Original Article

# Constructing Practical and Realistic Asset-Based Socioeconomic Status Assessment Scale Using Principal Component Analysis for Urban Population of Puducherry, India

**Vinayagamoorthy Venugopal, Amol R. Dongre[1], Poomathy Ponnusamy[2]**

Department of Community Medicine, Sri Manakula Vinayagar Medical College and Hospital, Puducherry, [1]Department of Extension Programme, Pramukhswami Medical College (PSMC), Karamsad, Gujarat, [2]Assistant Programme Manager, Deputy Director of Health Service, Erode District, Tamil Nadu, India

## Abstract

**Background:** Socioeconomic status (SES) is a key determinant of health. However, ascertaining the SES in developing countries is really challenging. Hence, we decided to develop an asset-based simple and rational SES tool for urban population of Puducherry and compare it with Modified Kuppuswamy's (MK) scale. **Materials and Methods:** Sequential mixed methods design was used. The list of local household assets to determine SES was created based on group interviews with stakeholders and review of literature. Then, survey was carried out among 500 urban households by trained medical interns after obtaining informed consent. EpiCollect-5, mobile-based software, was used to capture data. Principal component analysis (PCA) was carried out to construct a wealth index using SPSS version 24. The assets included in the final PCA were ranked based on their contribution to the index by linear regression. **Results:** The eigenvalue for the first principal component was 6.7 accounting for 33.6% of the variance in the original data. Finally, reduced 10-item-based SES scale was created and scoring system was formulated based on regression coefficient. The weighted kappa statistics and correlation coefficient measure of reliability between household quintiles on 20-item and 10-item reduced SES tool were 0.77 and 0.95, respectively. There was a moderate correlation between SES obtained from MK scale and newly constructed scale. **Conclusions:** The newly devised SES scale is context specific, reliable, easy to administer, and quick to ascertain the SES and thus can be used for a similar context in future health research.

**Keywords:** India, principal component analysis, socioeconomic status, urban population

## Introduction

Socioeconomic status (SES) is a comprehensive term relating to the social and economic factors that determine the position of individuals or groups within a hierarchical society.[1,2] The SES influences the accessibility, affordability, acceptability, and utility of available health facilities and hence an important determinant of health status of people.[3] SES is measured to ascertain the household development and to decide for the provision of social security and welfare services and to avail various SES-based government incentives. Consumption expenditure, education, income, occupation, participatory wealth ranking, and subjective measures were commonly used in epidemiological investigations to ascertain the SES, however, each has its own limitations.[4-8]

Nationwide representative surveys conducted all over the world use wealth index which is a household asset-based index as a measure to classify SES of people using principal component analysis (PCA). This has a higher predictive value than other methods of measuring SES.[7,9] However, data collection is time consuming and analysis and interpretation requires expertise. Notably, some of the assets used in the nationwide survey might not be context specific and outdated. Therefore, we decided to construct a household wealth quintile to ascertain the SES of urban population of Puducherry using

Access this article online

Quick Response Code:

Website: www.ijcm.org.in

DOI: 10.4103/ijcm.IJCM_925_20

**Address for correspondence:** Dr. Vinayagamoorthy Venugopal, Department of Community Medicine, Sri Manakula Vinayagar Medical College and Hospital, Puducherry - 605 001, India.
E-mail: drvinayagamoorthy@gmail.com



**How to cite this article:** Venugopal V, Dongre AR, Ponnusamy P. Constructing practical and realistic asset-based socioeconomic status assessment scale using principal component analysis for urban population of Puducherry, India. Indian J Community Med 2021;46:494-8.

**Received:** 24-10-20, **Accepted:** 13-04-21, **Published:** 13-10-21

the assets that are available locally and specific to urban context by PCA method. We also aimed to construct an SES tool with reduced assets and to check its correlation with the wealth quintile derived from the comprehensive list of assets and to compare the finding with commonly used modified Kuppuswamy's (MK) scale.

## Materials and Methods

### Study setting

The study was planned and carried out by the epidemiology unit of the department of community medicine in a tertiary care teaching hospital in Puducherry. The data collection was done in Villianur, a town that comprised approximately of 2000 households, which is the headquarters of the Villianur taluk in the Union Territory of Puducherry.

