higher levels of education and income than residents of Jackson County, MO. Obesity rates (15% vs 12%) and smoking exposure (27% vs 12%) were higher in Jackson County, while the rate of low birth weight was higher in Olmsted County (11% vs 6.5%).

Table 1: Sociodemographic characteristics and health outcomes of the study subjects

	Olmsted County, MN (n = 728)	Jackson County, MO (n = 701)
Demographic characteristics		
Age (years), median (25^{th} – 75^{th} %tiles)	10 (5 - 14)	10 (5 - 14)
Sex, female N (%)	358 (49%)	355 (51%)
Socioeconomic characteristics		
Parents' education		
Less than high school education	4 (0.5%)	18 (2.6%)
High~~s~~ school graduate	37 (5.1%)	103 (15%)
Some college, no degree	140 (19%)	173 (25%)
Associate/college degree	291 (40%)	229 (33%)
Graduate or professional degree	256 (35%)	178 (25%)
Family annual income		
Less than $24,999	9 (1.3%)	51 (7.8%)
$25,000 to $49,999	86 (12%)	139 (21%)
$50,000 to $74,999	144 (20%)	154 (24%)
$75,000 to $99,999	161 (22%)	136 (21%)
Over $100,000	316 (44%)	175 (27%)
Hollingshead index		
8 to 19	2 (0.3%)	2 (0.3%)
29 to 29	12 (1.6%)	35 (5.0%)
30 to 39	54 (7.4%)	109 (16%)
40 to 54	254 (35%)	268 (38%)
55 to 66	406 (56%)	287 (41%)
Nakao-Treas index		
0 to 12.5	0 (0.0%)	0 (0.0%)

12.6 to 25.1	2 (0.3%)	5 (0.7%)
25.2 to 37.7	53 (7.3%)	101 (14%)
37.8 to 50.3	79 (11%)	105 (15%)
50.4 to 62.9	109 (15%)	105 (15%)
63.0 to 75.5	184 (25%)	192 (27%)
75.6 to 88.1	230 (32%)	156 (22%)
88.2 to 100	71 (9.8%)	37 (5.3%)
Health outcomes		
Obesity, N (%)	71 (12%)	81 (15%)
Low birth weight, N (%)	78 (11%)	43 (6.5%)
Smoking exposure, N (%)	89 (12%)	188 (27%)

Identification of housing characteristics with most significant contributions for health outcomes: The results are summarized in Figure 1. Simultaneously considering all four housing characteristics used in the original HOUSES formula, assessed housing value and square footage had the greatest contribution to all three health outcomes in both counties (Figure 1). The combined contribution of assessed housing value and square footage ranged between 89% and 96% depending on health outcome evaluated. The contribution of number of bedrooms and bathrooms is negligible in the presence of assessed housing value and square footage (< 10% for all three outcomes). Therefore, a modified HOUSES index was formulated by summing z-scores for assessed housing value and square footage once normalized for the remaining analysis.

Correlation of modified HOUSES index with the original HOUSES index and other SES measures, and health outcomes: The results are summarized in Tables 2 and 3. The modified HOUSES index was highly correlated with original HOUSES based on 4 housing components (0.87 in Olmsted County, and 0.93 in Jackson County). As expected, correlations between the modified HOUSES index with other SES measures were comparable to the original HOUSES index for both counties (e.g., 0.43 [original HOUSES] vs 0.49 [modified HOUSES] with family annual income in Olmsted County, Table 2). Overall, the modified index performs similar to the original HOUSES index in inverse associations with health outcomes. The strongest health-related outcome association with the modified HOUSES Index was observed for smoking exposure (OR: 0.24 with 95% CI: 0.11 – 0.49 in Olmsted County and OR: 0.26

with 95% CI 0.16 – 0.44 in Jackson County for comparing subjects in highest [Q4] vs lowest [Q1] group; overall p < 0.001).

Table 2: The correlation between other SES measures and original and modified HOUSES: both original and modified HOUSES are comparable in correlation with other SES measures.

	Olmsted County, MN		Jackson County, MO	
	Original HOUSES	Modified HOUSES	Original HOUSES	Modified HOUSES
Parents' education	0.23	0.26	0.39	0.42
Family annual income	0.43	0.49	0.52	0.55
Hollingshead index	0.19	0.22	0.36	0.38
Nakao-Treas index	0.24	0.26	0.39	0.40

Table 3: Association between modified HOUSES and risk of childhood obesity, low birthweight, and smoking exposure

	Obesity, OR [95% CI]	Low birth weight, OR [95% CI]	Smoking exposure, OR [95% CI]
Olmsted County	[overall p= 0.01]	[overall p= 0.82]	[overall p < 0.001]
Q1	Reference	Reference	Reference
Q2	0.62 [0.33 – 2.28]	1.22 [0.51 – 2.91]	0.39 [0.21 – 0.73]
Q3	0.29 [0.14 – 0.63]	1.02 [0.41 – 2.51]	0.71 [0.41 – 1.22]
Q4	0.48 [0.25 – 0.94]	0.78 [0.30 – 2.03]	0.24 [0.11 – 0.49]
Jackson County	[overall p= 0.11]	[overall p= 0.08]	[overall p < 0.001]
Q1	Reference	Reference	Reference
Q2	0.58 [0.30 – 1.13]	0.42 [0.18 – 1.00]	0.60 [0.38 – 0.94]
Q3	0.75 [0.40 – 1.41]	0.38 [0.16 – 0.94]	0.47 [0.30 – 0.75]
Q4	0.45 [0.23 – 0.89]	0.54 [0.34 – 1.21]	0.26 [0.16 – 0.44]

DISCUSSION

This "modified" HOUSES index was associated with SES-related health outcomes, such as smoking exposure and obesity status. Also, the modified HOUSES was strongly correlated with the original four component HOUSES (assessed housing value, square footage, number of bedrooms and bathrooms) calculation in both counties evaluated. As the modified HOUSES index relies on only two housing-related variables (assessed housing value and square footage) that are consistently well captured within

publicly available real property data, it may be a suitable tool for health disparities research throughout diverse communities within the United States.

The modified HOUSES performs similarly to the original HOUSES in terms of correlating reasonably with other SES measures [Jackson County, Missouri: r=0.38-0.55 (modified) vs. 0.36-0.52 (original). Olmsted County, Minnesota: r=0.22-0.49 (modified) vs. 0.19-0.43 (original)] (Table 2). As for construct validity, which is regarding the association with health outcomes, smoking exposure status at home was strongly correlated with modified HOUSES although heterogeneity in the result with regard to study sites was noted for obesity. In support of these findings, assessed housing value and square footage alone showed the majority of contribution on the prediction and further supported our interpretation that assessed housing value and square footage might play a more significant role in predicting health outcomes of interest assessed than the number of bedrooms and bathrooms.

There are a few possible explanations for the relatively minimal influence of bedroom and bathroom count. First, the variance of assessed housing values and square footage in a continuous variable is larger than the number of bedrooms and bathrooms, a discrete variable with a small range. Therefore, it is possible that the impact of bathroom and bathroom counts may be minimal once assessed housing value and square footage are considered. Second, as missing information for bedroom count and bathroom count were imputed with the mean value in the original HOUSES index, their impact (variance) might have been reduced, compared to assessed housing value and square footage with minimal missing values. Finally, while positive correlation exists among number of bedrooms, square footage, number of bathrooms, and assessed housing value, conceptually, we postulate that assessed housing value and square footage might be better correlated with other SES measures than the number of bedrooms and bathrooms. For example, assessed housing value (r=0.52) and square footage (r=0.48) are more closely correlated with income, compared to the self-reported number of bedrooms (r=0.18) and bathrooms (r=0.37). In support of this postulation, in the original study, regardless of whether we used the imputed data or self-reported data on the number of bedrooms and bathrooms, these two different estimates did not

significantly influence the number of factors and factor loadings for original HOUSES index. Thus, the number of bedrooms and bathrooms conceptually seems to only add a finer granularity in capturing housing-based SES. Therefore, based on these conceptual and methodological aspects of the housing features, the modified HOUSES index might perform as well as the original HOUSES index.

There are a few noteworthy findings in our study. It would be important to determine whether removing two components (the number of bedrooms and bathrooms) from the original HOUSES results in differential estimate of SES between higher SES and lower SES group. We found that using two components (assessed housing value and square footage) might slightly overestimate the SES of individuals with lower SES while it might slightly underestimate the SES of individuals with higher SES, resulting in reducing variance overall (data not shown). This is not surprising given that numbers of bedrooms and bathrooms are discrete variables with narrower ranges of variance as opposed to assessed housing value and square footage, which are continuous variables with wider ranges of variance. In this regard, if the primary aim of studies is to compare outcomes which are less responsive to change of SES or such studies are being conducted in a community with (relatively) socioeconomic homogeneity, the original HOUSES index might be preferred as it detects SES with precision.

The paucity of SES-related data is a common but major challenge for existing large-scale datasets (e.g., disease registry, administrative dataset, etc.) that would otherwise be valuable for conducting health disparities research. Therefore, advantages of housing-based SES (both original and modified HOUSES) are promising as the address information, which is almost routinely collected in health care settings (e.g., medical records), is directly linked or geocoded to real property data to formulate the HOUSES index. However, considering that there are high missing rates of number of bedrooms and bathrooms in assessor's real property data or even unavailability of these two components in some regions, our study findings provide an important basis for using the modified HOUSES index as a potential alternative to the original HOUSES index in studying and addressing health disparities. Overall, the modified HOUSES

index provides an alternative approach for measuring SES when conventional data for characterizing SES is not available.

Strengths of the current study include that it was conducted in two study settings with diverse socioeconomic characteristics observed within each study population (external validity). The robust performance of modified HOUSES in these communities demonstrates that this approach to characterizing SES is both feasible and generalizable. Another strength for calculating the modified HOUSES index is to minimize effort for imputing missing information in the dataset. Limitations include self-report-based health outcomes and are therefore subject to reporting bias. However, the associations between self-reported health outcomes and SES have been well demonstrated in multiple independent investigations. Furthermore, these same health outcomes, defined by physician diagnosis or predetermined criteria, were also significantly associated with our original HOUSES index. [7-13] The performance of modified HOUSES for objective measure-based health outcomes can be expected to be similar to these findings. HOUSES, which is developed based on real property data for US taxation purposes, may not work well in other countries where housing data is not routinely collected or made publicly available in databases, or even in communities within the US where housing assessment is not done frequently or the quality of assessment is poor. Furthermore, the formulation of modified HOUSES was done relying on the relationship among four housing characteristics observed in the two Midwestern counties. Therefore, it is possible that the modified HOUSES may not work as well as the original HOUSES in communities where the relationship among those 4 characteristics is drastically different compared to the two Midwestern counties used in this study.

In conclusion, a modified HOUSES calculation, using two housing-related characteristics (assessed housing value and square footage) instead of four, correlates well with the ability of the original HOUSES index to represent SES status. The two modified HOUSES components of assessed housing value and square footage are well captured in assessor's housing data. As a result, a modified HOUSES improves the feasibility of comprehensively assessing SES status and expanding the application of this

tool into different geographic regions without complete real property data needed for HOUSES, especially in US Midwestern communities.

Acknowledgement: The authors would like to thank Ms. Kelly Okeson for her administrative assistance.

