Published in final edited form as:
*AIDS Behav*. 2013 February ; 17(2): 700–709. doi:10.1007/s10461-012-0236-8.

# Measuring socioeconomic inequality in the incidence of AIDS: rural-urban considerations

**Theophile Niyonsenga, PhD**[1] **[Associate Professor]**, **Mary Jo Trepka, MD, MSPH**[1] **[Associate Professor]**, **Spencer Lieb, MPH**[2] **[Senior Epidemiologist]**, and **Lorene M. Maddox, MPH**[3] **[Data Analysis Manager]**

[1]Department of Epidemiology and Biostatistics, Robert Stempel College of Public Health and Social Work Florida International University, Miami, Florida

[2]Florida Consortium for HIV/AIDS Research/The AIDS Institute, Tampa, Florida

[3]Bureau of HIV/AIDS, Florida Department of Health, Tallahassee, Florida

## Abstract

Low socioeconomic status (SES) influences the risk of acquiring human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS) and thus should be considered when analyzing HIV/AIDS surveillance data. Most surveillance systems do not collect individual level SES data but do collect residential ZIP code. We developed SES deprivation indices at the ZIP code tabulation area and assessed their predictive validity for AIDS incidence relative to individual neighborhood-level indicators in Florida using reliabililty analysis, factor analysis with principal component factorization, and structural equation modeling. For urban areas an index of poverty performed best, although the single factor poverty also performed well. For rural areas no index performed well, but the individual indicators of no access to a car and crowding performed well. In rural areas poverty was not associated with increased AIDS incidence. Users of HIV/AIDS surveillance data should consider urban and rural areas separately when assessing the impact of SES on AIDS incidence.

### Keywords

Socioeconomic status; Deprivation; Urban/rural neighborhood; Poverty index; Acquired immunodeficiency syndrome

## INTRODUCTION

In the United States (US) and other developed countries, there has been consistent evidence that the incidence of human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS) is higher among people and communities with lower socioeconomic status (SES) [1-4]. Low SES, viewed as a proxy of deprivation, can influence the risk of acquiring HIV infection and the progression to AIDS at both the individual and community levels. At the individual level lower SES may be associated with less access to health care [1, 5-7] and greater stress [8]. A deprived community exposes its residents to conditions such as high unemployment and may be more disorganized and chaotic creating a culture permitting higher risk sexual practices and illicit drug use [9, 10, 11].

Corresponding author Dr. Mary Jo Trepka, Associate Professor Department of Epidemiology and Biostatistics, Stempel College of Public Health and Social Work, Florida International University, University Park, HLSII, Room 595, 11200 SW 8th Street, Miami, FL 33199.


NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The HIV/AIDS surveillance systems throughout the US do not generally collect data about individual level SES such as income or education [12, 13]. For that reason, neighborhood socioeconomic data has been used as a proxy for individual SES by many researchers in studying AIDS incidence or survival in the US and other developed countries [1-7, 14-18] as well as in studying health disparities for other diseases and health service use inequalities [19-23]. Socioeconomic status is a complex construct that involves both material and social factors [23]. To measure SES at the neighborhood level, researchers have used the US Census variables: poverty, predominantly working class neighborhood, and low educational level [7, 21]; median annual income [16]; percentage of owner-occupied homes [24-26]; and percentage of people living below the poverty line, number of people living per square mile, and percentage of households with an annual income at least $150,000 to capture wealth [4, 27, 28]. While SES measured at the smallest geographic level such as census blocks and tracts is likely to give the most precise estimate of the conditions that individual people are exposed to, most surveillance systems do not have the resources to geo-code cases within the system [18, 25]. However, surveillance systems do collect the residential ZIP codes; thus rates by ZIP code can be easily calculated.

Because there are so many SES indicators available from the US Census, it is helpful for analyses to have SES indices composed of indicators that reflect various domains of neighborhood deprivation. Several indices have been proposed and used in Great Britain, Italy, Spain, France, Sweden, Japan, and Canada [19, 23, 29-33]. However, they may not be appropriate for the US economic and social situation. Furthermore, they may not work well in rural as well as in urban situations [20, 22, 34-36]. Indeed, the choice of indicators used in the indices and the spatial unit of observation depend upon the cultural, social and demographic contexts, as well as the sociodemographic data available at the geographic unit of analyses [25, 26]. For the US context, researchers developed SES indices which they validated using mortality and cancer incidence rates [25], sexually transmitted infections, tuberculosis and violence [25, 37], low birth and childhood lead poisoning data [26] and perinatal health outcomes [28]. However, indices have not been validated for AIDS incidence data and have not been assessed for rural and urban settings independently. The few studies that considered the rural/urban comparisons were performed in the UK [20, 22], and those looking at AIDS assessed SES-specific variables and not indices [4, 12].

The empirical indices assumed that SES indicators contribute equally to measure area-level deprivation. Indeed, the widely used indices, Townsend index [29, 31, 38] and Carstairs index [30, 31], are computed as an unweighted sum of the transformed scores of the indicators involved in each index. Newly developed indices adopted a factor-analytic model that takes into account the importance of each indicator in the index [23, 25, 26, 28, 39]. Some authors using the factor-analytic approaches only used the first principal component as the SES index [26, 28, 37, 39], assuming therefore that the SES construct has one single domain. Other authors identified and used 2 distinct components known as material and social deprivation indices [23]. The choice of SES indicators has been based on empirical evidence of relationships between each indicator and health outcomes [18, 23, 26], but varied substantially between studies [28] and may potentially be measuring different components of SES deprivation. Researchers assumed as well that SES indicators were reliable in measuring neighborhood deprivation. However, indicators may exhibit relatively different levels of reliability and homogeneity depending on the countries and geographic sub-unit within a given country (eg state, census tract, ZIP code) at which the census SES data has been collected from. There may also be more heterogeneity within rural areas than urban areas [20].

The first objective of this study was to develop SES indices at the ZIP code level. The second objective was to assess the predictive validity of the SES indices by correlating them with health outcomes well known to be associated with socioeconomic deprivation [4, 20, 25, 26, 39, 40], namely, the AIDS incidence data. Further objectives were to (1) estimate the reliability of indicators in measuring the SES deprivation for rural and urban areas, (2) identify clusters of indicators defining different components of SES deprivation, and (3) to determine if these deprivation indices would be better than any single SES indicator in explaining geographic disparities in AIDS incidence rates. Florida is a good state for this study because Florida has been particularly affected by the HIV/AIDS epidemic and in 2009 had the highest HIV case rate and the third highest AIDS case rate in the United States [41]. Although there have been no published reports about rural/urban differences in AIDS case rates in Florida and most of the cases are reported from several large urban counties, AIDS is widespread in Florida with 65 of the 67 counties reporting cases in 2009 [42].

## MATERIALS AND METHODS

### Unit of analysis

The choice of the neighborhood level of study is crucial as it defines the degree of homogeneity in the SES indicators [23, 25]. For this study, the ZIP code tabulation area (ZCTA) was selected as the unit of analysis, since it is the smallest geographic unit at which SES data are readily available from surveillance data in Florida.