### Study design

It is a sequential exploratory mixed method design, having two phases which are sequential in nature.[10] In the first phase, the qualitative research method was carried out and it was followed by the quantitative phase [Figure 1]. This study was a part of a larger study on health insurance and its determinants among middle-class family members and the Institute Ethical Committee clearance (Code no.: 51/2017) was obtained along with the main study.

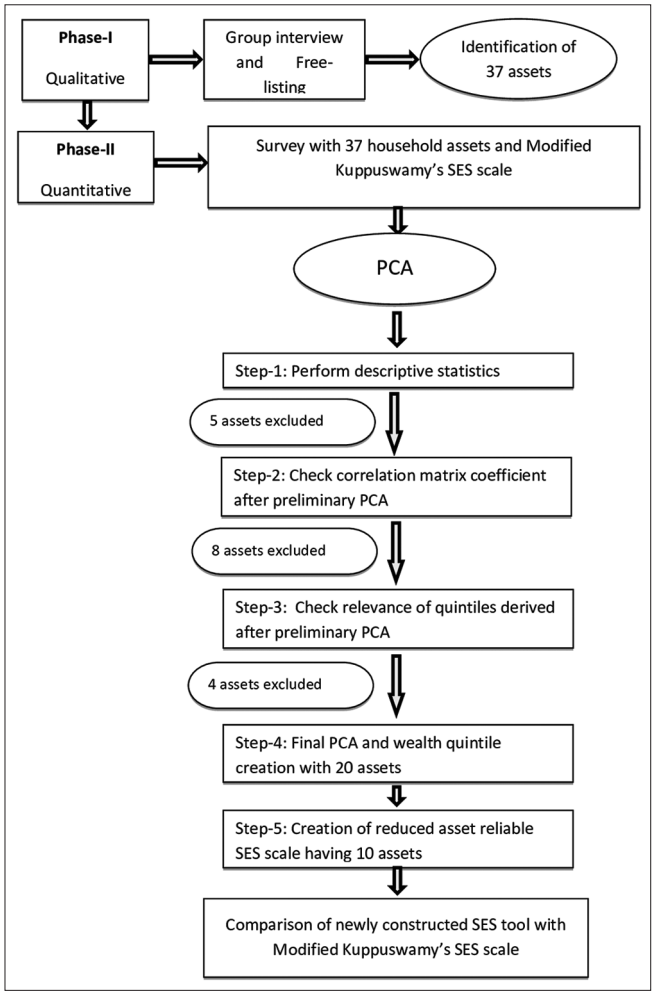


**Figure 1:** Flow diagram explaining the mixed-methods design and the steps followed in principal component analysis and validation of the study

### Sample size and sampling

As a rule of thumb, a minimum of 10 participants per variable is necessary to avoid the statistical difficulties in PCA analysis. We included a sample size of 500 which was considered an adequate sample that could yield meaningful interpretation statistically.[11] Systematic random sampling was adopted to identify the households. The sampling interval was calculated to be four (total household/desired sample size).

### Data collection

#### *Phase-I (qualitative data collection and analysis)*

Group interviews and free listing of assets were carried out among 10 study participants who were the head of households, mothers of under-five children, Anganwadi teachers, field workers, and staff nurses available in the study area. They were selected purposively on the basis of their perceived experience with the subject of interest (ability to discriminate SES based on the availability of household assets), willingness to participate, and the ability to communicate their opinions in an expressive and reflective manner.[12] The participants were informed to list the household assets that can discriminate the SES of the family. The assets used in the recent National Family Health Survey (NFHS-4) was used to assist participants to arrive at a comprehensive list.[13] The listed items were manually summarized and the commonly reported assets were included in the questionnaire. The same was then face validated with the field experts and was then pilot tested and finalized. Eventually, a 37-item questionnaire was developed to be used in the next phase of the study. The items are mentioned in Table 1.

#### *Phase-II (quantitative study)*

The questionnaire containing 37 assets identified in phase-I and MK scale was used to conduct a community-based survey. In the randomly selected household after obtaining written informed consent, the homemakers (women) were interviewed as they are better aware of the assets available. The survey was carried out by postgraduate and medical interns in the department of community medicine. Information was captured using an efficient mobile application, EpiCollect version 5.0.

