# Data Shop

Data Shop, a department of Cityscape, presents short articles or notes on the uses of data in housing and urban research. Through this department, the Office of Policy Development and Research introduces readers to new and overlooked data sources and to improved techniques in using well-known data. The emphasis is on sources and methods that analysts can use in their own work. Researchers often run into knotty data problems involving data interpretation or manipulation that must be solved before a project can proceed, but they seldom get to focus in detail on the solutions to such problems. If you have an idea for an applied, data-centric note of no more than 3,000 words, please send a one-paragraph abstract to david.a.vandenbroucke@hud.gov for consideration.

# Comparing Households in HUD Rental Assistance Programs With PUMS Data

Brent D. Mast

U.S. Department of Housing and Urban Development

The views expressed in this article are those of the author and do not represent the official positions or policies of the Office of Policy Development and Research or the U.S. Department of Housing and Urban Development.

## **Abstract**

In February 2012, the U.S. Department of Housing and Urban Development (HUD) launched a new Public Use Microdata Sample (PUMS) database, which contains household-level data for 5 percent of households in five of HUD's largest rental assistance programs. PUMS includes household characteristics as well as geographic information, enabling researchers to perform analyses not possible using other datasets.

This article describes the PUMS data and sampling methodology, and it compares the PUMS with other data sources. A data analysis example is presented using PUMS data comparing rent burdens across programs, adjusting for income differences using Barskey et al.'s (2002) nonparametric technique. Before accounting for income differences, rent burdens are much higher in the Housing Choice Voucher Program (HCVP) compared with HUD's other rental assistance programs. Results indicate that when comparing households that have similar incomes, rent burdens of HCVP households are even higher compared with those of other assisted households.

# Introduction

To advance the federal government's *Open Government Initiative*, in February 2012, the U.S. Department of Housing and Urban Development (HUD) launched a public database to help the research community better understand the characteristics of households receiving assistance under five of the Department's largest rental programs.

HUD's new Public Use Microdata Sample (PUMS) database<sup>1</sup> (HUD, 2012) includes household-level data for 5 percent of households assisted through five HUD programs: (1) the Housing Choice Voucher Program (HCVP), (2) public housing (PH), (3) Section 8 Project-Based (PBS8) Rental Assistance, (4) the Supportive Housing for the Elderly Section 202 program (S202), and (5) the Supportive Housing for Persons with Disabilities Section 811 program (S811).

HCVP tenants find rental units in the private market, paying at least 30 percent of their own income toward rent and utilities; the remainder of the rent is subsidized using a HUD voucher. PH tenants live in housing owned and operated by public housing agencies. PBS8 tenants live in privately owned and operated developments, where HUD subsidizes some, or all, housing units. S202 subsidizes privately owned and operated supportive housing developments for the elderly, and S811 subsidizes privately owned and operated supportive housing developments for persons with disabilities.

The 2009, 2010, and 2012 PUMS data currently are available, and annual updates are planned. The PUMS includes household characteristics and geographic information.

Using these household-level data, researchers can calculate results and statistical relationships at levels of demographic or geographic detail not available in HUD's tabular reports.

To protect tenant privacy, the PUMS contains no personally identifiable information, such as addresses, to ensure that it will not be possible to identify any individual or household. Some variables that have a high identification risk have been categorized, top-coded, or masked. The sample size is large enough to be representative of states and the nation as a whole, but it is small enough to preserve confidentiality.

The next section describes the PUMS data and sampling methodology. The following section compares the PUMS with other data sources. The section after that provides a data analysis example that compares rent burdens across programs that are adjusted for income differences, using Barskey et al.'s (2002) nonparametric method. The final section presents the author's conclusions.

# **Data Description**

This section describes the PUMS scope, sources, sampling methodology, and variables. More information is available in the data dictionaries on the PUMS HUD USER webpage (HUD, 2012).

<sup>&</sup>lt;sup>1</sup> Users must register before accessing PUMS data (HUD, 2012). Users must provide their names, institutions, and e-mail addresses and certify that they will use the data for legitimate research purposes and will in no way use the data to try to identify households.

## **Programs**

The PUMS contains separate datasets for households assisted by programs administered by HUD's Office of Public and Indian Housing (PIH) and Office of Housing. The PUMS contains data for the two largest PIH programs: PH and the HCVP. The PUMS contains data on three Office of Housing multifamily programs: PBS8, S202, and S811. Exhibit 1 reports 2009 sample sizes, weighted household counts, and percentage frequencies by program and variance estimates. Standard errors reported in this article are Taylor series approximations that account for sample design and have finite population corrections.

The 2009 PUMS contains data on 237,689 households representing an estimated 4.7 million assisted households: 2.1 million in the HCVP, 1.1 million in PH, 1.4 million in PBS8, 113,828 in S202, and 30,574 in S811.

#### Exhibit 1

## **PUMS Program Frequencies**

Program	Sample Size (N)	Weighted Households	Standard Deviation	Weighted Percent	Standard Error
Housing Choice Voucher Program	100,797	2,111,234	0.001	44.433	0.000
Public housing	61,687	1,137,413	0.002	23.938	0.000
Project-based Section 8	70,172	1,358,459	0.001	28.590	0.000
Section 202	3,955	113,828	0.000	2.396	0.000
Section 811	1,078	30,574	0.000	0.643	0.000
Total	237,689	4,751,508	0.003	100	

PUMS = Public Use Microdata Sample.