A ZCTA is a statistical entity developed by the US Census Bureau for the tabulation of summary statistics from the 2000 US Census [43]. It is built by aggregating the 2000 US Census blocks, whose addresses use a given ZIP code, into a ZCTA which gets that ZIP code assigned as its ZCTA code. The 1990 US Census reported data directly by ZIP codes, but since ZIP codes are determined by the US Postal Service and are not constructed from the same geographic building blocks as the US Census, ZCTAs were created in an attempt to reflect ZIP codes as much as possible using census geography. Contrary to a ZIP code, which contains on average 30,000 persons, there are no restrictions limiting how large or small a ZCTA can be in terms of either a minimum/maximum number of housing units or surface areas. As a result, ZCTAs may be as small as a few city blocks or cover many square miles.

A total number of 972 ZCTAs in Florida was considered. Those with correctional facilities were excluded because the ZIP code of a correctional facility may be different from the place where the person was actually exposed to HIV. Those ZCTAs with missing values on most of the indicators, especially ZCTAs with total population less than 30, were also excluded, leaving 854 ZCTAs (87.7%).

The ZCTAs were divided into rural and urban areas. The US Census Bureau classifies as urban all territory, population, and housing units located within an urbanized area (UA) or an urban cluster (UC). It delineates UA and UC boundaries to encompass densely settled territory, which consists of: (1) core census block groups or blocks that have a population density of at least 1,000 people per square mile and (2) surrounding census blocks that have an overall density of at least 500 people per square mile. The Census Bureau's classification of rural consists of all territory, population, and housing units located outside of UAs and UCs. The rural-urban status was assessed based on the Rural-Urban Categorization (RUCA) data codes developed by the University of Washington [44, 45]. The RUCA system was recommended because it is available at the zip code level. Therefore, it is more precise than alternative categorizations, and it provides a great deal of flexibility as the codes can be collapsed or combined in several different ways [45, 46]. In the RUCA algorithm, the categorization D was used as it is the simplest, defining urban as the Census Bureau does and including as well all the places that have 30% or more of their workers going to a US

Census Bureau-defined urbanized area. The ultimate sample of 854 ZCTAs consisted of 739 (86.5%) from the urban areas and 115 (13.5%) from rural areas.

## Socioeconomic variables and AIDS incidence data

A total of 13 variables drawn from numerous US Census measures of neighborhood socioeconomic conditions and related to the broad sub-domains of poverty, income and wealth, employment, housing and crowding, education, occupation and transportation, were the most commonly used indicators in various research studies [4, 21, 27, 28]. They were chosen to enable the measurement of SES in a comprehensive manner while taking into account its complexity [26, 37].

The following variables were selected based on their theoretical and empirically evidenced relevance (Table I): transportation (assessed by the percentage of households without access to a car) [29, 38]; crowding (percentage of households with more than one person per room) [21, 27-29, 38]; poverty (percentage of people below the federally-defined poverty line, a threshold varying by size and age composition of the household) [21, 25, 26]; wealth (percentage of owner-occupied homes worth at least $300,000) [25, 26]; median income (median household income in 1999) [21, 25-27]; low income (percentage of households with annual income less than $15,000) [25, 26]; high income (percentage of households with annual income of at least $150,000) [25, 26]; income disparity (a proxy of the measure of income inequality known as the Gini coefficient and defined as the logarithmic transformation of 100*ratio of the percentage of households with income less than $10,000 to the percentage of households with income of at least $50,000) [27]; low education (percentage of persons, aged 25 years and older, with less than $12^{th}$ grade education) [25, 26]; high education (percentage of persons, aged 25 years and older with a graduate or professional degree) [25, 26, 28]; renting (percentage of households who rent) [29, 38]; unemployment (percentage of persons, 16 years and older, who are unemployed) [21, 27-29, 38]; and low working class (percentage of persons employed in the predominantly low working class occupations) [21, 25, 26].

The ZCTA AIDS incidence data were drawn from the Florida HIV/AIDS Reporting System (HARS) and included all cases with a diagnosis of AIDS between 1998 and 2002, of which there were 20,528 (2.9% and 97.1% respectively in rural and urban ZCTAs). For each ZCTA the average 5-year (1998-2002) age-adjusted HIV/AIDS incidence rate was calculated using the Florida age-distribution of the population reported in the 2000 US Census. These years were chosen because a 5-year average could be calculated, and these were the closest years to the 2000 US Census.

## Data analysis

The analysis involved standardized scores of the SES indicators. As done for the Townsend deprivation indices [25, 26], and for others indices developed [21, 39], standardization was used to take into account the different measurement scales (eg, poverty, median income, and income disparity). The analysis involved the following steps: (1) reliability analysis of the whole SES scale, (2) factor analysis with principal components factorization, and (3) structural equation modeling framework with AIDS incidence as the outcome variable and the latent factor(s) defining SES indices as the predictors. Analysis was performed for rural and urban settings.

For the reliability analysis, the Cronbach's Alpha coefficient was used as a measure of the internal consistency of items in the scale [47, 48]. Cronbach's alpha coefficient is widely believed to indirectly indicate the degree to which a set of items measures a single unidimensional latent construct. The higher the coefficient, the more reliable the generated

NIH-PA Author Manuscript

NIH-PA Author Manuscript

summated scale is. A sequential variable's deletion based on “Cronbach's Alpha if item deleted” (high value) and “Corrected item-total correlation” (low value) was performed to select the most important indicators of the scale and to delete the less consistent within the total scale. Some of the indicators were reversed to align the items scoring in the same direction; that is, the higher the score on each indicator, the higher the SES index.

A factor analysis with principal components factorization, followed by a varimax rotation [49, 50], was used to identify clusters of inter-related indicators with larger loadings, defining clear theoretical and meaningful factors [51]. The number of factors retained was based on the number of eigenvalues greater than 1 [51]. Although the focus was on clustering variables into sub-domains of SES deprivation, the choice of the factorization was intentionally aimed to determine the minimum number of factors that will account for the maximum variance in the multivariate data, instead of the common variance shared by all the SES indicators [52].

A structural equation model, linking each latent factor (with its indicators) to AIDS incidence rates, was constructed to estimate the extent to which the SES indices correlate with AIDS incidence [39, 53]. In this SEM framework, the focus was more on the percentage of variability in the AIDS incidence explained by the latent factor, rather than the fit indices of a theoretical model that would fit the data at hand [53].

To evaluate the extent to which SES indicators and indices are associated with AIDS incidence, 2 measures were used: correlation coefficient (CC) and incidence rate ratio (IRR). The CC assumed a linear relationship between AIDS incidence and each variable. The use of IRR was intended to discriminate the extreme groups, the most deprived and the least deprived groups. For each SES indicator and each SES index, quartiles were computed. Poisson regression models [27, 37, 54] were fitted to AIDS incidence and reported IRR values compared the highest quartile to the lowest quartile.

Statistical analyses were performed using SPSS and Amos version 18 for reliability, factor analysis and SEM modeling, as well SAS 9.2 for extracting area-based AIDS incidence and HIV/AIDS deaths from the HARS data [55-57].

# RESULTS

## Reliability analysis and SES indices

There were 9 indicators with Cronbach's alpha of 0.923 for urban ZCTAs and 4 indicators with Cronbach's alpha of 0.889 for rural ZCTAs, chosen from the 13 indicators entered in the reliability analysis (Table II). Four items not selected in urban and rural ZCTAs correspond to indicators defining the Townsend index (no access to car, crowding, unemployment and renting a house) [29]. In rural ZCTAs, these indicators achieved a reliability coefficient of 0.403, which is lower than the usually accepted 0.7 cut-off. In urban ZCTAs, Cronbach's alpha coefficient was 0.728 for these 4 indicators.