### Statistical analysis

Descriptive analysis of assets was done in frequency and percentage. PCA technique adopted by the World Bank was used to develop asset-based socioeconomic index or wealth index.[14] Correlation between the assets is used to generate a set of uncorrelated principal components.[15] Since the aim is to construct a single measure of SES, the first principal component was used to define the asset index.[16]

Households were divided into quintiles based on the composite wealth index score. Linear regression was then carried out to find the contribution of each asset in determining the wealth index. Again the PCA was carried

out to find out the various composite wealth indices and wealth quintiles using the top 12, 10, 8, and 6 assets that were ranked based on regression coefficient. Pearson's correlation and weighted kappa agreement were carried out for statistical comparison of the quintiles generated using a different number of assets. Finally, a reliable and simple SES tool with reduced assets that can be quickly used in the field was prepared. The scoring system was developed based on the regression coefficient obtained from linear regression. The SES obtained based on a new reduced asset-based scale was compared with SES of the MK scale. Data analysis was performed in software SPSS version 24 (SPSS Inc, Chicago, IL, USA).

**Table 1: Description of final 37 assets included in the survey ($n$=500)**

| Assets | *n* (%) |
|---|---|
| **Water usage** | |
| Disinfected drinking water | 323 (64.6) |
| Wall-mounted water heater | 207 (41.4) |
| Motor to pump groundwater | 303 (60.6) |
| **Cooking related** | |
| LPG as cooking fuel | 497 (99.4) |
| Induction stove | 239 (47.8) |
| Rice cooker | 217 (43.4) |
| Microwave oven | 58 (11.6) |
| Wet grinder (self-owned) | 402 (80.4) |
| Wet grinder (government supply) | 259 (51.8) |
| Mixer grinder (self-owned) | 403 (80.6) |
| Mixer grinder (government supply) | 265 (53) |
| **Electronic goods** | |
| Refrigerator | 473 (94.6) |
| Washing machine | 326 (65.2) |
| Dish wash machine | 8 (1.6) |
| Air conditioner | 259 (51.8) |
| Inverter | 150 (30) |
| Iron box | 437 (87.4) |
| Ceiling fan | 494 (98.8) |
| **Recreational items** | |
| Black-and-white TV | 1 (0.2) |
| Color TV (old model) | 217 (43.4) |
| Color TV (LCD/LED) | 241 (48.2) |
| Cable connection (set top box) | 49 (9.8) |
| **Telecommunication** | |
| Landline | 133 (26.6) |
| Mobile phone (basic model) | 257 (51.4) |
| Mobile phone (smart model) | 383 (76.6) |
| Desktop/laptop (self-owned) | 207 (41.4) |
| Laptop (government supply) | 15 (3) |
| Internet connection | 213 (42.6) |
| **Housing** | |
| Own living house | 330 (66) |
| Pucca type of living house | 463 (92.6) |
| Availability of own land | 141 (28.2) |
| **Transportation** | |
| Bicycle | 177 (35.4) |
| Scooter | 230 (46) |
| Bike | 371 (74.2) |
| Car | 141 (28.2) |
| **Others** | |
| Sewing machine | 121 (24.2) |
| Livestock in home | 29 (5.8) |

LPG: Liquefied petroleum gas, LCD: Liquid crystal display, LED: Light emitting diode

## Selection of assets to run principal component analysis

The process of the final selection of assets to include in the PCA analysis is depicted in Figure 1. Assets that are correlated and unequally distributed between households will only be able to discriminate the SES better using PCA analysis.[16] Hence, five assets that were present in more than 95% or <5% of the households were excluded, following which preliminary PCA with 32 items was done. The correlation matrix of the PCA analysis which showed eight variables had either high ($r > 0.80$) or poor correlation ($r < 0.10$) and as a rule, these assets can also be not able to differentiate SES, so they were also removed.[15] Again, PCA with 24-items was carried out. Then, the wealth quintile generated was cross-tabulated with the assets included in the PCA. Four more variables that had meaningless distribution with wealth quintiles were also finally excluded [Figure 2]. After multiple PCAs and reiterations, 20 assets were ultimately identified to be able to discriminate the SES better.