Source: U.S. Department of Housing and Urban Development Public Use Microdata Sample 2009

#### **Data Sources**

Household characteristics for households in PIH programs were extracted from HUD's Inventory Management System/PIH Information Center (IMS/PIC) data system (HUD, 2013a). Household characteristics for households in Office of Housing programs were extracted from HUD's Tenant Rental Assistance Certification System (TRACS) data system (HUD, 2013b).

Both IMS/PIC and TRACS are transaction based. The PUMS is based on the most recent certification transaction for each household at the end of the calendar year. For most households, the record is at most 18 months old. For PH and HCVP households assisted by housing agencies in the Moving to Work demonstration program (HUD, 2013c), the record is at most 3 years old.

## Sample Design

The PUMS is a 5-percent sample, without replacement, of records for the 50 states, Washington, D.C., and Puerto Rico. The sample is stratified by state and program. The sample contains approximately 5 percent of records for each stratum.

The precise sample size for each stratum was determined by Neyman allocation, a statistical method giving a greater sample size for strata that have more diverse populations (Lohr, 1999). The standard deviation of adjusted household income was used as a proxy for diversity. The sample for each stratum is proportional to households multiplied by the standard deviation of adjusted household income.

The datasets contain a weight, which is the inverse of the sampling probability. The weight is the same for all households in a given stratum. Weighting makes the samples nationally representative.

#### **Variables**

The PUMS contains one variable to identify programs and two variables to identify states: state name and Federal Information Processing Standard, or FIPS, code. A variable also identifies strata.

The PUMS contains seven household-level variables. The household type variable reports household type in six categories. The number of household members is also reported, top-coded at 7. Number of bedrooms is reported, top-coded at 4. Household head sex is reported, along with a variable for household head race and ethnicity.

The 2009 PUMS reports annual household income measured in 13 categories. The first 12 categories report income in \$5,000 increments: \$0 to \$5,000, \$5,001 to \$10,000, ... \$55,001 to \$60,000. The 13th category is for households that have incomes between \$60,001 and \$90,000; income is top-coded at \$90,000.

Eligibility for HUD rental assistance programs is based on adjusted household income. Adjusted income is calculated by subtracting certain expenses from household income. Details of the calculation for PIH programs are reported on HUD form 50058 (HUD, 2013d) and on HUD form 50059 (HUD, 2013e) for Office of Housing programs. Annual adjusted household income is reported in the same categories as annual household income, and is also top-coded at \$90,000.

Rent burden is reported, which is defined as the household's contribution to gross rent (including utility costs) divided by monthly adjusted household income. Rent burden is undefined for households that have \$0 adjusted income. The 2009 PUMS reports rent burden in four categories: 0 to 31 percent, 32 to 39 percent, 40 to 49 percent, and 50 percent or more.

PUMS contains three geographic variables in addition to state: metropolitan status in three categories, an urban/rural indicator, and the poverty rate for the location's census tract in five categories.

# **Confidentiality**

To ensure confidentiality, some PUMS variables were masked for some households. State is suppressed for households having unique combinations of variables in the population. If a household is uniquely identified after suppressing state, one or more additional variables are suppressed. For sampling purposes, households that have a suppressed state variable were treated as a separate state.

# **Comparison of PUMS With Other Data Sources**

In this section, PUMS is compared with data available from other datasets: Picture of Subsidized Households and the American Housing Survey.

#### Picture of Subsidized Households

Picture of Subsidized Households (hereafter referred to as Picture; HUD, 2013f) reports a large amount of tabulated demographic<sup>2</sup> and geographic data on tenants in HUD rental assistance programs. Picture reports data on more programs than PUMS. Compared with PUMS, Picture also provides information at lower levels of geography (Core Based Statistical Area, county, city, and census tract). Picture information is also available by public housing agency for PIH programs and at the project level for projects in some programs.

Not all households have data reported in IMS/PIC and TRACS. Picture contains data for the entire population of assisted households reporting in IMS/PIC and TRACS, along with data on the percentage of housing units for which data are reported. By contrast, PUMS is limited to a 5-percent sample of reporting households.

Picture, however, does not allow for custom cross-tabulations. In the next section, a data analysis example is provided that uses PUMS data, which is not possible using Picture data.

# **American Housing Survey**

Before 2011, the usefulness of the American Housing Survey (AHS; HUD, 2013g) for analyzing HUD-assisted household data was extremely limited due to (1) a very small sample size and (2) reporting errors in both the fact and type of assistance (Casey, 1992; HUD, 2008; Mast, 2012; Rucinski and Athey, 1996; Shroder, 2002). In addition, only two HUD programs were identified: PH and the HCVP.