Funding: This work was supported by the National Heart Lung and Blood Institute (R01 HL126667), the National Institute of Allergy and Infectious Diseases (R21 AI101277, R21 AI116839), and T. Denny Sanford Foundation.

Competing interests: None declared

Contributors: ER, CW, and YJJ were responsible for the study design, initial manuscript drafting, and interpretation of the results. The original study design for HOUSES was done by ARW and YJJ. CW, PHW, JAS, BPY, TJB, ARW and YJJ collected the data and were responsible for formulating the original HOUSES. SSC and PHW contributed critically for manuscript drafting. ER and SMA conducted the statistical analysis in this paper. Earlier statistical analyses for the basis of HOUSES were done by ARW. All the authors had approved the final version of the manuscript.

Data sharing statement: Statistical codes and data are available from the authors upon request.

Figure legend:

Figure 1: Relative influence (percent) of 4 housing features (assessed housing value, , number of bedrooms, and number of bathrooms) for risk of obesity, low birthweight, and smoking exposure among subjects from Olmsted county (A) and Jackson county (B).

REFERENCES

1. Diez Roux AV. Conceptual approaches to the study of health disparities. *Annu Rev Public Health* 2012;33:41-58.
2. Warnecke RB, Oh A, Breen N, Gehlert S, Paskett E, Tucker KL, et al. Approaching health disparities from a population perspective: the National Institutes of Health Centers for Population Health and Health Disparities. *Am J Public Health* 2008;98(9):1608-15.
3. Braveman P. Health disparities and health equity: concepts and measurement. *Annu Rev Public Health* 2006;27:167-94.
4. Marmot M. Social determinants of health inequalities. *Lancet* 2005;365(9464):1099-104.
5. Oakes JM, Rossi PH. The measurement of SES in health research: current practice and steps toward a new approach. *Soc Sci Med* 2003;56(4):769-84.
6. Kawachi I, Berkman LF. *Neighborhoods and Health*: Oxford University Press, Inc., 2003.
7. Butterfield MC, Williams AR, Beebe T, Finnie D, Liu H, Liesinger J, et al. A two-county comparison of the HOUSES index on predicting self-rated health. *J Epidemiol Community Health* 2011;65(3):254-9.
8. Juhn YJ, Beebe TJ, Finnie DM, Sloan J, Wheeler PH, Yawn B, et al. Development and initial testing of a new socioeconomic status measure based on housing data. *J Urban Health* 2011;88(5):933-44.
9. Johnson MD, Urm SH, Jung JA, Yun HD, Munitz GE, Tsigrelis C, et al. Housing data-based socioeconomic index and risk of invasive pneumococcal disease: an exploratory study. *Epidemiol Infect* 2013;141(4):880-7.
10. Bang DW, Manemann SM, Gerber Y, Roger VL, Lohse CM, Rand-Weaver J, et al. A Novel Socioeconomic Measure Using Individual Housing Data in Cardiovascular Outcome Research. *Int J Env Res Pub He* 2014;11(11):11597-615.
11. Juhn Y, Krusemark E, Rand-Weaver J, Beebe T, Sloan J, Yawn B, et al. A novel measure of socioeconomic status using individual housing data in health disparities research for asthma in adults. *Allergy* 2014;69:327-27.
12. Harris MN, Lundien MC, Finnie DM, Williams AR, Beebe TJ, Sloan JA, et al. Application of a novel socioeconomic measure using individual housing data in asthma research: an exploratory study. *Npj Prim Care Resp M* 2014;24.
13. Ghawi H, Crowson CS, Rand-Weaver J, Krusemark E, Gabriel SE, Juhn YJ. A novel measure of socioeconomic status using individual housing data to assess the association of SES with rheumatoid arthritis and its mortality: a population-based case-control study. *BMJ Open* 2015;5(4):e006469.
14. Hollingshead A. Four Factor Index of Social Status: New Haven, CT: Yale University Department of Psychology, 1975.
15. Rieppi R, Greenhill LL, Ford RE, Chuang S, Wu M, Davies M, et al. Socioeconomic status as a moderator of ADHD treatment outcomes. *J Am Acad Child Adolesc Psychiatry* 2002;41(3):269-77.
16. Nakao K. The 1989 Socioeconomic index of occupations: Construction from the 1989 Occupational Prestige Scores (General Social Survey Methodological Report No 74): Chicago, IL: University of Chicago, National Opinion Research Center, 1992.
17. Wang Y. Cross-national comparison of childhood obesity: the epidemic and the relationship between obesity and socioeconomic status. *Int J Epidemiol* 2001;30(5):1129-36.
18. Parker JD, Schoendorf KC, Kiely JL. Associations between measures of socioeconomic status and low birth weight, small for gestational age, and premature delivery in the United States. *Ann Epidemiol* 1994;4(4):271-8.

19. Whitlock G, MacMahon S, Vander Hoorn S, Davis P, Jackson R, Norton R. Association of environmental tobacco smoke exposure with socioeconomic status in a population of 7725 New Zealanders. *Tob Control* 1998;7(3):276-80.
20. Reidgeway G. Generalized Boosted Models.
21. Atkinson EJ, Therneau TM, Melton LJ, 3rd, Camp JJ, Achenbach SJ, Amin S, et al. Assessing fracture risk using gradient boosting machine (GBM) models. *J Bone Miner Res* 2012;27(6):1397-404.
22. Natekin A, Knoll A. Gradient boosting machines, a tutorial. *Front Neurorobot* 2013;7:21.

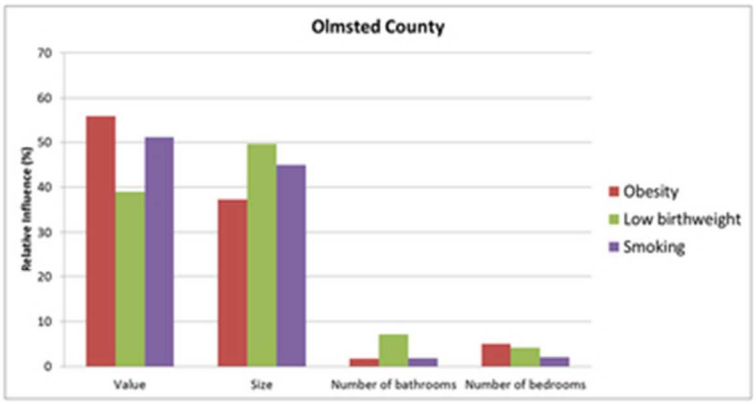

Figure 1: Relative influence (percent) of 4 housing features (assessed housing value, , number of bedrooms, and number of bathrooms) for risk of obesity, low birthweight, and smoking exposure among subjects from Olmsted county (A) and Jackson county (B).
32x17mm (300 x 300 DPI)

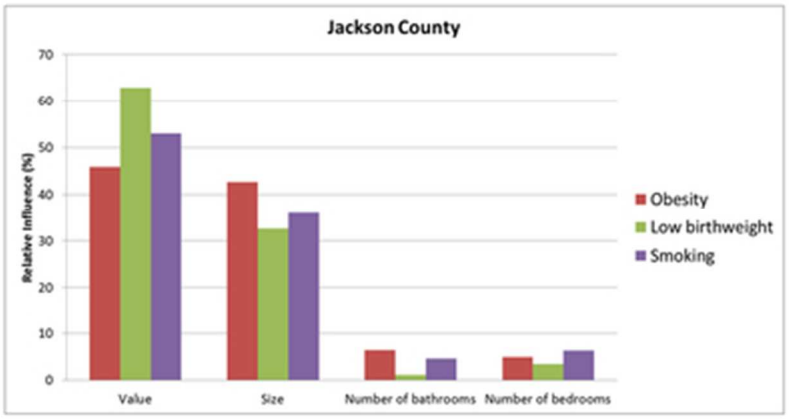

Figure 1: Relative influence (percent) of 4 housing features (assessed housing value, , number of bedrooms, and number of bathrooms) for risk of obesity, low birthweight, and smoking exposure among subjects from Olmsted county (A) and Jackson county (B).
33x18mm (300 x 300 DPI)

BMJ Open

Assessing health disparities in children using a modified housing-related socioeconomic status measure: A cross-sectional study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-011564.R1
Article Type:	Research
Date Submitted by the Author:	18-Apr-2016
Complete List of Authors:	Ryu, Euijung; Mayo Clinic, Health Sciences Research Wi, Chung-Il ; Mayo Clinic, Pediatric and Adolescent Medicine Crow, Sheri; Mayo Clinic, Pediatric and Adolescent Medicine Armasu, Sebastian; Mayo Clinic, Health Sciences Research Wheeler, Philip; Mayo Clinic, Pediatric and Adolescent Medicine Sloan, Jeff; MAYO clinic Yawn, Barbara; Olmsted Medical Center, Department of Research Beebe, Timothy ; Mayo Clinic Williams, Arthur; James A. Haley Veterans' Hospital Juhn, Young; Mayo Clinic, Community Pediatrics and Adolescent Medicine
Primary Subject Heading:	Paediatrics
Secondary Subject Heading:	Public health
Keywords:	socioeconomic status, HOUSES, health disparity

Assessing health disparities in children using a modified housing-related socioeconomic status measure: A cross-sectional study

Euijung Ryu [1,*], Chung-Il Wi [2,*], Sheri S. Crow [1,2], Sebastian M. Armasu [1], Philip H. Wheeler [2], Jeff A. Sloan [1], Barbara P. Yawn [3], Timothy J. Beebe [1], Arthur R. Williams[4,5], Young J. Juhn [1,2]

[1] Department of Health Sciences Research, Mayo Clinic, Rochester, MN;[2] Department of Pediatric and Adolescent Medicine, Mayo Clinic, Rochester, MN; [3] Olmsted Medical Center, Rochester, MN; [4]Center of Innovation on Disability and Rehabilitation Research (CINDRR), James A. Haley Veterans Affairs Hospital, Tampa, FL; [5] Department of Health Administration and Policy, George Mason University, Fairfax, VA

* These authors contributed equally

Correspondence to: Young J. Juhn, MD, MPH
Professor of Pediatrics
Department of Pediatric and Adolescent Medicine, Mayo Clinic
200 First Street SW,
Rochester MN, 55905
E-mail: Juhn.Young@mayo.edu
Phone: 507-538-1642
Fax: 507-284-9744

Keywords: Socioeconomic status, HOUSES, health disparity

Word counts: 2742

ABSTRACT

Objectives: Socioeconomic status (SES) is a well-established risk factor for many health outcomes. Recently, we developed an SES measure based on four housing-related characteristics (termed HOUSES) and demonstrated its ability to assess health disparities. In this study, we aimed to evaluate whether fewer housing-related characteristics could be utilized to provide a similar representation of SES status.

Study setting and participants: We performed a cross-sectional study using parents/guardians of children aged 1-17 years from two US Midwestern counties (n=728 in Olmsted County, Minnesota, and n=701 in Jackson County, Missouri).

Primary and secondary outcome measures: For each subject, housing-related characteristics used in the formulation of HOUSES (assessed housing value, square footage, number of bedrooms, and number of bathrooms) were obtained from the local government assessor's offices, and additional SES measures and health outcomes with known associations to SES (obesity, low birthweight, and smoking exposure) were collected from a telephone survey. Housing characteristics with the greatest contribution for predicting the health outcomes were added to formulate a modified HOUSES index.