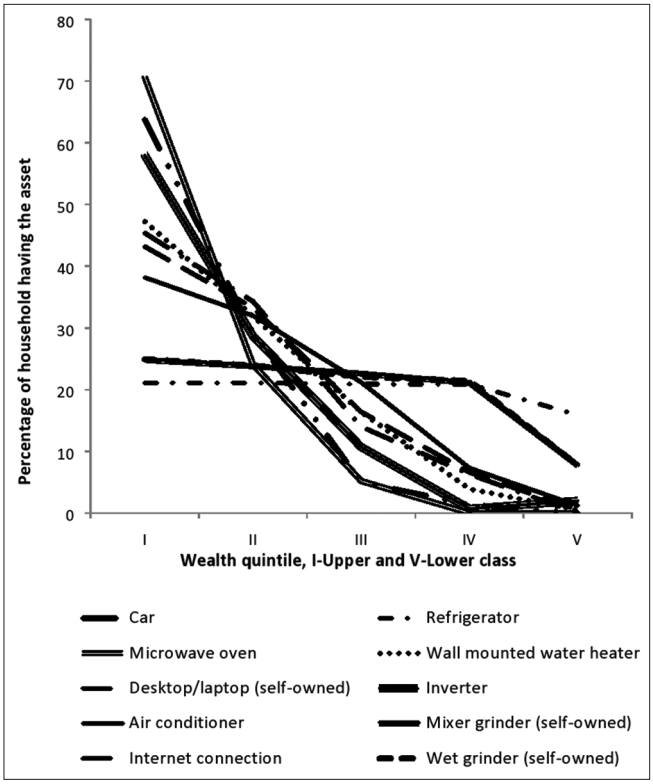


**Figure 2:** Distribution of top ten household assets across the wealth quintiles of the study participants

## Results

Liquid petroleum gas (LPG) as cooking fuel was present in 99.4% of the households. Similarly, a ceiling fan for ventilation was seen in 98.8% of the households. On the contrary, dish wash machine, black-and-white television, and government-supplied laptop were found in 1.6%, 0.5%, and 3% of the households [Table 1].

PCA with the finally selected 20 assets showed the eigenvalue for the first principal component as 6.7 accounting for 33.6% of the variance in the original data. The Kaiser–Meyer–Olkin measure of sampling adequacy was 0.856 ($P < 0.001$). Wealth quintile was prepared ultimately based on the PCA analysis with 20 assets. The linear regression analysis with the composite wealth index score as dependent variable showed car ranked first and presence of motor to pump groundwater in the household ranked the last in the list based on regression co-efficient value. The list of assets with corresponding weights is shown in Table 2.

The Spearman correlation coefficient for the 20-item and 10-item wealth quintile was 0.95 and was statistically significant ($P<0.001$) and the value of weighted kappa agreement between them was 0.77 (95% confidence interval: 0.74–0.80). Their agreement was better than the agreement between 8 and 6 items with 20 item-based SES quintile and the same as that of 12-asset and 20-asset-based SES quintiles. Finally, a simple and reliable 10-asset-based SES scale was developed and the scoring system was generated based on the regression coefficient of the assets, as displayed in Table 2. The scoring system of the simplified 10-item household asset-based socioeconomic scale is shown in Table 3.

**Table 2: List of assets ordered based on their priority in entrance to the linear regression model**

| Rank | Asset | Regression coefficient |
|---|---|---|
| 1 | Car | 0.238 |
| 2 | Refrigerator | 0.234 |
| 3 | Microwave oven | 0.223 |
| 4 | Wall-mounted water heater | 0.221 |
| 5 | Desktop/laptop (self-owned) | 0.219 |
| 6 | Inverter | 0.218 |
| 7 | Air conditioner | 0.216 |
| 8 | Mixer grinder (self-owned) | 0.207 |
| 9 | Internet connection | 0.206 |
| 10 | Wet grinder (self-owned) | 0.205 |
| 11 | Color TV (LCD/LED) | 0.204 |
| 12 | Washing machine | 0.195 |
| 13 | Mobile phone (smart model) | 0.186 |
| 14 | Color TV (old model) | 0.175 |
| 15 | Induction stove | 0.174 |
| 16 | Landline | 0.172 |
| 17 | Disinfected drinking water | 0.137 |
| 18 | Mobile phone (basic model) | 0.136 |
| 19 | Availability of own land | 0.132 |
| 20 | Motor to pump groundwater | 0.089 |

