FHEA Data Documentation

DRAFT

2013

Overview

With HUD's Office of Sustainable Housing and Communities asking its grant recipients to complete a Fair Housing Equity Assessment (FHEA), the agency is taking a more active role as a dynamic partner by providing more data and analytical tools to help quantify and interpret particular fair housing dynamics. This document outlines the data, methods, and sources behind the data that HUD is providing. HUD's Office of Policy Development & Research (PD&R) has compiled a set of neighborhood data and analysis that will be available to program participants to support local planning efforts. This document describes the data and analysis which accompanies three central equity principles: reducing segregation, eliminating racially/ethnically concentrated areas of poverty, increasing access to areas of high opportunity. This data package is by no means exhaustive, use of the data is optional and should not supplant more robust local data or knowledge. It represents a baseline effort to assemble consistent, nationally available data from a variety of sources in a single location and provide examples of possible analytical strategies to examine racially-concentrated areas of poverty, segregation and integration, and access to neighborhood opportunity.

1 Geographic Notes

Core data on race, ethnicity, and poverty is provided at two geographic levels: block groups from the Census 2010, and census tracts from the American Community Survey 2006-2010 estimates. Where census 2010 data is available it is used in place of survey data to allay concerns about sampling error. Data that incorporates economic cross-tabulations is from the American Community Survey 2006-2010. It is not provided below the census tract level due to concerns about sampling error.

2 Defining Racially/Ethnically-Concentrated Areas of Poverty

To assist communities in identifying racially/ethnically-concentrated areas of poverty (RCAPs/ECAPs), HUD PD&R has developed a census tract based definition for RCAP/ECAPs. The definition involves a racial/ethnic concentration threshold and a poverty test. The racial/ethnic concentration threshold is straightforward: RCAP/ECAPs must have a non-white population of 50 percent or more. Regarding the poverty threshold, Wilson (1980) defines neighborhoods of "extreme poverty" as census tracts with 40 percent or more of individuals living at or below the poverty line. Because overall poverty levels are much lower in many parts of the country, we supplement this with an alternate criterion. Thus, a neighborhood can be an RCAP/ECAP if it has a poverty rate that exceeds 40% or is three times the average tract poverty rate for the metro/micro area, whichever threshold is lower. Census tracts with this extreme poverty that satisfy the racial/ethnic concentration threshold are deemed RCAPs/ECAPs.

$$RCAP_i = yes \ldots if \ldots \begin{cases} PovRate_i \geqq \left[3 * \mu^{cbsa}_{PovRate}\right] \\ or \\ PovRate_i \geqq 0.4 \end{cases} \bigcap \left[\frac{(1-NHW_i)}{Pop_i}\right] \geqq .50$$

Where i indexes census tracts, $(\mu^{cbsa}_{povrate})$ is the metro (CBSA) mean tract poverty rate, PovRate is the ith tract poverty rate, (NHW_i) is the non-hispanic white population in tract i, and Pop is the population in tract i.

While this definition works well for tracts in metropolitan or micropolitan areas, places outside of these geographies are unlikely to have racial/ethnic concentrations as high as 50 percent. In these areas, we consequently set the racial/ethnic concentration threshold at 20 percent in applying the RCAP/ECAP definition.

3 Analyzing Segregation

To assist program participants in describing the level of residential segregation in their geography, HUD PD&R is constructing several common social science indices that measure segregation. These metrics will allow program participants to identify whether their area features high, moderate or low levels of segregation. The measures described below will be coupled with geospatial presentations of racial/ethnic patterns over time so that program participants can visualize the evolving patterns in their community.

3.1 Dissimilarity index

A primary metric for identifying segregation is the dissimilarity index. A dissimilarity index represents a summary measure of the extent to which the distribution of any two groups (frequently racial or ethnic groups) differs across census tracts or block-groups . It is calculated as:

$$D^{WB}_j = \frac{1}{2}\sum_{i=1}^{N}\left|\frac{W_i}{W_j} - \frac{B_i}{B_j}\right|$$

Where i indexes census block-groups, j is the jth jurisdiction, W is group one and B is group two, and N is the number of block-groups i in jurisdiction j. By construction, the index is bound between zero and one. A value of zero implies "perfect" integration, achieved if every census tract or block-group mirrors the two groups shares in the overall geography. A dissimilarity index of 1 reflects complete segregation, where each tract has exclusively one of the two groups.