Results: Among the four housing characteristics used in the original HOUSES, the strongest contributions for predicting health outcomes were observed from assessed housing value and square footage (combined contribution ranged between 89% and 96%). Based on this observation, these two were used to calculate a modified HOUSES index. Correlation between modified HOUSES and other SES measures was comparable to the original HOUSES for both locations. Consistent with the original HOUSES formula, the strongest association with modified HOUSES was observed with smoking exposure (OR=0.24 with 95% CI 0.11–0.49 for comparing subjects in highest HOUSES vs lowest group; overall p<0.001).

Conclusions: The modified HOUSES requires only two readily available housing characteristics thereby improving the feasibility of using this index as a proxy for SES in multiple communities, especially in the US Midwestern region.

Strengths and limitations of this study:

- Robust performance of the modified HOUSES demonstrates its application to different geographic regions with minimal property data.
- The modified HOUSES index does not rely on the quality of imputation of missing housing characteristics.
- This study is limited by self-reported health outcomes.
- The study is limited by testing the modified index in only two locations.
- This study may not work well in other countries where housing data is not routinely collected or not made publically available.

INTRODUCTION

The impact of socioeconomic status (SES) on health outcomes has been well documented in the US and elsewhere and included in assessments of frameworks for health disparities.[1-5] Overall, these frameworks for health disparities suggest that distal factors such as individual SES impact human health independently, jointly, and interactively with proximal factors (e.g., genetic predisposition or biological responses). Thus, SES and its definition and method of calculation can have important consequences on clinical practice, research, and health policy concerning health disparities.

SES reflects multifaceted assets or capacities of humans including materialistic, human, and social capital, making accurate measurement of SES a potentially formidable task that could be compared to the challenges in discovering biomarkers accurately predicting human diseases. One of the biggest challenges in health disparities research is the complexity of reporting individual's SES using commonly available data such as information found in medical records and administrative databases. To address this important gap in the ability to operationalize an SES definition using data available to health care researchers, our research group developed an individual housing-based SES measure termed HOUSES. HOUSES is a composite index consisting of assessed housing value, square footage, and the numbers of bedrooms and bathrooms available from property data found in the assessor's office of the county government. Using address information documented in medical records or administrative datasets linked to the county assessor's data, we were able to calculate an effective SES proxy without the need for specific educational or income levels which are rarely available in the medical record or administrative data. The HOUSES index predicts health outcomes in both adults and children that have previously been identified to be associated with SES (low birth weight, obesity, smoking exposure, asthma control status, pneumococcal diseases, post-Myocardial infarction mortality, and rheumatoid arthritis development and mortality risk).[6-12]

One of the challenges in calculating the original HOUSES index is the need for complex assessor's real property data generated for US taxation purposes. However, this data often does not include key variables of interest such as the number of bedrooms and bathrooms. For example, the 2013 real property data of Olmsted County, Minnesota, have 3 - 6% missing data on the number of bedrooms and bathrooms of single family housing units while assessed housing value and square footage is almost complete (<1% missing data). The rate of missing information on the number of bedrooms and bathrooms tend to vary depending on the age of the real property data and/or geographic regions of interest.

To address this concern, we explored whether the original HOUSES index could be modified using fewer housing-related characteristics, especially assessed housing value and square footage as these two components are consistently available in most counties and state property databases. Our aim was to evaluate whether a modified HOUSES index would provide an equivalent representation of SES status to the original HOUSES index.

METHODS

Study subjects and design: The original study enrolled parents/guardians of children aged 1 – 17 years living in Olmsted County, MN (n=746) or Jackson County, MO (n = 704) in 2006. The present study included those who had both successful geocoding of address with real property data and formulation of HOUSES index (728 subjects for Olmsted County, MN, and 701 subjects for Jackson County, MO). Detailed description of the study population and methodology for developing and validating the HOUSES index were previously reported by Juhn et al.[7] Briefly, subjects were originally recruited for the HOUSES derivation and validation study.[7] Data collection included sociodemographic characteristics and health-related information obtained through a telephone survey given to one parent or guardian who answered the phone. This information was then linked to the property data associated with each subject's address. Property data were acquired from the county assessor's offices. For comparison, additional SES

measures were included in the survey questionnaire, to describe parental education level (i.e., the highest educational level of either parent), family annual income, Hollingshead index (a family's composite index using education, occupation, sex, and marital status), and Nakao-Treas index (composite index using educational attainment and income of job incumbents corresponding to the 1980 census).[13-15]

For formulating HOUSES index, principal component factor analysis was performed using 7 housing-related features obtained in the real property data, including (1) square footage of housing unit, (2) assess housing value, (3) number of bathrooms, (4) number of bedrooms, (5) ownership of housing unit, (6) residential status for whether a housing unit is in a residential zoning, and (7) lot size of housing unit in acre, and 6 neighborhood characteristics collected from census-tract level data, including (1) percent of people speaking English as a second language, (2) percent of foreign-born people, (3) percent of households headed by female, (4) percent of households that are non-family households, (5) percent of people with less than high-school education, and (6) percent of families with family income below poverty level. The original HOUSES index was calculated using four housing-related characteristics (assessed housing value, square footage, number of bedrooms, and number of bathrooms) included in the first factor accounting for the largest proportion of total variance. These four housing components were transformed to standardized z-scores, and then summed up to formulate the HOUSES index. In the original study, it was demonstrated that a higher 4-item HOUSES score was related to a higher level of SES using other SES measures and also inversely associated with outcome measures assessed among subjects from both counties. While the HOUSES index developed in Olmsted County, MN, and was validated in Jackson County, MO in the original study, this index was further validated to a different tax jurisdiction and real property data system in Sioux Falls, SD.[11] Correlation of HOUSES with the health-related outcomes of childhood obesity, low birth weight, and smoking were evaluated because the association between these outcomes and SES has been well demonstrated. [16-18] The institutional review boards at both the University of Northern Iowa (for conducting telephone interviews) and Mayo Clinic approved the consent and study procedures.

To assess whether fewer housing-related characteristics could be used, the relative influence (RI) of each characteristic for predicting the health outcomes (obesity, low birth weight, and smoking exposure) was estimated using gradient boosting machine (GBM) models under logistic regression model frameworks. The GBM modeling approach is a machine learning technique for building a multivariable prediction model by incorporating all of the variables without variable selection.[3 19 20] RI is a measure of a given variable's importance, relative to that of other variables, in the model prediction process. The measure is based on the number of times a variable is selected for splitting in a decision tree, weighted by the improvement of the model fitting as a result of the split and further standardized so that the sum of RI from all variables adds up to 100%. The higher the RI value (maximum of 100%) of a characteristic, the more significant its contribution is to the model. Those with the greatest contribution were summed to formulate a modified HOUSES index for each county.

Statistical analysis*:* Descriptive statistics were used to represent socio-demographic characteristics and health-related outcomes for subjects for each county. Pearson correlation coefficients were calculated for correlations between original HOUSES and the modified HOUSES indices. For further analysis, both original HOUSES and modified HOUSES scores were collapsed into 4 groups using quartiles (Q1 [lowest] – Q4 [highest]). Spearman correlation coefficients were calculated to see whether correlations with other SES measures (parental education, family annual income, Hollingshead index, and Nakao-Treas index) were similar between original HOUSES. To evaluate whether the two non-independent correlation coefficients (original and modified HOUSES indices calculated on the same subjects) were similar, a t-test based on Hotelling's test accounting for dependency between two HOUSES indices was utilized.[21] In addition, logistic regression models were used to assess the association of the modified HOUSES with risk for health-related outcomes (obesity [body mass index at or above the 95th percentile for children of the same age and gender; yes vs no], low birth weight [<2,500g at birth; yes vs no], and smoking exposure [tobacco smoking status of household member; yes vs no]), using Q1 as a reference category. We used the 2002 National Health Interview Survey (NHIS) questions to obtain these

dependent variables ("What was child's birth weight?" "How much does child weight now?" "How tall is child now?" and "Does anyone in the household use cigarettes, cigars, or pipe tobacco?").

RESULTS

Characteristics of subjects: The results are summarized in Table 1. A total of 728 children from Olmsted County, MN, and 701 children from Jackson, MO, were included in the study analysis. The median age of children included in the study was 10 years (25^{th} – 75^{th} percentile: 5 - 14) with roughly 50% females in both counties (Table 1). Residents of Olmsted County, MN were more likely to have higher levels of education and income than residents of Jackson County, MO. Obesity rates (15% vs 12%) and smoking exposure (27% vs 12%) were higher in Jackson County, while the rate of low birth weight was higher in Olmsted County (11% vs 6.5%). The median HOUSES index was -0.44 (25^{th} – 75^{th} percentile: -1.11 to 0.91) in Olmsted County, MN, and -0.46 (-1.22 to -0.88) in Jackson County, MO.

Table 1: Sociodemographic characteristics and health outcomes of the study subjects

	Olmsted County, MN (n = 728)	Jackson County, MO (n = 701)
Demographic characteristics		
Age (years), median (25^{th} – 75^{th} %tiles)	10 (5 - 14)	10 (5 - 14)
Sex, female N (%)	358 (49%)	355 (51%)
Socioeconomic characteristics		
Parents' education		
Less than high school education	4 (0.5%)	18 (2.6%)
High school graduate	37 (5.1%)	103 (15%)
Some college, no degree	140 (19%)	173 (25%)
Associate/college degree	291 (40%)	229 (33%)
Graduate or professional degree	256 (35%)	178 (25%)
Family annual income		
Less than $24,999	9 (1.3%)	51 (7.8%)
$25,000 to $49,999	86 (12%)	139 (21%)
$50,000 to $74,999	144 (20%)	154 (24%)
$75,000 to $99,999	161 (22%)	136 (21%)
Over $100,000	316 (44%)	175 (27%)
Hollingshead index		
8 to 19	2 (0.3%)	2 (0.3%)
20 to 29	12 (1.6%)	35 (5.0%)
30 to 39	54 (7.4%)	109 (16%)

40 to 54	254 (35%)	268 (38%)
55 to 66	406 (56%)	287 (41%)
Nakao-Treas index		
0 to 12.5	0 (0.0%)	0 (0.0%)
12.6 to 25.1	2 (0.3%)	5 (0.7%)
25.2 to 37.7	53 (7.3%)	101 (14%)
37.8 to 50.3	79 (11%)	105 (15%)
50.4 to 62.9	109 (15%)	105 (15%)
63.0 to 75.5	184 (25%)	192 (27%)
75.6 to 88.1	230 (32%)	156 (22%)
88.2 to 100	71 (9.8%)	37 (5.3%)
Health outcomes		
Obesity, N (%)	71 (12%)	81 (15%)
Low birth weight, N (%)	78 (11%)	43 (6.5%)
Smoking exposure, N (%)	89 (12%)	188 (27%)

Identification of housing characteristics with most significant contributions for health outcomes: The results are summarized in Figure 1A (for Olmsted County) and 1B (for Jackson County). Simultaneously considering all four housing characteristics used in the original HOUSES formula, assessed housing value and square footage had the greatest contribution to all three health outcomes in both counties (Figure 1A and 1B). The combined contribution of assessed housing value and square footage ranged between 89% and 96% depending on health outcome evaluated. The contribution of number of bedrooms and bathrooms is negligible in the presence of assessed housing value and square footage (< 10% for all three outcomes). Therefore, a modified HOUSES index was formulated by summing z-scores for assessed housing value and square footage once normalized for the remaining analysis.