For rural areas, only one component was defined, grouping together poverty status, low income, income disparity and median income. This component, labeled as the “*poverty index*” (different from the poverty status indicator), explained 75.1% of the total variance in the indicators. In urban areas, a 2-factor solution was identified, and explained 81.8% of the total variance. The first cluster of indicators included poverty status, low income, income disparity , and low education . These SES indicators were used to create the “*poverty index*”, with high scores indicating the most deprived. The second cluster of indicators captured median income, high class work, high income, high education and wealth. They were used to create the “*affluence index*”, with high scores indicating the lack of affluence. Observed scores of the Kaiser-Meyer-Olkin (KMO) index of sampling adequacy were 0.836 and 0.810 for rural and urban areas, respectively, justifying the adequacy of the factor analysis. With the available sample sizes, the observed percentages of variance accounted for were detected with power of 99% [49].

### AIDS incidence with SES indicators and indices

In rural areas, there were weak and statistically non-significant correlations between AIDS incidence and individual SES indicators, except for income disparity and 3 of the Townsend indicators: no access to a car, crowding and renting a house (Table III). For these indicators, correlations ranged from 0.203 to 0.250 in absolute value and were statistically significant although still in the weak range. For urban ZCTAs, observed correlations between AIDS incidence and SES indicators ranged from 0.093 (wealth) to 0.621 (no access to a car), and were all statistically significant (Table III). The Townsend indicators were positively associated with the AIDS incidence with particularly high correlation values (Table III). Even though the correlations for urban ZCTAs were all statistically significant, some of them were small (eg for the items that form the affluence index and the affluence index itself) and significant due to the large sample size.

In terms of the ability of SES indicators to differentiate extreme groups, incidence rate ratio (IRR) values ranged from 1.710 (wealth) to 8.385 (no access to a car) in urban areas, meaning that the most deprived groups had higher AIDS incidence rates than less deprived groups. In rural areas, Townsend indicators discriminated positively the extreme groups with IRR values from 2.113 (no access to a car) to 3.966 (crowding). The remaining SES indicators showed opposite relationships with IRR values from 0.463 (income disparity) to 0.927 (poverty); that is, less deprived areas had higher AIDS incidence rates than the most deprived areas (Table III).

Structural equation models relating SES indices (Poverty, Affluence, and Townsend-like indices) to AIDS incidence rates are graphically illustrated in Figures 1 to 5, along with standardized coefficients, the estimated construct reliability (CR), the goodness-of-fit index (GFI), the comparative fit index (CFI) and the coefficient of determination ($R^2$). Other measures to assess the extent to which AIDS incidence correlates with SES indices, such as CC and IRR are presented in Table III.

The observed fit indices were 0.75 and over, establishing that, overall, the models are acceptable. However, acceptable fit indices do not imply the relationships are strong. Indeed, although models were acceptable, not all the paths were highly significant (Figures 1-5). The Townsend-like index had a very low construct reliability in rural areas (CR = 0.386). Other construct reliability coefficients ranged from 0.721 to 0.924, reflecting again adequate consistency among indicators defining each SES index.

In rural ZCTAs (Figure 1), the poverty index was negatively associated with AIDS incidence, implying that the more deprived the ZCTA is, the fewer incident cases of AIDS reported, as if deprivation lowers the risk of HIV/AIDS. Moreover, poverty index explained only 6.4% of the variance in AIDS incidence rates and showed less discrimination power as measured by IRR values (Table III). In urban ZCTAs, contrary to the rural area, poverty index was found positively associated with AIDS incidence with both CC value (0.584) and IRR value (8.558) statistically significant. Moreover, it explained 34.1% of the total variability in the AIDS incidence rates (Figure 2, Table III).

The affluence index did not perform well in urban areas. The observed positive correlation with AIDS incidence was less than 0.25, and the IRR value was less than 4 (Table III).

Although CC and IRR values were statistically significant, the model explained 2% of the total variance in incidence rates (Figure 3).

For the Townsend index, SEM models revealed very low and non-significant coefficients of variables for the factor for rural ZCTAs (Figure 4), as well as moderate to high significant loadings for urban ZCTAs (Figure 5). The Townsend index explained 62.7% of the total variance in the AIDS incidence data for urban ZCTAs, but only 25.8% in rural ZCTAs. In terms of the ability to discriminate extreme groups, IRR reached a value near 8 in urban ZCTAs, but only near 4 in rural ZCTAs (Table III). Townsend index and AIDS incidence were positively correlated, although not significantly for rural ZCTAs. Surprisingly, correlations were generally higher than those observed for the other constructs defined above (Table III).

To what extent did individual SES indicators outperform the developed indices? In terms of the power of discrimination (IRR values), crowding appeared to be better than poverty and Townsend indices for rural ZCTAs (Table III). In urban ZCTAs, the poverty index outperformed individual SES indicators but was only marginally better than poverty status. However, access to a car outperformed affluence and Townsend indices in urban ZCTAs. In terms of the correlation, no access to a car outperformed all the indices except the Townsend index in urban areas. In rural areas, renting a house outperformed all indices except the Townsend index.

## DISCUSSION

The study findings revealed that SES indicators do not consistently define one unique construct. Instead, they define various dimensions of the SES deprivation construct. Therefore, instead of using one single factor-index [26, 28, 39], poverty, affluence, and Townsend-like indices may define 3 dimensions of the SES deprivation concept. A more comprehensive and fully delineated SES deprivation construct should include various aspects of social and economic life [27]. Pampalon et al. [23] proposed 2 dimensions of the SES deprivation, material and social deprivation, tapping into indicators that measure the quality of social life, such as the state of being a single parent family, living alone, and being separated or divorced or widowed.

Both reliability and factor analyses were useful as exploratory tools, enabling discovery of various structures in the SES indicators, and as a measure of the internal consistency of the clusters defining dimensions of the SES deprivation. However, these analyses need to be applied to an exhaustive set of SES indicators. Studies have used a varying number of different SES indicators. For instance, Pampalon et al. [23] selected only 6 indicators for their study, while Messer et al. [28] retained 8 of the 20 indicators initially selected to compute the index. Krieger et al. [26] used 11 indicators while Singh and Siahpush [27] factor-analyzed 17 variables.

Our study findings suggested that rural and urban areas are different in terms of the importance of SES indicators, deprivation indices, and their associations with AIDS incidence. Barnet et al.'s [22] findings also indicated that SES deprivation may differ between urban and rural areas, suggesting the need to use and differentiate between SES indicators that are more sensitive to rural and urban settings. This emphasizes once again the need to tap into various aspects of the rural-urban socio-economic life and challenges.

Our study showed that SES deprivation indices were associated with AIDS incidence. Findings suggested close relationships and high discrimination ability of the proposed indices in urban areas, but not in rural areas. Haynes and Gale [20] reported similar findings in which both death and limiting long term illness rates were most closely predictable by deprivation in large urban centers, and much less predictable in inner rural areas. These rural-urban differences may be a function of either more heterogeneity in SES in rural areas and fewer indicators to account for this heterogeneity, or not enough variability in SES indicators in rural areas to explain individual-level variability in the risk of AIDS, or the existence of different factors (HIV/AIDS risk behaviors and other psychosocial behaviors) that put people at increased risk of HIV/AIDS in rural and urban areas. Rural-urban differences may reflect perhaps much more opportunity for transmission in urban locations, given the nature of HIV transmission (eg needle sharing, unprotected sexual intercourse).