PD&R will provide dissimilarity indices at the jurisdiction-level, for jurisdictions of similar size in the same census region, and for metropolitan/micropolitan level. At each level, PD&R calculates the index between the relevant racial/ethnic groups and the majority group, disabled populations by disability type relative to non-disabled persons, and the largest immigrant populations and the majority group in a jurisdiction. In addition, to help communities understand how their situation compares with others around the country, PD&R reports will provide a color-coding designation that will signify whether HUD interprets the value as high, moderate, or low. Table 1 is a tentative statistical designation for the dissimilarity index based on an examination of the literature and an inspection of the statistical distributions of these values across these categories.[1]

Table 1

Measure	Values	Description
Dissimilarity Index	< 0.40	Low Segregation
[min: 0, max: 1]	0.41-0.54	Moderate Segregation
	> 0.55	High Segregation

[1] See Massey and Denton (1993) or Glaeser and Vigdor (1999) for discussion of dissimilarity index values. HUD also examined the various statistical distributions of dissimilarity values across communities. For example, a dissimilarity index of 0.55 represented the 97th percentile of non-white/white segregation for CPD program participant geographies with sufficiently large (>10%) non-white populations using the ACS 05/09 estimates.

3.2 Isolation Indices

Another common approach to measuring segregation is the isolation index, which compares a group's share of the overall population in a jurisdiction to the average neighborhood share for members of that group. For example, suppose a jurisdiction is 20 percent Hispanic/Latino overall, but the average Hispanic/Latino resident of that jurisdiction lives in a neighborhood that is 60 percent Hispanic/Latino - the isolation index for Hispanics in this jurisdiction would take the value 0.4 (0.6-0.2). Similar to the dissimilarity index, the higher the value, the more segregated a community is. The isolation index cannot exceed 1 (or 100, depending on the scaling). Generalizing, for any group (M) in jurisdiction (j) the isolation index is calculated as follows:

$$Iso_M = \left[\sum_{i}^{N} \frac{M_i}{Mj} * \left(\frac{M_i}{T_i} \right) \right] - \frac{M_j}{T_j}$$

Where (i) indexes block-groups and (T) is the total population in block-group (i) or jurisdiction (j).

The isolation index is highly correlated with the dissimilarity index, and conceptually very similar, but it tends to provide a better characterization of residential segregation when minority populations are extremely small.

3.3 Predicted Racial/Ethnic Composition Ratio

For very small communities, there are generally too few census block-groups or minorities for statistical metrics such as a dissimilarity index or even the isolation index to be particularly informative. Instead, for these communities, PD&R calculates a predicted value for the racial/ethnic minority share for a jurisdiction and compares this to the actual composition. Predicted values are based on a metropolitan/micropolitan area's income distribution by race and ethnicity. For a jurisdiction, the metro-level racial share for each income category is multiplied by the number of households the jurisdiction has in that category. The totals are summed to determine the predicted number of minorities in a jurisdiction. This total is then compared with the actual number of minorities in a community by calculating a ratio of actual to predicted. For any jurisdiction j, the predicted total for subgroup M is defined as $\widehat{M}$, it is the number of households (H) in household income category l in jurisdiction j, multiplied by the metropolitan area (k) share of subgroup M in household income category l, summed across all income categories l to N:

$$\widehat{M_j} = \sum_{l}^{N} H_{lj} * \frac{M_{lk}}{H_{lk}}$$

Ratios near 1 indicate that the jurisdiction is close to its predicted level of minority composition. Those far less than 1 show that the jurisdiction has many fewer minorities than one might expect given income levels. Table 2 presents a stylized example of this procedure for two hypothetical jurisdictions in the same metropolitan area. As is clear, jurisdiction A has a non-white population far below what might be expected. The non-white population in jurisdiction B is close to what one might expect. Table 3 characterizes the value ranges of the measure.