LCD: Liquid crystal display, LED: Light emitting diode

The SES was calculated for all participants using the MK scale. The SES was calculated with the newly constructed 10-item asset-based SES scale (based on PCA analysis). Wealth quintile calculated by PCA with 20-asste scale and with newly constructed 10-asset scale was correlated and the co-efficient obtained was 0.90 ($P < 0.001$). Correlation co-efficient between the SES ascertained using the MK scale and the newly constructed 10-asset scale was 0.52 ($P < 0.001$).

## Discussion

Household asset-based simple SES scale with a scoring system for urban population of Puducherry was constructed. There was no under representation of context-specific household assets used to classify SES. The agreement between the wealth quintile generated using PCA based on 20 assets and reduced 10 assets was excellent. However, the correlation between SES obtained between MK scale and the newly constructed 10-asset-based scale was moderate.

In the derivation of asset-based wealth index, different assets were used worldwide by various authors.[14] The assets included in the NFHS-4 survey were used as a reference; however, many modifications were made to it. Assets specific for rural areas such as animal-drawn cart, tractor, and thresher were removed. Commonly found assets such as cot, chair, mattress, and table that are unlikely to differentiate SES were excluded. Furthermore, to keep it context specific, mixer grinder, wet grinder, and laptop that were supplied by the local government free of cost in the study area were included. We also included many new items as per the result of qualitative interview with homemakers. Hence, the tool is unique to the urban context, so its usage suits better for any urban geographic area.

**Table 3: The scoring system of the simplified 10-item household asset-based socioeconomic scale**

| Asset | Score |
|---|---|
| Car | 4 |
| Refrigerator | 3 |
| Microwave oven | 2 |
| Wall-mounted water heater | 2 |
| Desktop/laptop (self-owned) | 2 |
| Inverter | 2 |
| Air conditioner | 2 |
| Mixer grinder (self-owned) | 1 |
| Internet connection | 1 |
| Wet grinder (self-owned) | 1 |
| Total | 20 |
| **SES** | **Range of score** |
| I (upper) | 15-20 |
| II (upper middle) | 12-14 |
| III (lower middle) | 9-11 |
| IV (upper lower) | 7-8 |
| V (lower) | 0-6 |

SES: Socioeconomic status

The eigenvalue indicates the percentage of variation in the total data explained for each principal component. If it is lesser, then it could be due to less assets included or the complexity of correlations between them.[14] The higher value in the present study gives assurance that assets were comprehensively included.

Other statistical issues in PCA related to a number of assets included are clumping and truncation.[7,14,16] Clumping or clustering is described as households being grouped together in a small number of distinct clusters. Truncation indicates a more even distribution of SES but spreads over a narrow range resulting in difficulty to differentiate between socioeconomic groups. The histogram constructed in our study showed the absence of clumping and truncation that ensured the range of assets included was broad enough to avoid these issues in PCA.

Health researches carried out in India commonly used MK (1976) scale for urban setting; B G Prasad (1961) for urban and rural; and Uday Pareek scale (1964) for rural population.[17,18] B G Prasad scale is purely income based and Uday Pareek is for rural setting; hence, we used the MK scale to compare the finding of our newly constructed tool. In our present study, we got moderate correlation and poor agreement between the SES classes obtained between the newly constructed 10-item asset-based scale and the MK scale. This was similar to a previous study that compared the wealth indices estimated between various standard tools, with the agreement ranged between kappa of 0.01 and 0.15.[19]

The reasons could be firstly our scale is purely asset based and MK scale takes into consideration the education, occupation, and per capita family income. Second, inquiring about income is considered as a sensitive issue and so it suffers validity. Third, the theories governing their construction were different.[20] Besides these shortfalls, many variables involved in the MK scale are subjective as clear definitions of the same are not provided.[3]

In the current study, sampling bias was reduced by including a broad range of urban region-specific assets that were able to discriminate SES and the same was confirmed by a higher and significant statistical measure of sampling adequacy. There was excellent and significant agreement between the wealth quintiles calculated based on the total of 20 assets and the final reduced 10 assets. As the study was based on a representative sample of 500 urban households, it has adequate discriminative power, and hence, the results are statistically generalizable and can be used in different surveys and studies among the urban population.