The structural equation model-based approach provided a way to maximize SES data available by constructing latent variables, allowing different weights for SES indicators. However, this does not necessary lead to perfect indices capable of explaining the variability in AIDS incidence and/or discriminating highly deprived areas from least deprived areas better than single SES indicators. Study results showed that SES indices behaved only slightly better than some of the single indicators (eg poverty status, low income and transportation). These findings are consistent with results from Krieger et al. [25, 26] that suggest poverty status would be a better simple measure to monitor socioeconomic disparities. In terms of income, Simon et al. [12] found that 3 income-related indicators (income disparity, low income and median income) showed relatively strong relationships with AIDS rates, although some were outperformed by the Townsend-related SES indicators.

Our study findings bear some limitations. First, the selection of SES measures is limited by what is collected by the US Census, and the ZCTAs used by the US Census Bureau are relatively large areas, likely more heterogeneous, rendering less easy the task of capturing the meaning of contexts being measured and pinpointing existing socioeconomic gradients. Moreover, people living in the so-called rural areas may be in a so-called poor ZCTA but also be relatively close to a more affluent ZCTA (adjacency to small and large urban centers). There is no clear-cut delineation of SES indicators following rural and urban ZCTAs.

Second, for AIDS incidence rates, the location of reported case is not necessarily where a person was infected because the time between HIV infection and AIDS onset can be very long. People sometimes move from the place they were infected with HIV to a place where they might get better care. A study using a national probability sample of HIV-infected people found that about 17% reported moving to a non-adjacent county or farther after HIV diagnosis; 85% of these moves were from one urban county to another urban county, 8% were from urban to rural areas, and 6% from rural to urban areas [58]. Moreover, ZIP code location obtained by self-report or from address in medical reports cannot be assumed to correspond to census-defined ZCTAs.

Third, excluding zip codes with correctional facilities may have led to the exclusion of cases that were not incarcerated. However, to attribute the SES of the zip code to the environment of the incarcerated people would almost certainly have led to misclassification. Finally, the small sample size of rural areas limits the study of SES deprivation indices and their association with AIDS incidence.

Although the small sample size of rural areas is a concern, we feel that it is adequate for what we aimed to do. Yurdugul [59] showed that if the largest eigenvalue is between 3.00 and 6.00 (which was the case in this study), the required minimum sample size, n = 100, will be adequate for an unbiased estimator of the coefficient alpha. Moreover, based on the ratio of indicator variables, the minimum sample sizes would be 100 for the structural equations modeling (SEM) [49, 52, 59]. However, for sample sizes less than 200, problems of nonconvergence to a proper solution and improper solutions may appear in SEM, but may be reduced by strong internal consistency [60].

## CONCLUSION

In conclusion, this study provides evidence that: (1) SES deprivation is defined by more than one dimension; (2) there is an inverse association between SES indices and AIDS incidence; (3) rural and urban areas are different in terms of SES indicators, SES deprivation indices, and their association with AIDS incidence; (4) there is a need for more data on SES indicators, especially in rural areas, to tap into various aspects of the rural-urban socioeconomic life; and (5) the availability of SES data at smaller geographic levels would ensure a higher degree of homogeneity in the socioeconomic conditions, enhancing SES deprivation indices and their predictive ability. However, more research work is needed to obtain a consensus on all the SES indicators pertinent to socioeconomic inequalities, tapping into all aspects of rural and urban daily life.

The validation process needs to be repeatedly undertaken until such time as we obtain valid and stable SES indicators and the estimates of their relative importance (weights) in defining the various domains of SES deprivation. This would allow for the calculation of SES deprivation scores by plugging the indicators' values into the formulas. In the meantime one can consider using poverty as a single variable or poverty index when analyzing the association between SES and AIDS incidence. However, caution is advised in the use of these variables for analyses involving rural areas. More importantly, HIV/AIDS surveillance systems should focus on collecting individual-level SES data, instead of relying on neighborhood SES data as proxy of individual SES measures.

## Acknowledgments

The project described was supported by Award Number R01MD004002 from the National Institute on Minority Health and Health Disparities. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Minority Health and Health Disparities or the National Institutes of Health. The authors wish to thank Khaleeq Lutfi, MPH, and Elena Pavlova, MD, MHSA, for assistance in the dataset preparation.

## References

1. Fife D, Mode C. AIDS incidence and income. J Acquir Immune Defic Syndr. 1992; 5(11):1105–1110. [PubMed: 1403639]
2. Murrain M, Barker T. Investigating the relationship between economic status and HIV risk. J Health Care Poor Underserved. 1997; 8(4):416–423. [PubMed: 9334534]
3. Wallace RG. AIDS in the HAART era: New York's heterogeneous geography. Soc Sci Med. 2003; 56(6):1155–1171. [PubMed: 12600355]
4. Zierler S, Krieger N, Tang Y, et al. Economic deprivation and AIDS incidence in Massachusetts. Am J Public Health. Jul; 2000 90(7):1064–1073. [PubMed: 10897184]
5. Arnold M, Hsu L, Pipkin S, McFarland W, Rutherford GW. Race, place and AIDS: the role of socioeconomic context on racial disparities in treatment and survival in San Francisco. SocSci Med. 2009; 69(1):121–128.
6. Cunningham WE, Hays RD, Duan N, et al. The effect of socioeconomic status on the survival of people receiving care for HIV infection in the United States. JHealth Care Poor Underserved. 2005; 16(4):655–676. [PubMed: 16311491]
7. Katz MH, Hsu L, Lingo M, Woelffer G, Schwarcz SK. Impact of socioeconomic status on survival with AIDS. Am J Epidemiol. 1998; 148(3):282–291. [PubMed: 9690366]
8. Adler NE, Boyce T, Chesney MA, Folkman S, Syme SL. Socioeconomic inequalities in health: no easy solution. JAMA. 1993; 269(24):3140–3145. [PubMed: 8505817]