Table 2

Panel A. Hypothetical Jurisdiction A					
Income category	Metro area racial share	Total Jurisdiction Population	Predicted racial pop.	Actual racial pop.	Actual/ Predicted Ratio
Less than $50,000	0.32	10,000	3,200	1,300	
$50,000 to $100,000	0.34	6,000	2,040	500	
Greater than $100,000	0.28	2,000	560	200	
Total			5,800	2,000	0.3448

Panel B. Hypothetical Jurisdiction B					
Income category	Metro area racial share	Total Jurisdiction Population	Predicted racial pop.	Actual racial pop.	Actual/ Predicted Ratio
Less than $50,000	0.32	10,000	3,200	3,400	
$50,000 to $100,000	0.34	6,000	2,040	2,280	
Greater than $100,000	0.28	2,000	560	400	
Total			5,800	6,080	1.0483

Note: This table is illustrative, the income categories presented here are notional.

Table 3

Measure	**Values**	**Description**
Predicted Racial/Ethnic Composition	0.0-50%	Non-White Share Extremely Below Predicted
(Across Jurisdiction, Non-Economic, Racial/Ethnic Segregation)	50-70%	Non-White Share Moderately Below Predicted
(ratio of predicted non-white share over actual non-white share)	70-90%	Non-White Share Slightly Below Predicted
	90-110%	Non-White Share Approximates Predicted
	110%+	Non-White Share Above Predicted

4 Analyzing Neighborhood Opportunity Indicators

HUD has developed a two-stage process for analyzing disparities in access to neighborhood opportunity. The first stage involves quantifying the degree to which a neighborhood offers features commonly associated with opportunity. This stage uses metrics that rank each neighborhood along a set of key dimensions. In the second stage, HUD combines these dimension rankings with data on where people in particular subgroups live to develop a measure of that group's general access or exposure to each opportunity dimension. These summary measures can then be compared across subgroups to characterize disparities in access to opportunity.

HUD considers "opportunity" a multi-dimensional notion. To focus the analysis, HUD developed methods to quantify a selected number of the important "stressors" and "assets" in every neighborhood. These dimensions were selected because existing research suggests they have a bearing on a range of individual outcomes. In particular, HUD has selected six dimensions upon which to focus:

- Neighborhood School Proficiency
- Poverty

- Labor Market Engagement
- Job Accessibility
- Health Hazards Exposure
- Transit Access

Invariably, these dimensions do not capture everything that is important to the well-being of individuals and families. In quantifying indicators of neighborhood opportunity, HUD is not making a definitive assessment of one's life chances based on geography. HUD is quantifying features of neighborhoods for the purpose of assessing whether significant disparities exist in the spatial access or exposure of particular groups to these quality of life factors.

While these important dimensions capture a number of key concepts identified by research as important to quality of life, the measures are not without limitations. PD&R constrained the scope of HUD-provided items to those that are closely linked to neighborhood geographies and could be measured consistently at small area levels across the country. For example, HUD's measure of school performance only reflects elementary school proficiency. It does not capture academic achievement for higher-grades of schooling, which are important to a community's well-being, but likely less geographically-tied to individual neighborhoods than elementary schools. Similarly, the health hazard exposure measure only captures outdoor toxins, missing indoor exposures. The national-availability restriction is a necessity given that all HUD program participants must complete an Assessment of Fair Housing. HUD realizes that there are other assets and stressors that are relevant for opportunity, such as neighborhood crime or housing unit lead and radon levels. However, these lack consistent neighborhood-level data across all program participant geographies. As a consequence, HUD encourages program participants to supplement the data it provides with robust locally-available data on these other assets and stressors, so that the analysis is as all-encompassing as possible. Each dimension is described below.

4.1 Neighborhood School Proficiency Index

The neighborhood school proficiency index uses school-level data on the performance of students on state exams to describe which neighborhoods have high-performing elementary schools and which have lower performing elementary schools. The proficiency index is a function of the percent of elementary school students proficient in read (r) and math (m) on state test scores for the ith school associated with the neighborhood ($i = 1, 2, ..n$)where N is the maximum number of schools in any block-group in the state-distribution, and school enrollment s:

$$School_i = \sum_{i}^{N} \frac{s_i}{\sum^n s_i} * \frac{1}{2} * r_i + \frac{1}{2} * m_i$$

Elementary schools are linked with block-groups based on a geographic mapping of attendance area zones from School Attendance Boundary Information System (SABINS), where available, or within-district proximity matches of up to the four-closest schools within a mile. In cases with multiple school matches, an enrollment-weighted score is calculated following the equation above.