Correlation of modified HOUSES index with the original HOUSES index and other SES measures, and health outcomes: The results are summarized in Tables 2 and 3 and Figure 2. The modified HOUSES index was highly correlated with original HOUSES based on 4 housing components (0.87 in Olmsted County, and 0.93 in Jackson County). Correlations between the modified HOUSES index with other SES measures were comparable or slightly higher to the original HOUSES index for both counties (Table 2).

Overall, the modified index performs similar to the original HOUSES index in inverse associations with health outcomes (Table 3 and Figure 2). Figure 2 depicted odds ratios (ORs) and their 95% CIs for association between the modified HOUSES (using Q1 as a reference group) and each of three health outcomes (Panel A for Olmsted County and Panel B for Jackson County). In addition, Figure 2 included the association results with the original HOUSES for comparison. The 95% CIs for the modified HOUSES were overlapped with those for the original HOUSES for all three outcomes and both counties, which implied that the association results between two HOUSES measures were similar. The strongest health-related outcome association with the modified HOUSES Index was observed for smoking exposure (OR: 0.24 with 95% CI: 0.11 – 0.49 in Olmsted County and OR: 0.26 with 95% CI 0.16 – 0.44 in Jackson County for comparing subjects in highest [Q4] vs lowest [Q1] group; overall $p < 0.001$). The risk of childhood obesity was also inversely associated with the modified HOUSES, although not statistically significant in Jackson County (overall p-value for the modified HOUSES = 0.01 in Olmsted County; 0.11 in Jackson County). The association for the risk of low birthweight was inconsistent between two counties (Table 3). As mentioned in the original paper describing the HOUSES index, the insignificant association in Olmsted County might be potentially due to 1) a low incidence of low birth weight in Olmsted County, 2) a low incidence of low birth weight among the low-income population in Olmsted County such as Somali immigrants in the county, and 3) a high incidence of multiple gestations (associated with low birth weight) from in-vitro fertilization among a relatively higher SES group.

Table 2: The correlation between other SES measures and original and modified HOUSES: both original and modified HOUSES are comparable in correlation with other SES measures.

	Olmsted County, MN			Jackson County, MO		
	Original HOUSES	Modified HOUSES	P-values*	Original HOUSES	Modified HOUSES	P-values*
Parents' education	0.23	0.26	0.16	0.39	0.42	0.07
Family annual income	0.43	0.49	<0.001	0.52	0.55	0.06
Hollingshead index	0.19	0.22	0.17	0.36	0.38	0.24

Nakao-Treas index	0.24	0.26	0.39	0.39	0.40	0.56

* P-values comparing a correlation coefficient of each other SES measure with original HOUSES and that with modified HOUSES.

Table 3: Association (unadjusted odds ratio [OR] and 95% CI) between modified HOUSES and risk of childhood obesity, low birthweight, and smoking exposure.

	Obesity, OR [95% CI]	Low birth weight, OR [95% CI]	Smoking exposure, OR [95% CI]
Olmsted County	[overall p= 0.01]	[overall p= 0.82]	[overall p < 0.001]
Q1 (lowest SES)	Reference	Reference	Reference
Q2	0.62 [0.33 – 1.18]	1.22 [0.51 – 2.91]	0.39 [0.21 – 0.73] **
Q3	0.29 [0.14 – 0.63] **	1.02 [0.41 – 2.51]	0.71 [0.41 – 1.22]
Q4 (highest SES)	0.48 [0.25 – 0.94] *	0.78 [0.30 – 2.03]	0.24 [0.11 – 0.49] **
Jackson County	[overall p= 0.11]	[overall p= 0.08]	[overall p < 0.001]
Q1 (lowest SES)	Reference	Reference	Reference
Q2	0.58 [0.30 – 1.13]	0.42 [0.18 – 1.00] *	0.60 [0.38 – 0.94] *
Q3	0.75 [0.40 – 1.41]	0.38 [0.16 – 0.94] *	0.47 [0.30 – 0.75] **
Q4 (highest SES)	0.45 [0.23 – 0.89] *	0.54 [0.34 – 1.21]	0.26 [0.16 – 0.44] **

*: p-values between 0.01 and 0.05
**: p-values less than 0.01

DISCUSSION

In this study, we examined the utility of the "modified" HOUSES index as a suitable tool for health disparities research. The modified HOUSES relies on only two housing-related variables (assessed housing value and square footage), data usually publicly available. We made four main observations, providing reliability, validity, predictability and generalizability of the modified HOUSES index as an alternative SES measure. First, the modified HOUSES was strongly correlated with the original four component HOUSES (assessed housing value, square footage, number of bedrooms and bathrooms) index in both counties evaluated and the degree of the correlations was similar in both counties (i.e., reliability across different geographic settings). Second, the degree of correlations between the modified HOUSES index with other SES measures (e.g., parental education) were comparable to the original HOUSES index for both counties (i.e., validity). Third, this "modified" HOUSES index was consistently

associated with health outcomes linked to SES, such as smoking exposure (i.e., predictability). Finally, the modified HOUSES index utilized 2 housing features well captured in real property data throughout diverse communities within the United States, and provided consistent results in the two geographically different regions (i.e., generalizability).

The observed relatively minimal influence of not including bedroom and bathroom count in the modified HOUSES has several potential explanations. First, the variance of assessed housing values and square footage in a continuous variable is larger than the number of bedrooms and bathrooms, a discrete variable with a small range. Therefore, it is possible that the impact of bedroom and bathroom counts may be minimal once assessed housing value and square footage are considered. Second, as missing information for bedroom count and bathroom count were imputed with the mean value in the original HOUSES index, their impact (variance) might have been reduced, compared to assessed housing value and square footage with minimal missing values. Finally, while positive correlation exists among number of bedrooms, square footage, number of bathrooms, and assessed housing value, we found that assessed housing value and square footage are better correlated with other SES measures than the number of bedrooms and bathrooms. For example, assessed housing value (r=0.52) and square footage (r=0.48) are more closely correlated with income, compared to the self-reported number of bedrooms (r=0.18) and bathrooms (r=0.37). In support of this finding, in the original study, regardless of whether we used the imputed data or self-reported data on the number of bedrooms and bathrooms, these two different estimates did not significantly influence the number of factors and factor loadings for original HOUSES index. Thus, the number of bedrooms and bathrooms conceptually seems to only add a finer granularity in capturing housing-based SES. Therefore, based on these conceptual and methodological aspects of the housing features, the modified HOUSES index performed as well as the original HOUSES index.

The paucity of SES-related data is a common but major challenge for existing large-scale datasets (e.g., disease registry, administrative datasets, etc.) that would otherwise be valuable for conducting health disparities research. Therefore, advantages of housing-based SES (both original and modified

HOUSES) are promising as the address information, which is almost routinely collected in health care settings (e.g., medical records), is directly linked or geocoded to real property data to formulate the HOUSES index. However, considering that there are high missing rates of number of bedrooms and bathrooms in assessor's real property data or even unavailability of these two components in some regions, our study findings provide an important basis for using the modified HOUSES index as a potential alternative to the original HOUSES index in studying and addressing health disparities. Overall, the modified HOUSES index provides an alternative approach for measuring SES when conventional data for characterizing SES is not available.

Strengths of the current study include that it was conducted in two study settings with diverse socioeconomic characteristics observed within each study population (external validity). The robust performance of modified HOUSES in these communities demonstrates that this approach to characterizing SES is both feasible and generalizable. Another strength for calculating the modified HOUSES index is to minimize effort for imputing missing information in the dataset.

Limitations include self-report-based health outcomes and are therefore subject to reporting bias. However, the associations between self-reported health outcomes and SES have been well demonstrated in multiple independent investigations. Furthermore, these same health outcomes, defined by physician diagnosis or predetermined criteria, were also significantly associated with our original HOUSES index. [6-12] The performance of modified HOUSES for objective measure-based health outcomes can be expected to be similar to these findings. HOUSES, which is developed based on real property data for US taxation purposes, may not work well in other countries where housing data is not routinely collected or made publicly available in databases, or even in communities within the US where housing assessment is not done frequently or the quality of assessment is poor. Furthermore, the formulation of modified HOUSES was done relying on the relationship among four housing characteristics observed in the two Midwestern counties. Therefore, it is possible that the modified HOUSES may not work as well as the original HOUSES in communities where the relationship among those 4 characteristics is drastically different

compared to the two Midwestern counties used in this study. Additionally, the use of assessed values without other more objective measures for housing features may make the modified HOUSES index more susceptible to a potential bias when comparisons are made among communities in which widely different assessment procedures are used. Further research might focus on assessing and reducing any potential biases. The modified HOUSES index is likely to be more robust when used in a single community for determining SES of individuals and families.

In conclusion, a modified HOUSES calculation, using two housing-related characteristics (assessed housing value and square footage) instead of four, highly correlates with the ability of the original HOUSES index to represent SES status. The two modified HOUSES components of assessed housing value and square footage are well captured in assessor's housing data. As a result, a modified HOUSES improves the feasibility of comprehensively assessing SES status and expanding the application of this tool into different geographic regions without complete real property data needed for HOUSES, especially in US Midwestern communities.

Acknowledgement: The authors would like to thank Ms. Kelly Okeson and Ms. Elizabeth Krusemark for their editorial peer review for the manuscript. Collection and validation of the Jackson County data was facilitated by Joan Pu, City of Kansas City, Missouri, and Mark Funkhouser, PhD, then Mayor of Kansas City, Missouri.

Funding: This work was supported by the National Heart Lung and Blood Institute (R01 HL126667), the National Institute of Allergy and Infectious Diseases (R21 AI101277, R21 AI116839), and T. Denny Sanford Foundation.

Competing interests: None declared

Contributors: ER, CW, and YJJ were responsible for the study design, initial manuscript drafting, and interpretation of the results. The original study design for and formulation of HOUSES was done by

ARW and YJJ. CW, PHW, JAS, BPY, TJB, ARW and YJJ were responsible for procedures used to collect the data for the original HOUSES. SSC and PHW contributed critically for manuscript drafting. ER and SMA conducted the statistical analysis in this paper. Earlier statistical analyses for the basis of HOUSES were done by ARW. All the authors had approved the final version of the manuscript.

Data sharing statement: No additional data are available.

Figure legend:

Figure 1: Relative influence (percent) of 4 housing features (assessed housing value, square footage, number of bedrooms, and number of bathrooms) for risk of obesity, low birthweight, and smoking exposure among subjects from Olmsted County (Panel A) and Jackson County (Panel B).

Figure 2: Association comparisons between modified HOUSES (odds ratio and 95% Cis with dotted line) and original HOUSES (odds ratio and 95% Cis with solid line) for three health-related outcomes (obesity, low birthweight, and smoking exposure) among subjects from Olmsted County (Panel A) and Jackson County (Panel B).