9. Cohen D, Mason K, Bedimo A, Scribner R, Basolo V, Farley TA. Neighborhood physical conditions and health. Am J Public Health. 2003; 93(3):467–471. [PubMed: 12604497]
10. Farley TA. Sexually transmitted diseases in the southeastern United States: location, race, and social context. Sex Transm Dis. 2006; 33(7):S58–S64. [PubMed: 16432486]
11. Poundstone KE, Strathdee SA, Celentano DD. The social epidemiology of human immunodeficiency virus/acquired immunodeficiency syndrome. Epidemiol Rev. 2004; 26(1):22–35. [PubMed: 15234945]
12. Simon PA, Hu DJ, Diaz T, Kerndt PR. Income and AIDS rates in Los Angeles County. AIDS. 1995; 9(3):281–284. [PubMed: 7755917]
13. Krieger N, Chen JT, Ebel G. Can we monitor socioeconomic inequalities in health? A survey of U.S. health departments' data collection and reporting practices. Public Health Rep. 1997; 112(6): 481–491. [PubMed: 10822475]
14. Harrison KM, Ling Q, Son R, Hall I. County-level socioeconomic status and survival after HIV diagnosis, United States. Ann Epidemiol. 2008; 18(12):919–927. [PubMed: 19041591]
15. Joy R, Druyts EF, Brandson EK, et al. Impact of neighborhood-level socioeconomic status on HIV disease progression in a universal health care setting. J Acquir Immune Defic Syndr. Apr 1; 2008 47(4):500–505. [PubMed: 18197117]
16. McFarland W, Chen S, Hsu L, Schwarcz S, Katz M. Low socioeconomic status is associated with a higher rate of death in the era of highly active antiretroviral therapy, San Francisco. J Acquir Immune Defic Syndr. 2003; 33(1):96–103. [PubMed: 12792361]
17. Rapiti E, Porta D, Forastiere F, Fusco D, Perucci CA. Socio-economic status and survival of persons with AIDS before and after the introduction of highly active antiretroviral therapy. Epidemiology. 2000; 11(5):496–501. [PubMed: 10955400]
18. Zierler S, Krieger N. Reframing women's risk: social inequalities and HIV infection. Annu Rev Public Health. 1997; 18:401–436. [PubMed: 9143725]
19. Tello JE, Jones J, Bonizzato P, Mazzi M, Amaddeo F, Tansella M. A census-based socio-economic status (SES) index as a tool to examine the relationship between mental health services use and deprivation. Soc Sci Med. 2005; 61(10):2096–2105. [PubMed: 15922500]
20. Haynes R, Gale S. Deprivation and poor health in rural areas: inequalities hidden by averages. Health Place. 2000; 6(4):275–285. [PubMed: 11027953]
21. Diez-Roux AV, Kiefe CI, Jacobs DR Jr. et al. Area characteristics and individual-level socioeconomic position indicators in three population-based epidemiologic studies. Ann Epidemiol. 2001; 11(6):395–405. [PubMed: 11454499]
22. Barnett S, Roderick P, Martin D, Diamond I, Wrigley H. Interrelations between three proxies of health need at the small area level: an urban/rural comparison. Journal of Epidemiol Community Health. 2002; 56(10):754–761.
23. Pampalon R, Hamel D, Gamche P, Raymond G. A deprivation index for health planning in Canada. Chronic Dis Can. 2009; 29(4):178–191. [PubMed: 19804682]
24. Rice RJ, Roberts PL, Handsfield HH, Holmes KK. Sociodemographic distribution of gonorrhea incidence: implications for prevention and behavioral research. Am J Public Health. 1991; 81(10): 1252–1258. [PubMed: 1928521]
25. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter? The public health disparities geocoding project (US). Am J Epidemiol. 2002; 156(5):471–482. [PubMed: 12196317]
26. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: the public health disparities heocoding project (US). J Epidemiol Community Health. 2003; 57(3):186–199. [PubMed: 12594195]
27. Singh GK, Siahpush M. Increasing inequalities in all-cause and cardiovascular mortality among US adults aged 25-64 years by area socioeconomic status, 1969-1998. Int J Epidemiol. 2002; 31:600–613. [PubMed: 12055162]
28. Messer LC, Laraia BA, Kaufman JS, et al. The development of a standardized neighborhood deprivation index. J Urban Health. 2006; 83(6):1041–1062. [PubMed: 17031568]

29. Townsend P. Deprivation. J Soc Policy. 1987; 16:125–146.
30. Carstairs V, Morris R. Deprivation and mortality: an alternative to social class? Community Med. 1989; 11(3):210–219. [PubMed: 2605887]
31. Carstairs V. Deprivation indices: Their interpretation and use in relation to health. J Epidemiol Community Health. 1995; 49(Suppl 2):S3–S8. [PubMed: 8594130]
32. Benach J, Yasui Y. Geographical patterns of excess mortality in Spain explained by two indices of deprivation. J Epidemiol Community Health. 1999; 53(7):423–431. [PubMed: 10492736]
33. Noble, M.; McLennan, D.; Wilkinson, K., et al. The English Indices of Deprivation 2007. Office of the Deputy Prime Minister; London: 2008.
34. Niggebrugge A, Haynes R, Jones A, Lovett A, Harvey I. The index of multiple deprivation 2000 access domain: a useful indicator for public health? Soc Sci Med. 2005; 60(12):2743–2753. [PubMed: 15820584]
35. Martin D, Brigham P, Roderick P, Barnett S, Diamond I. The (mis)representation of rural deprivation. Environment and Planning A. 2000; 32(4):735–751.
36. Gilthorpe MS, Wilson RC. Rural/urban differences in the association between deprivation and healthcare utilization. Soc Sci Med. 2003; 57(11):2055–2063. [PubMed: 14512237]
37. Krieger N, Waterman PD, Chen JT, Soobader MJ, Subramanian SV. Monitoring socioeconomic inequalities in sexually transmitted infections, tuberculosis, and violence: geocoding and choice of area-based socioeconomic measures - the public health disparities geocoding project (US). Public Health Rep. 2003; 118(3):240–260. [PubMed: 12766219]
38. Townsend P, Simpson D, Tibbs N. Inequalities in health in the city of Bristol: a preliminary review of statistical evidence. Int J Health Serv. 1985; 15(4):637–663. [PubMed: 3841091]
39. Hogan JW, Tchernis R. Bayesian factor analysis for spatially correlated data, with application to summarizing area-level material deprivation from census data. Journal of the American Statistical Association. 2004; 99(466):314–324.
40. Salmond C, Crampton P, Sutton F. NZDep91: A New Zealand index of deprivation. Australian and New Zealand Journal of Public Health. 1998; 22(7):835–837. [PubMed: 9889455]
41. Centers for Disease Control and Prevention.. [October 22, 2011] HIV Surveillance Report, 2009; vol. 21. Feb. 2011 http:/www.cdc.gov/hiv/topics/surveillance/resources/reports/.
42. Florida Department of Health. [April 03, 2012] Acquired immune deficiency syndrome/human immunodeficiency virus. The Florida Division of Disease Control Surveillance 2009 Report. http://www.doh.state.fl.us/disease_ctrl/aids/trends/epiprof/AIDS_09.pdf.
43. Census 2000 ZCTAs™. [October 22, 2011] ZIP Code Tabulation Areas. Technical Documentation; 2000. http://www.census.gov/geo/ZCTA/zcta_tech_doc.pdf.
44. RUCA. [October 22, 2011] Rural-urban categorization. WWAMI Rural Health Research Center; 2005. http://depts.washington.edu/uwruca/ruca-uses.php.
45. Hart LG, Larsen EH, Lishner DM. Rural definitions for health policy and research. Am J Public Health. 2005; 95(7):1149–1155. [PubMed: 15983270]
46. Washington State Department of Health. [April 03, 2012] Guidelines for using rural-urban classification systems for public health assessment; 2010. http://www.doh.wa.gov/Guidelines/RuralUrban.htm.
47. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951; 16(3):297–334.
48. Bland JM, Altman DG. Statistics notes: Cronbach's alpha. BMJ. 1997; 314:572. [PubMed: 9055718]
49. Stevens, JP. 4th Ed.. Lawrence Erlbaum Associates Publishers; New Jersey: 2002. Applied Multivariate Statistics for the Social Sciences.
50. Hair, JF., Jr.; Black, WC.; Babin, BJ., et al. Multivariate Data Analysis. 6th Ed.. Pearson Prentice Hall Publications; New Jersey: 2006.
51. Velicer WF, Jackson DN. Component analysis versus common factor analysis: some issues in selecting an appropriate procedure. Multivariate Behavioral Research. 1990; 25:1–28.
52. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. Psychol Methods. 1999; 4(3):272–299.