4.2 Poverty Index

HUD created a simple poverty index to capture the depth and intensity of poverty in a given neighborhood. The index uses family poverty rates and public assistance receipt[2] to operationalize both aspects. The index is a linear combination of two vectors: the family poverty rate (pv) and the percentage of households receiving public assistance (pa). Where means (μ_{pv}, μ_{pa}) & standard errors (σ_{pv}, σ_{pa})are estimated over the metropolitan area distribution or balance of state in non-metros.

$$Pov_i = (\frac{pv_i - \mu_{pv}}{\sigma_{pv}}) - 1 + (\frac{pa_i - \mu_{pa}}{\sigma_{pa}}) * -1$$

[2] Public assistance is cash-welfare, such as Temporary Assistance for Needy Families (TANF).

4.3 Job Access Model

The job access index summarizes the accessibility of a given residential neighborhood as a function of its distance to all job locations, with distance to larger employment centers weighted more heavily. Specifically, a gravity model is used, where the accessibility (A_i) of a given residential block-group is a summary description of the distance to all job locations, with the distance from any single job location positively weighted by the size of employment (job opportunities) at that location and inversely weighted by the labor supply (competition) to that location. More formally, the model has the following specification:

$$A_i = {}_{i=1}^{n} \frac{E_j d_{ij}^{-\beta}}{\sum L_j}$$

Where i indexes residential locations and j indexes job locations, and distance, d, is measured as "as the crow flies" or by commute time (depending on availability) between block-groups i and j. E represents the number of jobs in tract j and L is the number of workers. The term β is a distance friction, it characterizes how rapidly a job opportunity should be "discounted" as distance increases. This discount factor is estimated parametrically by modeling the observed commute patterns in a region. The contribution of distance in the commute model below is the coefficient β on the variable distance (d):

$$C_{ij} = \alpha L_i^{\delta} E_j^{\vartheta} d_{ij}^{\beta}$$

Where again i indexes residential locations and j indexes job locations, C is the observed number of commuters for block-group pairing ij.

4.4 Labor Market Engagement Index

The labor market engagement index provides a summary description of the relative intensity of labor market engagement and human capital in a neighborhood. This is based upon the level of employment, labor force participation and educational attainment in that neighborhood. Formally, the labor market engagement index is a linear combination of three standardized vectors: unemployment rate (u), labor-force participation rate (l), and percent with bachelor's or higher (b), using the following formula

$$LBM_i = \quad (\frac{u_i - \mu_u}{\sigma_u}) * -1 \quad + (\frac{l_i - \mu_l}{\sigma_l}) + (\frac{b_i - \mu_b}{\sigma_b})$$

Where means (μ_u,μ_l, μ_b) and standard errors(σ_u,σ_l, σ_b) are estimated over the metropolitan area distribution or balance of state in non-metros.

4.5 Environmental Health Hazard Exposure Index

HUD constructed a health hazards exposure index to summarize potential exposure to harmful toxins at a neighborhood level.[3] Potential health hazards exposure is modeled in a given block-group as a function of the volume of toxic industrial releases from the EPA's Toxic Release Inventory (R), the EPA toxicity assessment of the release chemicals (X), and the distance to the toxic release facility (d).[4] Again i indexes residential locations, and f indexes facilities.

$$HazExp_i = -1 * \qquad R_f X / d_{if}^2$$

4.6 Transit Access

HUD has constructed a transit access index where available data exists to support local analysis. HUD uses data on over 200 transit agencies that provide data through GTFS Exchange (http://www.gtfs-data-exchange.com/) to assess relative accessibility within metro areas (or balance of state). The appendix contains a list of metropolitan areas where GTFS data was available and used. The GTFS-based accessibility

[3]HUD anticipates deriving detailed health hazard exposure data from EPA's to-be-released C-FERST tool when available.
[4]See www.epa.gov/tri to learn more about the Toxic Release Inventory program.

index is designed to model relative accessibility to amenities via bus or trains within a metro. Because standardized data on the location of amenities is not uniformly available at a granular level, HUD uses the number of jobs in retail (NAICS 44-45), arts entertainment & recreation (NAICS 71), and food & accommodations (NAICS 72) as proxies for the magnitude of amenities at the block-group level from the Local Employment Dynamics dataset published by the census bureau[5]. First, HUD identified the number of jobs in these sectors within 1/2 mile of each bus stop and 3/4 mile of each rail transit stop and summed them. Then for each trip in the transit system, HUD calculated a stop-specific measure of the additional amenities accessed in each ensuing stop on that route, which it then divided by (deflated) the additional travel time to each ensuing stop. Mathematically, this can be expressed in several terms.