REFERENCES

1. Martyn M, Weaver AL, Jacobson RM, Juhn YJ. Characterization of the duration from onset of asthma symptoms to asthma disease. *Ann Allergy Asthma Immunol* 2008;100(6):589-95.
2. Dhillon RK, Yawn BP, Yoo KH, Boyce TG, Jacobson RM, McGree ME, et al. Impact of Asthma on the Severity of Serious Pneumococcal Disease. *Epidemiology* 2011;Suppl 3.
3. Juhn YJ, Kita H, Yawn BP, Boyce TG, Yoo KH, McGree ME, et al. Increased risk of serious pneumococcal disease in patients with asthma. *J Allergy Clin Immunol* 2008;122(4):719-23.
4. Bjur KA, Lynch RL, Fenta YA, Yoo KH, Jacobson RM, Li X, et al. Assessment of the association between atopic conditions and tympanostomy tube placement in children. *Allergy Asthma Proc* 2012;33(3):289-96.
5. Capili CR, Hettinger A, Rigelman-Hedberg N, Fink L, Boyce T, Lahr B, et al. Increased risk of pertussis in patients with asthma. *J Allergy Clin Immunol* 2012;129(4):957-63.
6. Butterfield MC, Williams AR, Beebe T, Finnie D, Liu H, Liesinger J, et al. A two-county comparison of the HOUSES index on predicting self-rated health. *J Epidemiol Community Health* 2011;65(3):254-9.
7. Juhn YJ, Beebe TJ, Finnie DM, Sloan J, Wheeler PH, Yawn B, et al. Development and initial testing of a new socioeconomic status measure based on housing data. *J Urban Health* 2011;88(5):933-44.
8. Johnson MD, Urm SH, Jung JA, Yun HD, Munitz GE, Tsigrelis C, et al. Housing data-based socioeconomic index and risk of invasive pneumococcal disease: an exploratory study. *Epidemiol Infect* 2013;141(4):880-7.
9. Bang DW, Manemann SM, Gerber Y, Roger VL, Lohse CM, Rand-Weaver J, et al. A Novel Socioeconomic Measure Using Individual Housing Data in Cardiovascular Outcome Research. *Int J Env Res Pub He* 2014;11(11):11597-615.
10. Juhn Y, Krusemark E, Rand-Weaver J, Beebe T, Sloan J, Yawn B, et al. A novel measure of socioeconomic status using individual housing data in health disparities research for asthma in adults. *Allergy* 2014;69:327-27.
11. Harris MN, Lundien MC, Finnie DM, Williams AR, Beebe TJ, Sloan JA, et al. Application of a novel socioeconomic measure using individual housing data in asthma research: an exploratory study. *Npj Prim Care Resp M* 2014;24.
12. Ghawi H, Crowson CS, Rand-Weaver J, Krusemark E, Gabriel SE, Juhn YJ. A novel measure of socioeconomic status using individual housing data to assess the association of SES with rheumatoid arthritis and its mortality: a population-based case-control study. *BMJ Open* 2015;5(4):e006469.
13. Hollingshead A. Four Factor Index of Social Status: New Haven, CT: Yale University Department of Psychology, 1975.
14. Rieppi R, Greenhill LL, Ford RE, Chuang S, Wu M, Davies M, et al. Socioeconomic status as a moderator of ADHD treatment outcomes. *J Am Acad Child Adolesc Psychiatry* 2002;41(3):269-77.
15. Nakao K. The 1989 Socioeconomic index of occupations: Construction from the 1989 Occupational Prestige Scores (General Social Survey Methodological Report No 74): Chicago, IL: University of Chicago, National Opinion Research Center, 1992.
16. Wang Y. Cross-national comparison of childhood obesity: the epidemic and the relationship between obesity and socioeconomic status. *Int J Epidemiol* 2001;30(5):1129-36.
17. Parker JD, Schoendorf KC, Kiely JL. Associations between measures of socioeconomic status and low birth weight, small for gestational age, and premature delivery in the United States. *Ann Epidemiol* 1994;4(4):271-8.

18. Whitlock G, MacMahon S, Vander Hoorn S, Davis P, Jackson R, Norton R. Association of environmental tobacco smoke exposure with socioeconomic status in a population of 7725 New Zealanders. *Tob Control* 1998;7(3):276-80.
19. Kang CI, Rouse MS, Patel R, Kita H, Juhn YJ. Allergic airway inflammation and susceptibility to pneumococcal pneumonia in a murine model with real-time in vivo evaluation. *Clin Exp Immunol* 2009;156(3):552-61.
20. Molis WE, Bagniewski S, Weaver AL, Jacobson RM, Juhn YJ. Timeliness of diagnosis of asthma in children and its predictors. *Allergy* 2008;63(11):1529-35.
21. Williams EJ. The Comparison of Regression Variables. *Journal of the Royal Statistical Societies* 1959;21 (2):396-99.

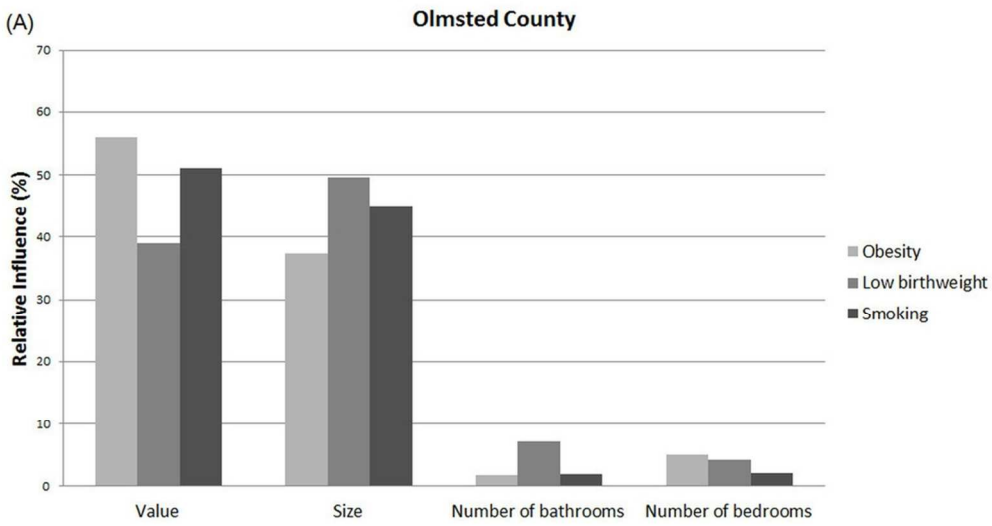

Figure 1A
89x47mm (300 x 300 DPI)

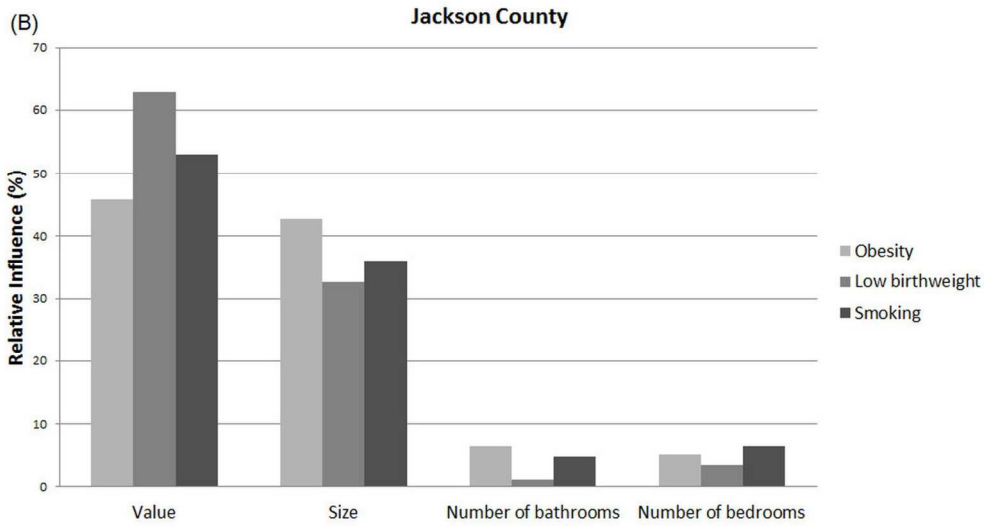

Figure 1B
166x90mm (300 x 300 DPI)

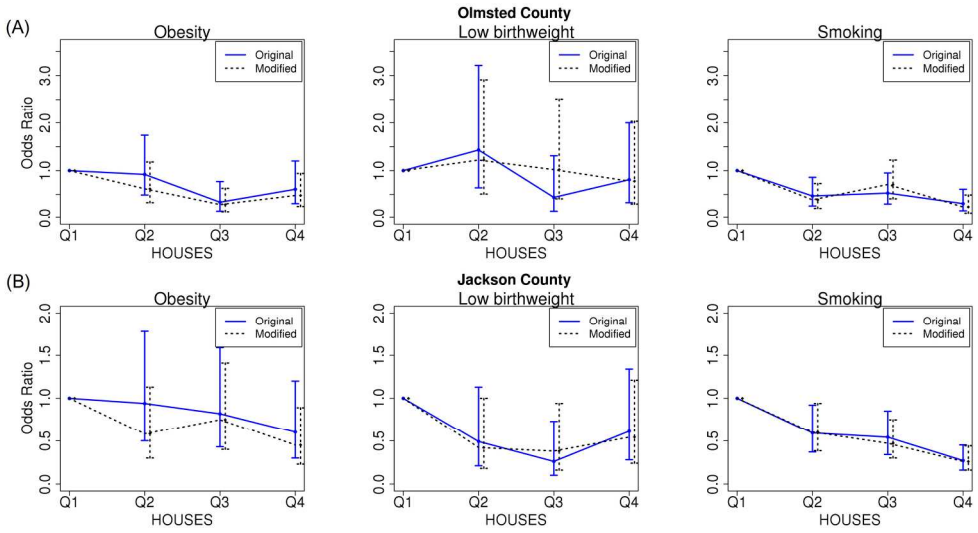

Figure 2
203x108mm (300 x 300 DPI)

BMJ Open

Assessing health disparities in children using a modified housing-related socioeconomic status measure: A cross-sectional study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-011564.R2
Article Type:	Research
Date Submitted by the Author:	23-May-2016
Complete List of Authors:	Ryu, Euijung; Mayo Clinic, Health Sciences Research Wi, Chung-Il ; Mayo Clinic, Pediatric and Adolescent Medicine Crow, Sheri; Mayo Clinic, Pediatric and Adolescent Medicine Armasu, Sebastian; Mayo Clinic, Health Sciences Research Wheeler, Philip; Mayo Clinic, Pediatric and Adolescent Medicine Sloan, Jeff; MAYO clinic Yawn, Barbara; Olmsted Medical Center, Department of Research Beebe, Timothy ; Mayo Clinic Williams, Arthur; James A. Haley Veterans' Hospital Juhn, Young; Mayo Clinic, Community Pediatrics and Adolescent Medicine
Primary Subject Heading:	Paediatrics
Secondary Subject Heading:	Public health
Keywords:	socioeconomic status, HOUSES, health disparity

Assessing health disparities in children using a modified housing-related socioeconomic status measure: A cross-sectional study

Euijung Ryu [1,*], Chung-Il Wi [2,*], Sheri S. Crow [1,2], Sebastian M. Armasu [1], Philip H. Wheeler [2], Jeff A. Sloan [1], Barbara P. Yawn [3], Timothy J. Beebe [1], Arthur R. Williams[4,5], Young J. Juhn [1,2]

[1] Department of Health Sciences Research, Mayo Clinic, Rochester, MN;[2] Department of Pediatric and Adolescent Medicine, Mayo Clinic, Rochester, MN; [3] Olmsted Medical Center, Rochester, MN; [4]Center of Innovation on Disability and Rehabilitation Research (CINDRR), James A. Haley Veterans Affairs Hospital, Tampa, FL; [5] Department of Health Administration and Policy, George Mason University, Fairfax, VA

* These authors contributed equally

Correspondence to: Young J. Juhn, MD, MPH
Professor of Pediatrics
Department of Pediatric and Adolescent Medicine, Mayo Clinic
200 First Street SW,
Rochester MN, 55905
E-mail: Juhn.Young@mayo.edu
Phone: 507-538-1642
Fax: 507-284-9744

Keywords: Socioeconomic status, HOUSES, health disparity

Word counts: 2742

ABSTRACT

Objectives: Socioeconomic status (SES) is a well-established risk factor for many health outcomes. Recently, we developed an SES measure based on four housing-related characteristics (termed HOUSES) and demonstrated its ability to assess health disparities. In this study, we aimed to evaluate whether fewer housing-related characteristics could be utilized to provide a similar representation of SES.