## Conclusions

A simple and reliable household asset-based SES scale using the PCA technique was created for the urban population of Puducherry. This is a better way of estimating SES at field studies as it has fewer assets and avoids those problems associated with income and consumption-based methods such as recall bias, social desirability bias, seasonality, and data collection time.


## Financial support and sponsorship

Nil.


## Conflicts of interest

There are no conflicts of interest.

## References


1. Krieger N, Williams DR, Moss NE. Measuring social class in US public health research: Concepts, methodologies, and guidelines. Annu Rev Public Health 1997;18:341-78.
2. Krieger N. A glossary for social epidemiology. J Epidemiol Community Health 2001;55:693-700.
3. Sharma R, Saini NK. A critical appraisal of kuppuswamy's socioeconomic status scale in the present scenario. J Fam Med Prim Care 2014;3:3-4.
4. Galobardes B, Shaw M, Lawlor DA, Lynch JW, Davey Smith G. Indicators of socioeconomic position (part 1). J Epidemiol Community Health 2006;60:7-12.
5. Galobardes B, Shaw M, Lawlor DA, Lynch JW, Davey Smith G. Indicators of socioeconomic position (part 2). J Epidemiol Community Health 2006;60:95-101.
6. Montgomery MR, Gragnolati M, Burke KA, Paredes E. Measuring living standards with proxy variables. Demography 2000;37:155-74.
7. Filmer D, Pritchett LH. Estimating wealth effects without expenditure data-or tears: An application to educational enrollments in states of India. Demography 2001;38:115-32.
8. Howe LD, Galobardes B, Matijasevich A, Gordon D, Johnston D, Onwujekwe O, *et al*. Measuring socio-economic position for epidemiological studies in low- and middle-income countries: A methods of measurement in epidemiology paper. Int J Epidemiol 2012;41:871-86.
9. Bollen KA, Glanville JL, Stecklov G. Socio-economic status, permanent income, and fertility: A latent-variable approach. Popul Stud (Camb) 2007;61:15-34.
10. Creswell JW, Clark VL. Designing and Conducting Mixed Methods Research. California: SAGE; 2011. p. 489.
11. Tabachnick B, Fidell L. Using Multivariate Statistics. 4th ed.. Boston: Allyn and Bacon; 2001. p. 588.
12. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015;42:533-44.
13. National Family Health Survey IV: Household Questionnaire. International Institute for Population Sciences. Available from: http://rchiips.org/NFHS/NFHS4/schedules/NFHS-4Household.pdf. [Last accessed on 2021 Jan 20].
14. Vyas S, Kumaranayake L. Constructing socio-economic status indices: How to use principal components analysis. Health Policy Plan 2006;21:459-68.
15. Hjelm L, Mathiassen A, Miller D, Wadhwa A. Creation of a Wealth Index. World Food Programme. Available from: https://docs.wfp.org/ api/documents/WFP-0000022418/download/?_ga=2.249741894.11059 8194 2.1556043003-679515948.1556043003. [Last accessed on 2021 Jan 20].
16. McKenzie D. Measuring inequality with asset indicators. J Popul Econ 2005;18:229-60.
17. Holyachi S, Santosh A. Socioeconomic status scales-An update. Ann Community Health 2013;1:24-7.
18. Singh T, Sharma S, Nagesh S. Socio-economic status scales updated for 2017. Int J Res Med Sci 2017;5:3264-7.
19. Kattula D, Venugopal S, Velusamy V, Sarkar R, Jiang V, S MG, *et al*. Measuring poverty in southern India: A comparison of socio-economic scales evaluated against childhood stunting. PLoS One 2016;11:e0160706.
20. Deshmukh PR, Akkilagunta S. Socioeconomic status: A theoretical framework for the development and use of assessment tools. Indian J Community Fam Med 2020;6:4.