Let (s_{ij}) represent the accessibility of stop i on trip j, a is the amenity radius of a stop (the total jobs mentioned above), and T is the marginal travel time with each stop. Each stop of each trip takes on a value equal to the sum of the amenity radius of each ensuing stop divided be the time to that next stop for all stops on a trip.

$$s_{ij} = \sum_{i}^{N} \frac{a_{i+1}}{T_{t+1}}$$

These stop-journey specific (s_{ij}) values are then summed over all journeys j (where a journeys in opposite direction are counted as two trips) made in 24-hours to create a single aggregate accessibility value for each stop in the system (where k is the total stops in the system).

$$A_i = \sum_{j}^{k} s_{ij}$$

To translate these stop accessibility values (A_i) to block-groups, HUD then calculates the distance between each stop and the population-weighted centroid of each block-group. The three highest accessibility stops within 3/4 of a mile are summed to generate a block-group value for accessibility. Finally, these values are placed into deciles (10-percentile) buckets within-metro or balance of state, and are scaled up by a factor of 10 to align with the other indices. Block-groups that are not within 3/4 of a mile of either a bus or transit stop are normalized to a value of 1 – the lowest accessibility score. For communities with fixed rail, but no available GTFS data, HUD calculates a simple access measure as the distance of the block-group centroid to the nearest fixed-rail.

[5]For states without Work Area Characteristics files in the LED data, population was used as a proxy

4.7 Sources

Table 3 below details the sources of each data point.

Table 3

Opportunity Dimensions	**Input Variables**	**Source**
Poverty Index	Family Poverty Rate	ACS 2006-2010
	Pct. Households Receiving Public Assistance	ACS 2006-2010
Neighborhood School Proficiency Index	School Math Proficiency / State Math Proficiency	Dept. of Education
	School Reading Proficiency / State Reading Proficiency	Dept. of Education
Labor Market Engagement Index	Unemployment Rate	ACS 2006-2010
	Labor force Participation Rate	ACS 2006-2010
	Pct. with a Bachelor's or higher	ACS 2006-2010
Job Access Index	block-group-level Job Counts	LED, 2010
	block-group-level Job Worker Counts	LED, 2010
	Origin-Destination Flows	LED, 2010
	Distance	GIS-Derived
Transit Access Index	Distance to stops and accessibility of stops (defined above)	GTFS
	Distance to Nearest Rail or BRT Station	DOT
Health Hazards Exposure Index	TRI Facilities, Releases	EPA, 2009
	RSEI Toxicity Assessment	EPA, 2007
	Distance	GIS-Derived

4.8 Access or Exposure to Opportunity Dimensions

To identify disparities in opportunity, HUD PD&R calculates exposure indices for each opportunity dimension across a range of subgroups. The exposure index calculates a weighted average for a given characteristic. The generic access for subgroup M to opportunity dimension R in city j is calculated as:

$$Exp_M^R = \sum_i^N \frac{M_i}{Mj} * R_i$$

Where i indexes block-groups in city j for subgroup M to opportunity dimension R. Again, N is the total number of block-groups in city j. The raw values for the opportunity dimensions are placed into 100 percentile buckets, based on the within-metro (or non-metro balance of state) ranking.[6] For each dimension, the higher the percentile, the more favorable the neighborhood condition along that given dimension.

It is useful to provide an example of this in practice (Table 4). Consider a hypothetical jurisdiction with three neighborhoods. Given the poverty dimension values and population distributions as shown (and abstracting away from the deciles issue for the moment), one can calculated the total group score for both white and Hispanic children using the exposure index formula. The results indicate that there is a disparity between white children and Hispanic children with respect to poverty, with Hispanic children on average exposed to higher poverty levels.

[6] There are a few exceptions, with some percentiles calculated over the state distribution (school proficiency) and national distribution (healh hazards exposure). Also for metro areas with fewer than 100 blockgroups, the blockgroups were placed into decile rankings, which were then scaled up by a factor of ten (10).