Study setting and participants: We performed a cross-sectional study using parents/guardians of children aged 1-17 years from two US Midwestern counties (n=728 in Olmsted County, Minnesota, and n=701 in Jackson County, Missouri).

Primary and secondary outcome measures: For each subject, housing-related characteristics used in the formulation of HOUSES (assessed housing value, square footage, number of bedrooms, and number of bathrooms) were obtained from the local government assessor's offices, and additional SES measures and health outcomes with known associations to SES (obesity, low birthweight, and smoking exposure) were collected from a telephone survey. Housing characteristics with the greatest contribution for predicting the health outcomes were added to formulate a modified HOUSES index.

Results: Among the four housing characteristics used in the original HOUSES, the strongest contributions for predicting health outcomes were observed from assessed housing value and square footage (combined contribution ranged between 89% and 96%). Based on this observation, these two were used to calculate a modified HOUSES index. Correlation between modified HOUSES and other SES measures was comparable to the original HOUSES for both locations. Consistent with the original HOUSES formula, the strongest association with modified HOUSES was observed with smoking exposure (OR=0.24 with 95% CI 0.11–0.49 for comparing subjects in highest HOUSES vs lowest group; overall p<0.001).

Conclusions: The modified HOUSES requires only two readily available housing characteristics thereby improving the feasibility of using this index as a proxy for SES in multiple communities, especially in the US Midwestern region.

Strengths and limitations of this study:

- Robust performance of the modified HOUSES demonstrates its application to different geographic regions with minimal property data.
- The modified HOUSES index does not rely on the quality of imputation of missing housing characteristics.
- This study is limited by self-reported health outcomes.
- The study is limited by testing the modified index in only two locations.
- This study may not work well in other countries where housing data is not routinely collected or not made publically available.

INTRODUCTION

The impact of socioeconomic status (SES) on health outcomes has been well documented in the US and elsewhere and included in assessments of frameworks for health disparities.[1-5] Overall, these frameworks for health disparities suggest that distal factors such as individual SES impact human health independently, jointly, and interactively with proximal factors (e.g., genetic predisposition or biological responses). Thus, SES, its definition, and method of calculation can have important consequences on clinical practice, research, and health policy concerning health disparities.

SES reflects multifaceted assets or capacities of humans including materialistic, human, and social capital, making accurate measurement of SES a potentially formidable task. One of the biggest challenges in health disparities research is the complexity of reporting individual's SES using commonly available data such as information found in medical records and administrative databases. To address this important gap in the ability to operationalize SES using data available to health care researchers, our research group developed an individual housing-based SES measure termed HOUSES.[6] HOUSES is a composite index consisting of assessed housing value, square footage, and the numbers of bedrooms and bathrooms available from property data found in the assessor's office of the county government. Using address information documented in medical records or administrative datasets linked to the county assessor's data, we were able to calculate an effective SES proxy without the need for specific educational or income levels, which are rarely available in the medical record or administrative data. The HOUSES index predicts health outcomes in both adults and children that have previously been identified to be associated with SES (low birth weight, obesity, smoking exposure, asthma control status, pneumococcal diseases, post-Myocardial infarction mortality, rheumatoid arthritis (RA), and post-RA mortality).[6-12]

One of the challenges in calculating the original HOUSES index is the need for complex assessor's real property data generated for US taxation purposes. However, this data often does not

include key variables of interest such as the number of bedrooms and bathrooms. For example, the 2013 real property data of Olmsted County, Minnesota, have 3 - 6% missing data on the number of bedrooms and bathrooms of single family housing units while assessed housing value and square footage is almost complete (<1% missing data). The rates of missing information on the number of bedrooms and bathrooms tend to vary depending on the age of the real property data and/or geographic regions of interest.

To address this concern, we explored whether the original HOUSES index could be modified using fewer housing-related characteristics, especially assessed housing value and square footage as these two components are consistently available in most counties and state property databases. Our aim was to evaluate whether a modified HOUSES index would provide an equivalent representation of SES status to the original HOUSES index.

METHODS

Study subjects and design: The original study enrolled parents/guardians of children aged 1 – 17 years living in Olmsted County, MN (n=746) or Jackson County, MO (n=704) in 2006. The present study included those who had both successful geocoding of address with real property data and formulation of HOUSES index (728 subjects for Olmsted County, MN, and 701 subjects for Jackson County, MO). Detailed description of the study population and methodology for developing and validating the HOUSES index were previously reported by Juhn et al.[6] Briefly, subjects were originally recruited for the HOUSES derivation and validation study.[6] Data collection included sociodemographic characteristics and health-related information obtained through a telephone survey given to the one parent or guardian who answered the phone. This information was then linked to the property data associated with each subject's address. Property data were acquired from the county assessor's offices. For comparison, additional SES measures were included in the survey questionnaire, to describe parental education level (i.e., the highest educational level of either parent), family annual income, Hollingshead index (a family's

composite index using education, occupation, sex, and marital status), and Nakao-Treas index (composite index using educational attainment and income of job incumbents corresponding to the 1980 census).[13-15]

For formulating the HOUSES index, principal component factor analysis was performed using 7 housing-related features obtained in the real property data, including (1) square footage of housing unit, (2) assessed housing value, (3) number of bathrooms, (4) number of bedrooms, (5) ownership of housing unit, (6) residential status (whether a housing unit is in a residential zone), and (7) lot size of housing unit in acres, and 6 neighborhood characteristics collected from census-tract level data, including (1) percent of people speaking English as a second language, (2) percent of foreign-born people, (3) percent of households headed by a female, (4) percent of households that are non-family households, (5) percent of people with less than a high-school education, and (6) percent of families with family income below poverty level. The original HOUSES index was calculated using four housing-related characteristics (assessed housing value, square footage, number of bedrooms, and number of bathrooms) included in the first factor, accounting for the largest proportion of total variance. These four housing components were transformed to standardized z-scores, and then summed to formulate the HOUSES index. In the original study, it was demonstrated that a higher 4-item HOUSES score was related to a higher level of SES using other SES measures and also inversely associated with outcome measures assessed among subjects from both counties. While the HOUSES index developed in Olmsted County, MN, and was validated in Jackson County, MO in the original study, this index was further validated to a different tax jurisdiction and real property data system in Sioux Falls, SD.[11] Correlation of HOUSES with the health-related outcomes of childhood obesity, low birth weight, and smoking were evaluated because the association between these outcomes and SES has been well demonstrated. [16-18] The institutional review boards at both the University of Northern Iowa (for conducting telephone interviews) and Mayo Clinic approved the consent and study procedures.

To assess whether fewer housing-related characteristics could be used, the relative influence (RI) of each characteristic for predicting the health outcomes (obesity, low birth weight, and smoking exposure) was

estimated using gradient boosting machine (GBM) models under logistic regression model frameworks. The GBM modeling approach is a machine learning technique for building a multivariable prediction model by incorporating all of the variables without variable selection.[3 19 20] RI is a measure of a given variable's importance, relative to that of other variables, in the model prediction process. The measure is based on the number of times a variable is selected for splitting in a decision tree, weighted by the improvement of the model fitting as a result of the split and further standardized so that the sum of RI from all variables adds up to 100%. The higher the RI value (maximum of 100%) of a characteristic, the more significant its contribution is to the model. Those with the greatest contribution were summed to formulate a modified HOUSES index for each county.

Statistical analysis*:* Descriptive statistics were used to represent socio-demographic characteristics and health-related outcomes for subjects for each county. Pearson correlation coefficients were calculated for correlations between original HOUSES and the modified HOUSES indices. For further analysis, both original HOUSES and modified HOUSES scores were collapsed into 4 groups using quartiles (Q1 [lowest] – Q4 [highest]). For each HOUSES index, Spearman correlation coefficients were calculated for correlations with other SES measures (parental education, family annual income, Hollingshead index, and Nakao-Treas index). To evaluate whether the two non-independent correlation coefficients (original and modified HOUSES indices calculated on the same subjects) were similar, a t-test based on Hotelling's test accounting for dependency between two HOUSES indices was utilized.[21] In addition, logistic regression models were used to assess the association of the modified HOUSES with risk for health-related outcomes (obesity [body mass index at or above the 95th percentile for children of the same age and gender; yes vs no], low birth weight [<2,500g at birth; yes vs no], and smoking exposure [tobacco smoking status of household member; yes vs no]), using Q1 as a reference category. We used the 2002 National Health Interview Survey (NHIS) questions to obtain these dependent variables ("What was child's birth weight?" "How much does child weight now?" "How tall is child now?" and "Does anyone in the household use cigarettes, cigars, or pipe tobacco?").

RESULTS

Characteristics of subjects: The results are summarized in Table 1. A total of 728 children from Olmsted County, MN, and 701 children from Jackson, MO, were included in the study analysis. The median age of children included in the study was 10 years (25^{th} – 75^{th} percentile: 5 - 14) with roughly 50% females in both counties (Table 1). Residents of Olmsted County, MN were more likely to have higher levels of education and income than residents of Jackson County, MO. Obesity rates (15% vs 12%) and smoking exposure (27% vs 12%) were higher in Jackson County, while the rate of low birth weight was higher in Olmsted County (11% vs 6.5%). The median HOUSES index was -0.44 (25^{th} – 75^{th} percentile: -1.11 to 0.91) in Olmsted County, MN, and -0.46 (-1.22 to 0.88) in Jackson County, MO.