53. Kline, RB. Principles and practices of structural equation modeling. 2nd Ed.. The Guilford Press; New York: 2005.

54. Coxe S, West S, Aiken L. The analysis of count data: a gentle introduction to poisson regression and its alternatives. J Pers Assess. 2009; 91(2):121–136. [PubMed: 19205933]

55. SPSS. SPSS PASW Statistics 18.0 Brief Guide. SPSS Inc.; Chicago, Illinois: 2009.

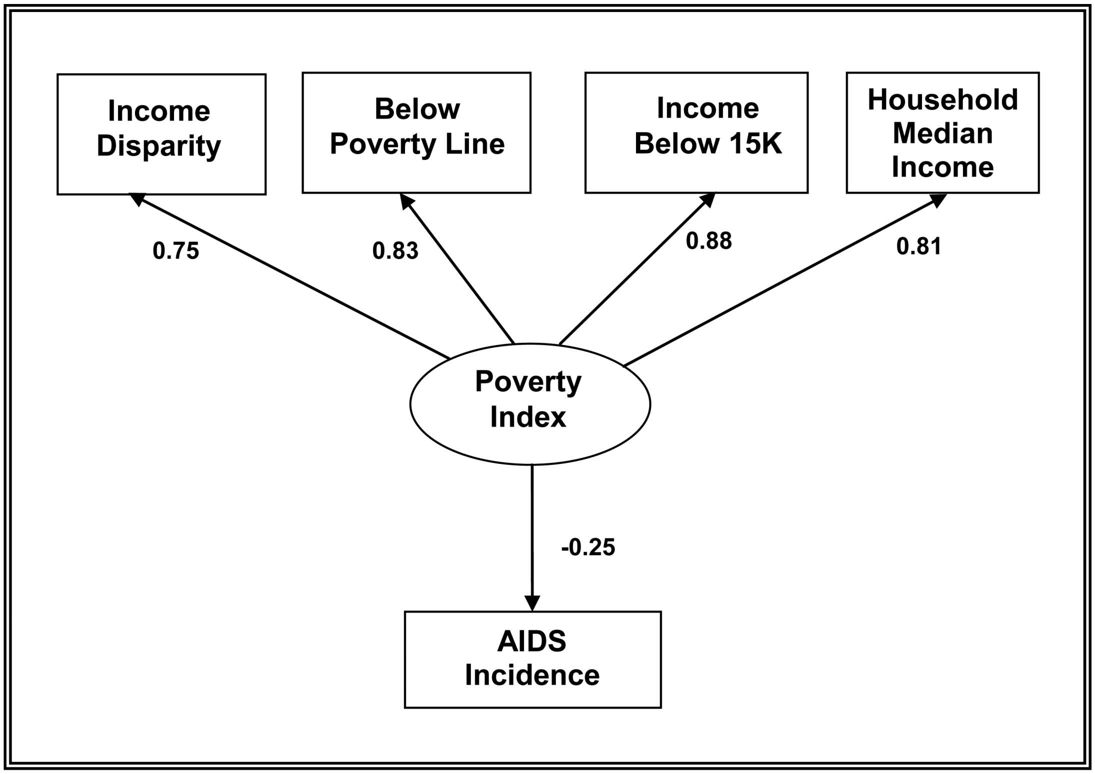
56. SAS Institute Inc.. SAS 9.1.3. 2000-2004. Help and Documentation. SAS Institute Inc.; Cary, NC: 2004.

57. Arbuckle, JL. AMOS 18 User's Guide. SPSS Inc.; Chicago, Illinois: 2009.

58. Berk ML, Schur CL, Dunbar JL, Bozzette S, Shapiro M. Short report: migration among persons living with HIV. Soc Sci Med. 2003; 57(6):1091–1097. [PubMed: 12878108]

59. Yurdugul H. Minimum sample size for Cronbach's coefficient alpha: a Monte-Carlo study. H U Journal of Education. 2008; 35:397–405. http://www.scribd.com/doc/52309780/Minimum-Sample-Size-For-Cronbach-Coefficient-Alpha.

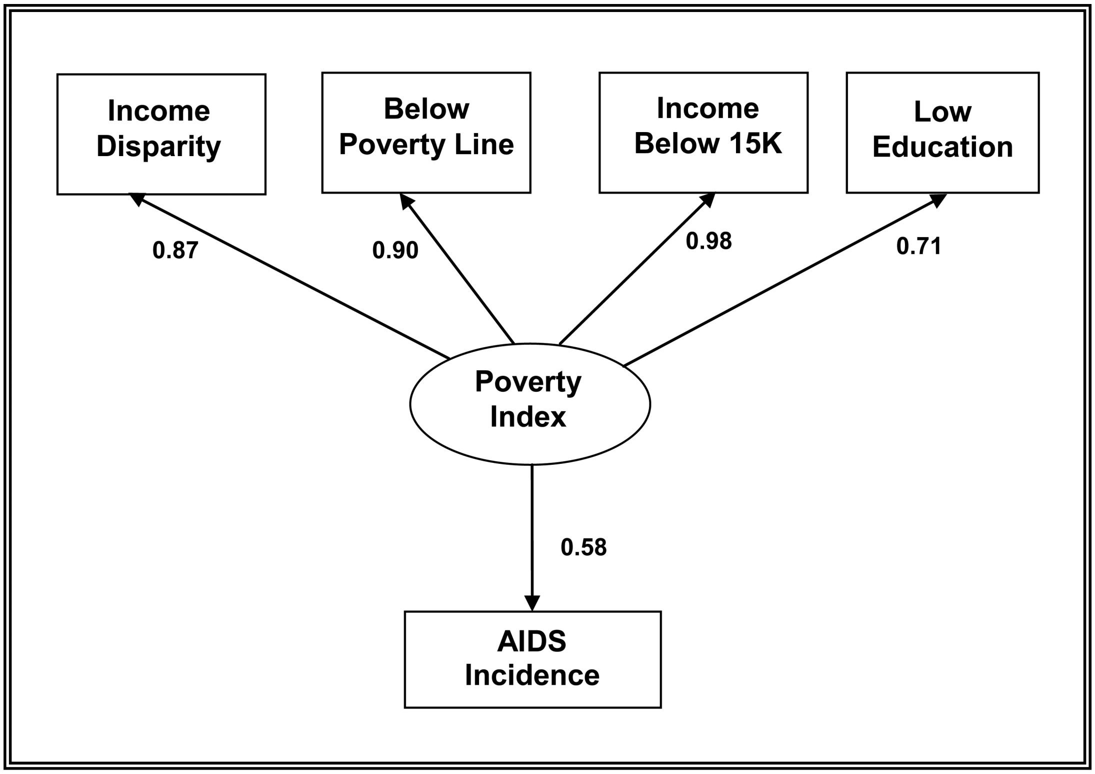
60. Westland JC. Lower bounds on sample size in structural equation modeling. Electron Commer R A. 2010; 9(6):476–487.




Notes: Model fit indices: CR = 0.89; GFI = 0.992; CFI = 1.000; $R^2$ (%) = 6.4% (R-Square).