Table 4							
	Dimension	**White Children**			**Hispanic Children**		
Neighborhoods	Poverty Index Value	White Children Pop	Share	Exposure Index [(1)*(3)]	Hispanic Chil-dren Pop	Share	Exposure Index [(1)*(6)]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A	80	400	0.4	32	100	0.2	16
B	50	400	0.4	20	150	0.3	15
C	20	200	0.2	4	250	0.5	10
Total		1000	1	**56**	500	1	**41**

Using these exposure indices, it's possible to compare the access or exposure to the above opportunity dimensions across protected classes and identify disparities. Column 4 illustrates the exposure index calculation for white children, with a summary value of 56. The corresponding value for Hispanic children is 41 Disparity values are calculated as the simple difference in average exposure to a given opportunity dimension across two groups. In the example above, the disparity between white and Hispanic children in exposure to poverty is -15 (41-56=-15). In this example, neighborhood A has the lowest-poverty and thus the highest ranking (80th percentile) among the three neighborhoods.

To account for differences in household income across groups, PD&R also provides these exposure indices across protected classes for persons in poverty. This assists jurisdictions in understanding whether there are differences in exposure to opportunity across groups that cannot be explained by differences in income.

PD&R provides these exposure calculations for each non-white group (overall and in poverty) and the disparity relative to the white population (overall and in poverty). PD&R also calculates basic significance tests (at the 0.05 significance level) to identify whether these disparities are statistically discernible from random noise.

References

[1] Glaeser, Edward, David Cutler, and Jacob Vigdor. 1999. "The Rise and Decline of the American Ghetto." *Journal of Political Economy* 107(3): 455-506.
[2] Massey, Douglas and Nancy A. Denton. 1993. American Apartheid: Segregation and the Making of the Underclass. Cambridge: Harvard University Press.

Poverty

[3] Aaronson, Aaron. 1998. "Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes." *The Journal of Human Resources*. Vol. 33, No. 4, pp. 915-946.
[4] Jacob, Brian A., Jens Ludwig and Douglas L. Miller. 2011. "The Effects of Housing and Neighborhood Conditions on Child Mortality." NBER Working Paper #17369.
[5] Jens Ludwig, Lisa Sanbonmatsu, Lisa Gennetian, Emma Adam, Greg J. Duncan, Lawrence F. Katz., Ronald C. Kessler, Jeffrey R. Kling, Stacy Tessler Lindau, Robert C. Whitaker, and Thomas W. McDade, Ph.D. 2011. "Neighborhoods, Obesity, and Diabetes — A Randomized Social Experiment." *New England Journal of Medicine*; 365:1509-1519.

Schools

[6] Schwartz, Heather. 2010. "Housing Policy is School Policy" Rand Corporation.
[7] Deming, David, Justine Hastings, Tom Kane and Doug Staiger. 2011. "School Choice, School Quality and Academic Achievement" revise and resubmit, *American Economic Revie*w.

Labor Market Engagement

[8] Borjas, George. 1995. "Ethnicity, Neighborhoods, and Human-Capital Externalities," *American Economic Review.*
[9] Ioannides, Yannis. 2003. "Empirical Nonlinearities and Neighborhood Effects in the Intergenerational Transmission of Human Capital," *Applied Economics Letters*

Job Access

[10] Ihlanfeldt, Keith. 2006. "Spatial Mismatch Within Urban Labor Markets," Chapter 24 in A Companion to Urban Economics, edited by Richard Arnott and Dan McMillen, Blackwell Publishing, Inc.
[11] Raphael, Steven. 1998. "The Spatial Mismatch Hypothesis of Black Youth Joblessness: Evidence from the San Francisco Bay Area," *Journal of Urban Economics.*

Transit

[12] Quigley, John, Harry Holzer, and Steven Raphael. 2003. "Public Transit and the Spatial Distribution of Minority Employment: Evidence from a Natural Experiment," *Journal of Policy Analysis and Management*, 22(3): 415-442.

Health Hazards

[13] Moretti, Enrico and Janet Currie and Michael Greenstone. 2011. "Superfund Cleanups and Infant Health," *American Economic Review, Papers and Proceedings*.