Table 1: Sociodemographic characteristics and health outcomes of the study subjects

	Olmsted County, MN (n = 728)	Jackson County, MO (n = 701)
Demographic characteristics		
Age (years), median (25^{th} – 75^{th} %tiles)	10 (5 - 14)	10 (5 - 14)
Sex, female N (%)	358 (49%)	355 (51%)
Socioeconomic characteristics		
Parents' education		
Less than high school education	4 (0.5%)	18 (2.6%)
High school graduate	37 (5.1%)	103 (15%)
Some college, no degree	140 (19%)	173 (25%)
Associate/college degree	291 (40%)	229 (33%)
Graduate or professional degree	256 (35%)	178 (25%)
Family annual income		
Less than $24,999	9 (1.3%)	51 (7.8%)
$25,000 to $49,999	86 (12%)	139 (21%)
$50,000 to $74,999	144 (20%)	154 (24%)
$75,000 to $99,999	161 (22%)	136 (21%)
Over $100,000	316 (44%)	175 (27%)
Hollingshead index		
8 to 19	2 (0.3%)	2 (0.3%)
20 to 29	12 (1.6%)	35 (5.0%)
30 to 39	54 (7.4%)	109 (16%)
40 to 54	254 (35%)	268 (38%)
55 to 66	406 (56%)	287 (41%)
Nakao-Treas index		
0 to 12.5	0 (0.0%)	0 (0.0%)

12.6 to 25.1	2 (0.3%)	5 (0.7%)
25.2 to 37.7	53 (7.3%)	101 (14%)
37.8 to 50.3	79 (11%)	105 (15%)
50.4 to 62.9	109 (15%)	105 (15%)
63.0 to 75.5	184 (25%)	192 (27%)
75.6 to 88.1	230 (32%)	156 (22%)
88.2 to 100	71 (9.8%)	37 (5.3%)
Health outcomes		
Obesity, N (%)	71 (12%)	81 (15%)
Low birth weight, N (%)	78 (11%)	43 (6.5%)
Smoking exposure, N (%)	89 (12%)	188 (27%)

Identification of housing characteristics with most significant contributions for health outcomes: The results are summarized in Figure 1A (for Olmsted County) and 1B (for Jackson County). Simultaneously considering all four housing characteristics used in the original HOUSES formula, assessed housing value and square footage had the greatest contribution to all three health outcomes in both counties (Figure 1A and 1B). The combined contribution of assessed housing value and square footage ranged between 89% and 96% depending on health outcome evaluated. The contribution of number of bedrooms and bathrooms is negligible in the presence of assessed housing value and square footage (< 10% for all three outcomes). Therefore, assessed housing value and square footage were transformed to standard z-scores and then summed to formulate a modified HOUSES index.

The modified HOUSES index for presenting SES status: The results are summarized in Tables 2 and 3 and Figure 2. The modified HOUSES index was highly correlated with original HOUSES based on 4 housing components (0.87 in Olmsted County, and 0.93 in Jackson County). Correlations between the modified HOUSES index with other SES measures were comparable or slightly higher to the original HOUSES index for both counties (Table 2).

Overall, the modified index performs similar to the original HOUSES index in inverse associations with health outcomes (Table 3 and Figure 2). Figure 2 depicted odds ratios (ORs) and their 95% CIs for association between the modified HOUSES (using Q1 as a reference group) and each of three health outcomes (Panel A for Olmsted County and Panel B for Jackson County). In addition, Figure

2 includes the association results with the original HOUSES for comparison. The 95% CIs for the modified HOUSES were overlapped with those for the original HOUSES for all three outcomes and both counties, which implied that the association results between two HOUSES measures were similar. The strongest health-related outcome association with the modified HOUSES Index was observed for smoking exposure (OR: 0.24 with 95% CI: 0.11 – 0.49 in Olmsted County and OR: 0.26 with 95% CI 0.16 – 0.44 in Jackson County, comparing subjects in highest [Q4] vs lowest [Q1] group; overall p < 0.001for both locations). The risk of childhood obesity was also inversely associated with the modified HOUSES, although not statistically significant in Jackson County (overall p=0.01 in Olmsted County; p=0.11 in Jackson County). The association for the risk of low birthweight was inconsistent between two counties (Table 3). We postulate that the lack of significant association in Olmsted County might be partly due to unique characteristics of the Olmsted County population, such as a relatively high prevalence of recent Somali immigrants with low SES. As mentioned in the original paper describing the HOUSES index, the incidence of low birth weight in Somali population was lower than the US average.[22] In addition, a high incidence of multiple gestations (associated with low birth weight) from in-vitro fertilization participants among a relatively higher SES group in Olmsted County might also contribute to the results.

Table 2: Comparison of the correlation coefficients of two HOUSES indices with other SES measures

	Olmsted County, MN			Jackson County, MO		
	Original HOUSES	Modified HOUSES	P-values*	Original HOUSES	Modified HOUSES	P-values*
Parents' education	0.23	0.26	0.16	0.39	0.42	0.07
Family annual income	0.43	0.49	<0.001	0.52	0.55	0.06
Hollingshead index	0.19	0.22	0.17	0.36	0.38	0.24
Nakao-Treas index	0.24	0.26	0.39	0.39	0.40	0.56

* Each p-value represents statistical significance for the difference between the two Spearman correlation coefficients (r1 and r2): (r1: the correlation coefficient between the original HOUSES index and each individual SES measure) and (r2: the correlation coefficient between the modified HOUSES index and each individual SES measure). Lack of statistical significant difference (p>0.05) means no difference between r1 and r2.

Table 3: Associations (unadjusted odds ratio [OR] and 95% CI) between modified HOUSES and risk of childhood obesity, low birthweight, and smoking exposure.

	Obesity, OR [95% CI]	Low birth weight, OR [95% CI]	Smoking exposure, OR [95% CI]
Olmsted County	[overall p= 0.01]	[overall p= 0.82]	[overall p < 0.001]
Q1 (lowest SES)	Reference	Reference	Reference
Q2	0.62 [0.33 – 1.18]	1.22 [0.51 – 2.91]	0.39 [0.21 – 0.73] **
Q3	0.29 [0.14 – 0.63] **	1.02 [0.41 – 2.51]	0.71 [0.41 – 1.22]
Q4 (highest SES)	0.48 [0.25 – 0.94] *	0.78 [0.30 – 2.03]	0.24 [0.11 – 0.49] **
Jackson County	[overall p= 0.11]	[overall p= 0.08]	[overall p < 0.001]
Q1 (lowest SES)	Reference	Reference	Reference
Q2	0.58 [0.30 – 1.13]	0.42 [0.18 – 1.00] *	0.60 [0.38 – 0.94] *
Q3	0.75 [0.40 – 1.41]	0.38 [0.16 – 0.94] *	0.47 [0.30 – 0.75] **
Q4 (highest SES)	0.45 [0.23 – 0.89] *	0.54 [0.34 – 1.21]	0.26 [0.16 – 0.44] **

*: p<0.05
**: p<0.01

DISCUSSION

In this study, we examined the utility of the "modified" HOUSES index as a suitable tool for health disparities research. The modified HOUSES relies on only two housing-related variables (assessed housing value and square footage), data usually publicly available. We made four main observations, providing reliability, validity, predictability and generalizability of the modified HOUSES index as an alternative SES measure. First, the modified HOUSES was strongly correlated with the original four component HOUSES (assessed housing value, square footage, number of bedrooms and bathrooms) index in both counties evaluated and the correlation coefficients between two HOUSES indices was similar in both counties (i.e., reliability across different geographic settings). Second, the correlation coefficients between the modified HOUSES index with other SES measures (e.g., parental education) were comparable to the original HOUSES index for both counties (i.e., validity). Third, the associations of this modified HOUSES index with health outcomes or related known risk factors linked to SES, such as smoking exposure, were consistent with those of the original HOUSES (i.e., predictability). Finally, the modified HOUSES index utilized 2 housing features commonly captured in real property data

throughout diverse communities within the United States, and provided consistent results in the two geographically different regions (i.e., generalizability).

The observed relatively-minimal influence of not including bedroom and bathroom count in the modified HOUSES has several potential explanations. First, the variances of assessed housing values and square footage, continuous variables, are larger than those for number of bedrooms and bathrooms, discrete variables with smaller ranges. Therefore, it is possible that the impact of bedroom and bathroom counts may be minimal once assessed housing value and square footage are considered. Second, as missing information for bedroom count and bathroom count were imputed with the mean value in the original HOUSES index, their impact (variance) might have been reduced, compared to assessed housing value and square footage, for which there were few missing values. Finally, while positive correlations exists among number of bedrooms, square footage, number of bathrooms, and assessed housing value, we found that assessed housing value and square footage better correlate with other SES measures than the number of bedrooms and bathrooms. For example, assessed housing value (r=0.52) and square footage (r=0.48) are more closely correlated with income, compared to the self-reported number of bedrooms (r=0.18) and bathrooms (r=0.37). Thus, the number of bedrooms and bathrooms conceptually seems to only add a finer granularity in capturing housing-based SES. Therefore, based on these conceptual and methodological aspects of the housing features, the modified HOUSES index performed as well as the original HOUSES index.

The paucity of readily accessible SES-related data is a common but major challenge for existing large-scale datasets (e.g., disease registry, administrative datasets.) rendering them less valuable for conducting health disparities research. Therefore, use of housing-based SES (i.e., both original and modified HOUSES) is promising as address information, is almost routinely collected in health care settings (e.g., medical records), and is directly linked or geocoded to real property data.. Considering that there are high missing rates of number of bedrooms and bathrooms in assessor's real property data, our study findings provide important evidence supporting use of the modified HOUSES index as a potential

alternative to the original HOUSES index in studying and addressing health disparities. Overall, the modified HOUSES index provides an alternative approach for measuring SES when conventional data for characterizing SES is not available.

One strength of the current study is that it was conducted in two study settings with diverse socioeconomic characteristics (external validity). The robust performance of modified HOUSES in these communities demonstrates that this approach to characterizing SES is both feasible and generalizable. Although it is not a strength of the current study, utilizing the modified HOUSES index minimizes effort for imputing missing information in the real property dataset.

A limitation of the study is that self-reported health outcomes are subject to reporting bias. The associations between self-reported health outcomes and SES, however, have been well demonstrated in multiple independent investigations. Furthermore, these same health outcomes, defined by physician diagnosis or predetermined criteria, were also significantly associated with our original HOUSES index. [6-12] The performance of modified HOUSES for objective measure-based health outcomes can be expected to be similar to these findings. HOUSES, which is developed based on real property data for US taxation purposes, may not work well in other countries where housing data is not routinely collected or made publicly available in databases, or even in communities within the US where housing assessments are infrequent or of poor quality. Furthermore, the formulation of modified HOUSES was done relying on the relationship among four housing characteristics observed in the two Midwestern counties. Therefore, it is possible that the modified HOUSES may not work as well as the original HOUSES in communities where the relationship among those 4 characteristics is drastically different compared to the two counties used in this study. Additionally, the use of assessed values without other more objective measures for housing features may make the modified HOUSES index more susceptible to a potential bias when comparisons are made among communities in which widely different assessment procedures are used. Further research might focus on assessing and reducing any potential biases. The modified HOUSES

index is likely to be more robust when used in a single community for determining SES of individuals and families.

In conclusion, a modified HOUSES calculation, using two housing-related characteristics (assessed housing value and square footage) instead of four, highly correlates with the ability of the original HOUSES index to represent SES status, especially in US Midwestern communities. The two modified HOUSES components are commonly-captured in assessor's housing data. As a result, the modified HOUSES improves the feasibility of comprehensively assessing SES status, expanding the application of this tool into different geographic regions that do not routinely collect the real property data needed for the original HOUSES index.

Acknowledgement: The authors would like to thank Ms. Kelly Okeson and Ms. Elizabeth Krusemark for their editorial peer review for the manuscript. Collection and validation of the Jackson County data was facilitated by Joan Pu, City of Kansas City, Missouri, and Mark Funkhouser, PhD, then Mayor of Kansas City, Missouri.

Funding: This work was supported by the National Heart Lung and Blood Institute (R01 HL126667), the National Institute of Allergy and Infectious Diseases (R21 AI101277, R21 AI116839), and T. Denny Sanford Foundation.