- CR: Construct reliability; GFI: Goodness-of-fit index; CFI: Comparative fit index; $R^2$ (or R-Square): % of variability that is explained.
- Path coefficients are standardized coefficients.

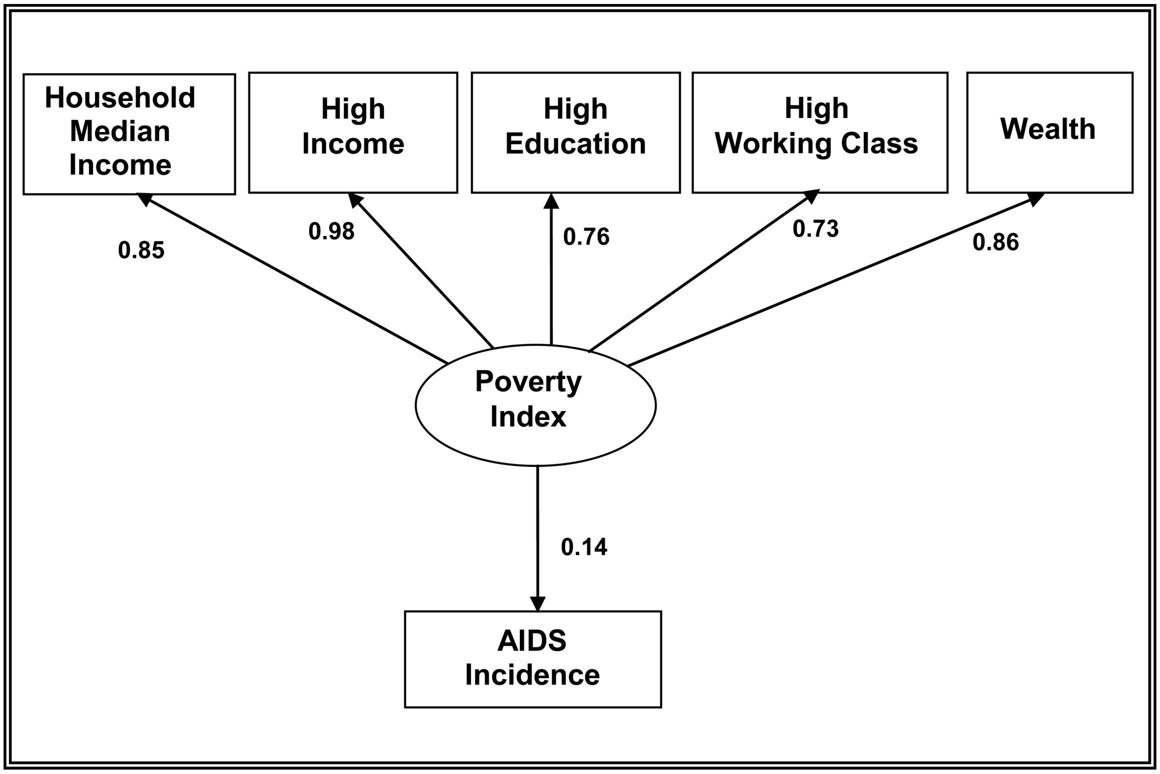
**Figure 1.**
Poverty index as a predictor of AIDS incidence in rural areas




Notes: Model fit indices: CR = 0.924; GFI = 0.928; CFI = 0.958; $R^2$ (%) = 34.1% (R-Square)

- CR: Construct reliability; GFI: Goodness-of-fit index; CFI: Comparative fit index; $R^2$ (or R-Square): % of variability that is explained.
- Path coefficients are standardized coefficients.

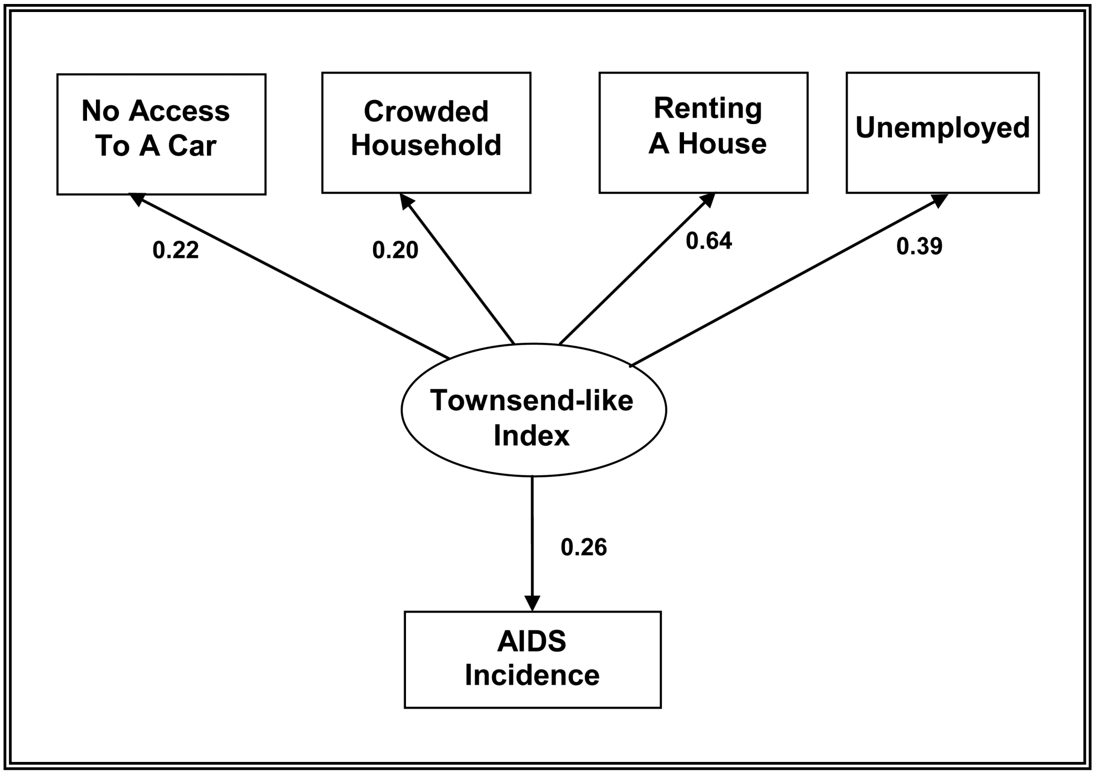
**Figure 2.**
Poverty index as a predictor of AIDS incidence in urban areas




Notes: Model fit indices: CR = 0.923; GFI = 0.750; CFI = 0.773; $R^2$ (%) = 2.0% (R-Square).

- CR: Construct reliability; GFI: Goodness-of-fit index; CFI: Comparative fit index; $R^2$ (or R-Square): % of variability that is explained.
- Path coefficients are standardized coefficients.

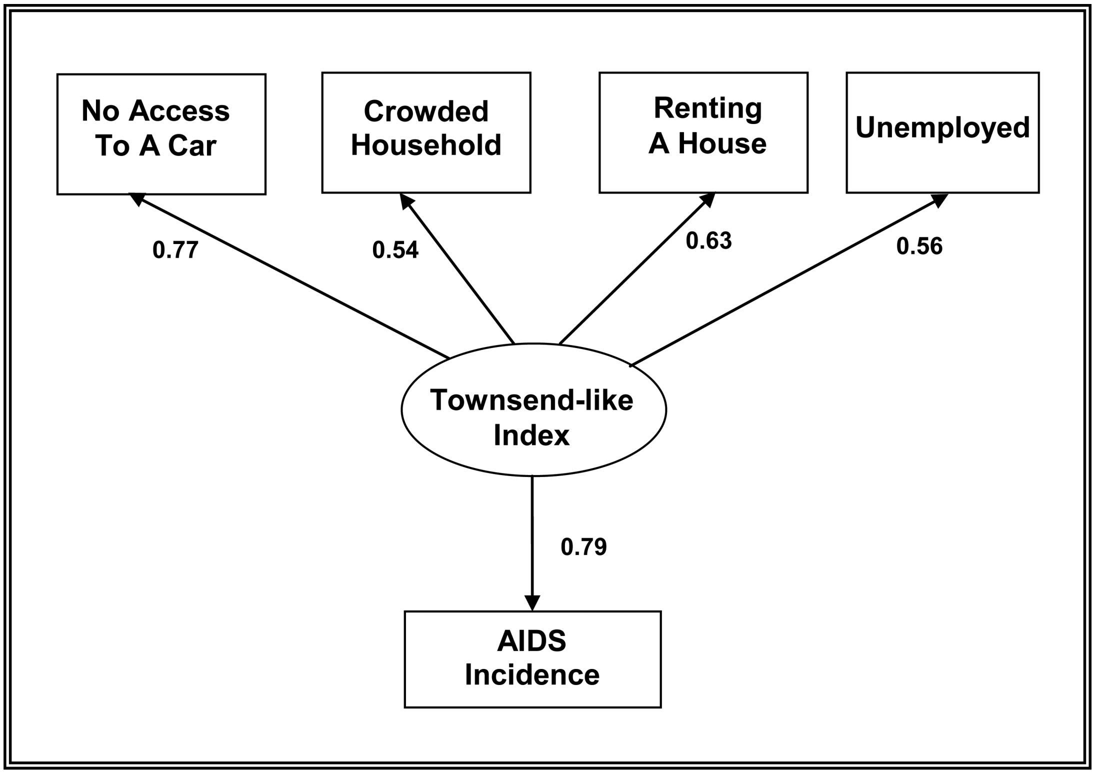
**Figure 3.**
Affluence index as a predictor of AIDS incidence in urban areas




Notes: Model fit indices: CR = 0.386; GFI = 0.964; CFI = 0.749; $R^2$ (%) = 25.8% (R-Square).

- CR: Construct reliability; GFI: Goodness-of-fit index; CFI: Comparative fit index; $R^2$ (or R-Square): % of variability that is explained.
- Path coefficients are standardized coefficients.

**Figure 4.**
Adapted Townsend index as a predictor of AIDS incidence in rural areas




Notes: Model fit indices: CR = 0.721; GFI = 0.939; CFI = 0.893; $R^2$ (%) = 62.7% (R-Square).

- CR: Construct reliability; GFI: Goodness-of-fit index; CFI: Comparative fit index; $R^2$ (or R-Square): % of variability that is explained.
- Path coefficients are standardized coefficients.

**Figure 5.**
Adapted Townsend index as a predictor of AIDS incidence in urban areas

NIH-PA Author Manuscript
NIH-PA Author Manuscript

**Table I**

Area-based (ZIP code tabulation areas) socio-economic (SES) measures that were used

| **SES Indicators**[a] | **Brief Description** |
|---|---|
| No access to a car | % of households without access to a car |
| Crowding | % of households with ≥ 1 person per room |
| Poverty status | % of people living below the federally-defined poverty line |
| Wealth[b] | % of owner-occupied home worth ≥ $300,000 |
| Median income[b] | Median household income in 1999 |
| Low income | % of households with annual income < $15,000 |
| High income[b] | % of households with annual income ≥ $150,000 |
| Income disparity[c] | Ratio of low income to high income |
| Low education | % of population aged 25 and over with less than 12th grade education |
| High education[b] | % of population aged 25 and over with a graduate or professional degree |
| Renting a house | % of households living in rented house |
| Unemployment | % of population age 16 and over unemployed |
| High class work[b] | % of population aged 16 and over employed in high working class occupations. |

***Notes***:

[a]Indicators were standardized (Z-scores)

[b]variable was reversed to align all the items scoring in the same direction.

[c]Income disparity was defined as the logarithmic of 100*ratio the % of households with annual income less than $10,000 to the % of households with annual income of at least $50,000.

**Table II**

Reliability of socioeconomic status scales: Cronbach's Alpha coefficients (k is the number of variables)

| | Rural ZCTA | Urban ZCTA |
|---|---|---|
| All SES indicators | 0.781 (k = 13) | 0.917 (k = 13) |
| Best SES indicators | 0.889 (k = 4) | 0.924 (k = 9) |
| Townsend indicators * | 0.403 (k = 4) | 0.728 (k = 4) |

| ***Townsend indicators*** **(k = 4)** | ***Selected indicators in rural areas*** **(k = 4)** | ***Selected indicators in urban areas*** **(k = 9)** |
|---|---|---|
| No access to a car | Income disparity | Income disparity |
| Crowding | Poverty status | Poverty status |
| Renting a house | Low income | Low income |
| Unemployment | Household median income | Low education |
| | | Household median income |
| | | High income |
| | | High education |
| | | High class work |
| | | Wealth |

***Notes***: The unselected SES indicators included: no access to a car, crowding, renting a house, and unemployment. They constitute the Townsend index. k is the number of variables.

## Table III

AIDS and SES variables: correlation coefficient and incidence rate ratio

| Variables | Rural ZCTA | | Urban ZCTA | |
|---|---|---|---|---|
| | CC[(a)] | IRR[(b)] | CC[(a)] | IRR[(b)] |
| Income disparity | −0.250 | 0.463 | 0.427 | 6.806 |
| Poverty status | −0.050 (NS) | 0.927 (NS) | 0.578 | 8.247 |
| Low income | −0.180 (NS) | 0.676 (NS) | 0.530 | 7.955 |
| Low education | 0.121 (NS) | 0.771 (NS) | 0.434 | 6.672 |
| Median income | −0.126 (NS) | 0.512 (NS) | 0.416 | 6.136 |
| High income | 0.008 (NS) | 0.606 (NS) | 0.176 | 2.990 |
| High education | −0.059 (NS) | 0.767 (NS) | 0.143 | 2.938 |
| High class work | −0.096 (NS) | 0.726 (NS) | 0.253 | 3.463 |
| Wealth | −0.015 (NS) | 0.561 (NS) | 0.093 | 1.710 |
| No access to a car | 0.203 | 2.113 | 0.621 | 8.385 |
| Crowding | 0.226 | 3.966 | 0.422 | 5.046 |
| Renting a house | 0.269 | 3.504 | 0.549 | 6.779 |
| Unemployment | 0.061 (NS) | 0.580 (NS) | 0.494 | 6.384 |
| Poverty index | −0.253 | 0.694 | 0.584 | 8.558 |
| Townsend-like index | 0.508 | 3.573 | 0.792 | 8.276 |
| Affluence index | ------- | ------- | 0.140 | 3.732 |

***Notes***:

The CC assumes linear relationships; for the IRR, quartiles are used and reported IRR values compare the highest quartile to the lowest (the most deprived to the least deprived group). *NS* indicates statistically non-significant results.

*(a)*CC is the correlation coefficient

*(b)*IRR is the incidence rate ratio.