Competing interests: None declared

Contributors: ER, CW, and YJJ were responsible for the study design, initial manuscript drafting, and interpretation of the results. The original study design for and formulation of HOUSES was done by ARW and YJJ. CW, PHW, JAS, BPY, TJB, ARW and YJJ were responsible for procedures used to collect the data for the original HOUSES. SSC and PHW contributed critically for manuscript drafting. ER and SMA conducted the statistical analysis in this paper. Earlier statistical analyses for the basis of HOUSES were done by ARW. All the authors had approved the final version of the manuscript.

Data sharing statement: No additional data are available.

Figure legend:

Figure 1: Relative influence (percent) of 4 housing features (assessed housing value, square footage, number of bedrooms, and number of bathrooms) for risk of obesity, low birthweight, and smoking exposure among subjects from Olmsted County (Panel A) and Jackson County (Panel B).

Figure 2: Association comparisons between modified HOUSES (odds ratio and 95% confidence intervals with dotted line) and original HOUSES (odds ratio and 95% confidence intervals with solid line) for three health-related outcomes (obesity, low birthweight, and smoking exposure) among subjects from Olmsted County (Panel A) and Jackson County (Panel B).

REFERENCES

1. Martyn M, Weaver AL, Jacobson RM, Juhn YJ. Characterization of the duration from onset of asthma symptoms to asthma disease. *Ann Allergy Asthma Immunol* 2008;100(6):589-95.
2. Dhillon RK, Yawn BP, Yoo KH, Boyce TG, Jacobson RM, McGree ME, et al. Impact of Asthma on the Severity of Serious Pneumococcal Disease. *Epidemiology* 2011;Suppl 3.
3. Juhn YJ, Kita H, Yawn BP, Boyce TG, Yoo KH, McGree ME, et al. Increased risk of serious pneumococcal disease in patients with asthma. *J Allergy Clin Immunol* 2008;122(4):719-23.
4. Bjur KA, Lynch RL, Fenta YA, Yoo KH, Jacobson RM, Li X, et al. Assessment of the association between atopic conditions and tympanostomy tube placement in children. *Allergy Asthma Proc* 2012;33(3):289-96.
5. Capili CR, Hettinger A, Rigelman-Hedberg N, Fink L, Boyce T, Lahr B, et al. Increased risk of pertussis in patients with asthma. *J Allergy Clin Immunol* 2012;129(4):957-63.
6. Juhn YJ, Beebe TJ, Finnie DM, Sloan J, Wheeler PH, Yawn B, et al. Development and initial testing of a new socioeconomic status measure based on housing data. *J Urban Health* 2011;88(5):933-44.
7. Butterfield MC, Williams AR, Beebe T, Finnie D, Liu H, Liesinger J, et al. A two-county comparison of the HOUSES index on predicting self-rated health. *J Epidemiol Community Health* 2011;65(3):254-9.
8. Johnson MD, Urm SH, Jung JA, Yun HD, Munitz GE, Tsigrelis C, et al. Housing data-based socioeconomic index and risk of invasive pneumococcal disease: an exploratory study. *Epidemiol Infect* 2013;141(4):880-7.
9. Bang DW, Manemann SM, Gerber Y, Roger VL, Lohse CM, Rand-Weaver J, et al. A Novel Socioeconomic Measure Using Individual Housing Data in Cardiovascular Outcome Research. *Int J Env Res Pub He* 2014;11(11):11597-615.
10. Juhn Y, Krusemark E, Rand-Weaver J, Beebe T, Sloan J, Yawn B, et al. A novel measure of socioeconomic status using individual housing data in health disparities research for asthma in adults. *Allergy* 2014;69:327-27.
11. Harris MN, Lundien MC, Finnie DM, Williams AR, Beebe TJ, Sloan JA, et al. Application of a novel socioeconomic measure using individual housing data in asthma research: an exploratory study. *Npj Prim Care Resp M* 2014;24.
12. Ghawi H, Crowson CS, Rand-Weaver J, Krusemark E, Gabriel SE, Juhn YJ. A novel measure of socioeconomic status using individual housing data to assess the association of SES with rheumatoid arthritis and its mortality: a population-based case-control study. *BMJ Open* 2015;5(4):e006469.
13. Hollingshead A. Four Factor Index of Social Status: New Haven, CT: Yale University Department of Psychology, 1975.
14. Rieppi R, Greenhill LL, Ford RE, Chuang S, Wu M, Davies M, et al. Socioeconomic status as a moderator of ADHD treatment outcomes. *J Am Acad Child Adolesc Psychiatry* 2002;41(3):269-77.
15. Nakao K. The 1989 Socioeconomic index of occupations: Construction from the 1989 Occupational Prestige Scores (General Social Survey Methodological Report No 74): Chicago, IL: University of Chicago, National Opinion Research Center, 1992.
16. Wang Y. Cross-national comparison of childhood obesity: the epidemic and the relationship between obesity and socioeconomic status. *Int J Epidemiol* 2001;30(5):1129-36.
17. Parker JD, Schoendorf KC, Kiely JL. Associations between measures of socioeconomic status and low birth weight, small for gestational age, and premature delivery in the United States. *Ann Epidemiol* 1994;4(4):271-8.

18. Whitlock G, MacMahon S, Vander Hoorn S, Davis P, Jackson R, Norton R. Association of environmental tobacco smoke exposure with socioeconomic status in a population of 7725 New Zealanders. *Tob Control* 1998;7(3):276-80.
19. Kang CI, Rouse MS, Patel R, Kita H, Juhn YJ. Allergic airway inflammation and susceptibility to pneumococcal pneumonia in a murine model with real-time in vivo evaluation. *Clin Exp Immunol* 2009;156(3):552-61.
20. Molis WE, Bagniewski S, Weaver AL, Jacobson RM, Juhn YJ. Timeliness of diagnosis of asthma in children and its predictors. *Allergy* 2008;63(11):1529-35.
21. Williams EJ. The Comparison of Regression Variables. *Journal of the Royal Statistical Societies* 1959;21 (2):396-99.
22. Flynn PM, Foster EM, Brost BC. Indicators of acculturation related to Somali refugee women's birth outcomes in Minnesota. *J Immigr Minor Health* 2011;13(2):224-31.

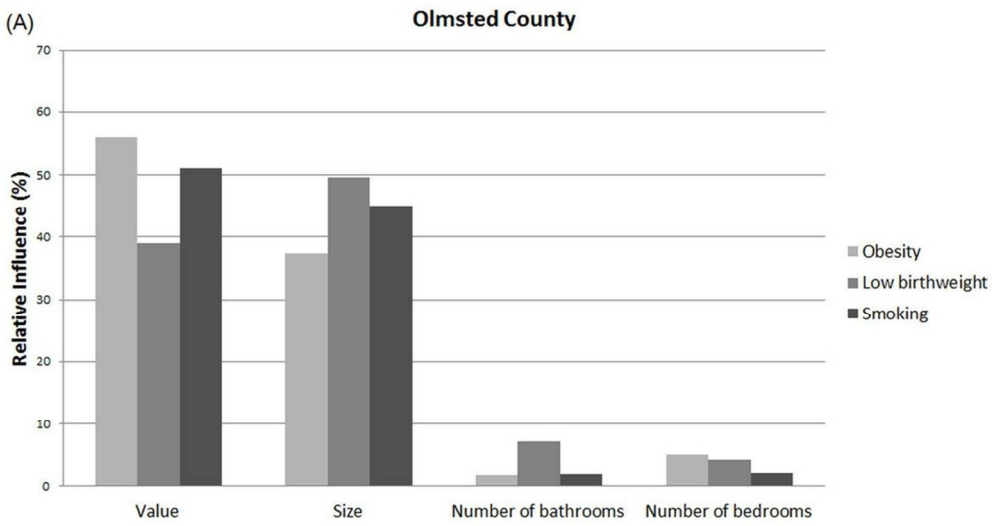

Figure 1A
89x47mm (300 x 300 DPI)

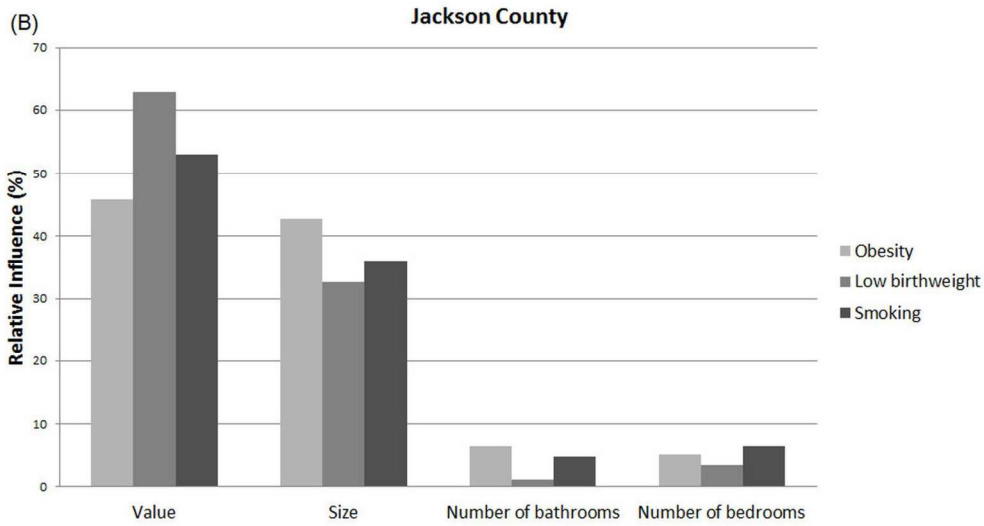

Figure 1B
166x90mm (300 x 300 DPI)

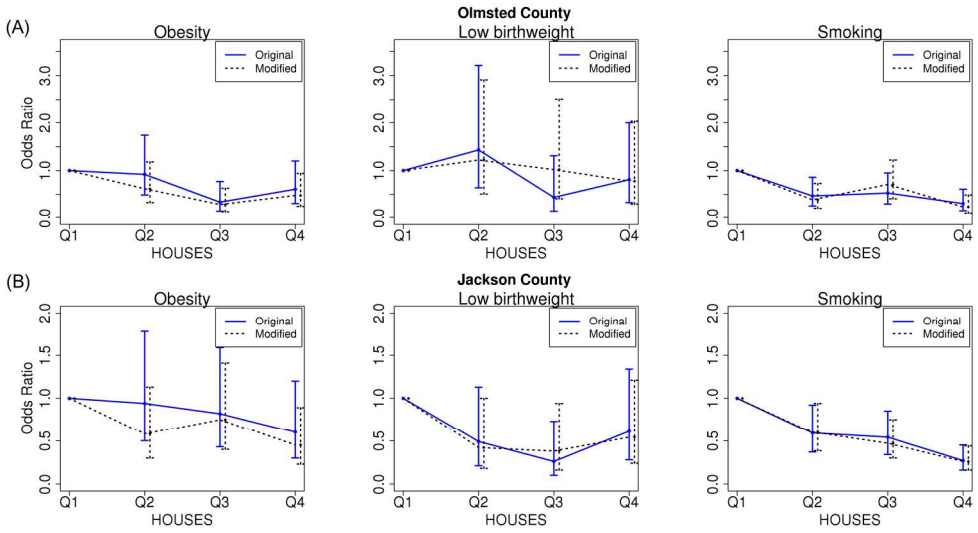

Figure 2
203x108mm (300 x 300 DPI)