In 2011, AHS began oversampling HUD-assisted households and verifying assistance status and program type. AHS now identifies households in three rental assistance categories: PH, the HCVP, and other "privately owned subsidized housing" (HUD, 2013h: 390), which can include housing subsidized by a variety of HUD multifamily rental assistance programs administered by the Office of Housing. PUMS and Picture identify specific HUD multifamily programs, but AHS does not.

Exhibit 2 reports AHS 2011-verified assistance status estimates by program. AHS 2011 contains data on 8,987 housing units that have verified assistance, representing an estimated 3.6 percent of total occupied housing units. The 2011 estimated counts of assisted households based on AHS data were 2.1 million HCVP households, 985,225 PH households, and 1.1 million households in multifamily programs.

Compared with PUMS, AHS contains much more demographic information for both people and households. AHS also contains a wealth of data on housing characteristics and neighborhood conditions not available in PUMS. The only geographic areas identified in AHS, however, are select metropolitan areas.

<sup>&</sup>lt;sup>2</sup> PUMS contains one variable—rent burden—that is not contained in Picture. Although Picture reports mean household contribution to gross rent and mean adjusted household income, mean rent burden cannot be computed with these variables due to Jensen's Inequality (the mean of [X/Y] does not equal mean[X]/mean[Y]). Mean household contribution to gross rent divided by mean adjusted household monthly income can be used as an estimator of mean rent burden.

Exhibit 2

American Housing Survey Assisted Household Estimates

Verified Rental Assistance Status	Sample Size (N)	Weighted Housing Units	Standard Deviation	Weighted Percent	Standard Error
Housing Choice Voucher Program	2,562	2,059,291	66,647	1.792	0.058
Public housing	2,303	985,225	44,415	0.857	0.039
Multifamily	4,122	1,127,739	41,004	0.981	0.036
Total subsidized	8987	4,172,255	89,526	3.631	0.077
Not subsidized	42,452	33,900,398	262,877	29.502	0.207
Not applicable (owner-occupied)	83,479	76,834,533	390,164	66.867	0.213
Total	134,918	114,907,187	428,882	100	

Source: American Housing Survey 2011

# **Data Analysis**

This section provides a data analysis example using 2009 PUMS data that is not possible using Picture or AHS data. Rent burdens are compared across households in the HCVP, PH, and PBS8, making adjustments for differences in adjusted household income.

The S202 and S811 households are excluded from the analysis because some income categories in the 2009 PUMS contain no households in these programs, which would complicate the analysis. Suitable methods for comparisons, including these programs, are discussed at the end of this section.

Exhibit 3 reports weighted percentages of households by rent burden category and program, with standard errors. Households in the HCVP tend to have much higher burdens compared with those in PH and PBS8. In 2009, an estimated 31.5 percent of HCVP households had burdens greater than 31 percent. The corresponding estimates for PH and PBS8 are 3.1 and 9.0 percent, respectively. An estimated 13.6 percent of voucher families had burdens of at least 40 percent. By sharp contrast, 2.4 percent of PH households and 7.4 percent of PBS8 households were estimated to have burdens that were 40 percent or more.

High rent burdens may be a greater economic hardship for households that have relatively lower incomes. For policy purposes, it may be useful to compare rent burdens across programs after adjusting for income differences. This adjustment is made using Barskey et al.'s (2002) simple and powerful nonparametric propensity score weighting method.<sup>3</sup>

Exhibit 4 reports sample sizes and percentage frequencies for the 13 income categories by program. Compared with HCVP households, PH families were overrepresented in the upper income categories, but PBS8 families were underrepresented. An estimated 7.0 percent of voucher households had adjusted annual incomes of at least \$25,001 compared with an estimated 8.9 percent of PH households and an estimated 2.9 percent of PBS8 households.

<sup>&</sup>lt;sup>3</sup> Barskey et al.'s (2002) reweighting technique is similar to that of DiNardo, Fortin, and Lemieux (1996).

Exhibit 3

## Rent Burden Percentage Frequencies by Program

#### **Housing Choice Voucher Program**

Rent Burden	N	Weighted Percent	Standard Error
0–31%	58,830	59.449	0.149
32-39%	18,588	17.895	0.119
40-49%	8,974	8.399	0.085
50% or more	5,446	5.159	0.068
Missing	8,959	9.097	0.087

#### **Public Housing**

Rent Burden	N	Weighted Percent	Standard Error	Income-Adjusted Percent	Standard Error
0-31%	53,679	85.253	0.148	88.093	0.118
32-39%	497	0.780	0.037	0.776	0.034
40-49%	450	0.690	0.034	0.701	0.032
50% or more	1,032	1.665	0.055	1.558	0.047
Missing	6,029	11.612	0.138	8.872	0.103

#### **Project-Based Section 8**

Rent Burden	N	Weighted Percent	Standard Error	Income-Adjusted Percent	Standard Error
0–31%	59,943	85.772	0.128	88.506	0.103
32–39%	1,295	1.594	0.045	1.506	0.040
40-49%	1,083	1.365	0.042	1.233	0.036
50% or more	4,111	5.994	0.089	4.596	0.067
Missing	3,740	5.275	0.084	4.160	0.065

Notes: Rent burden is defined as the family's contribution to gross rent, including utilities, divided by adjusted monthly income. Rent burden is missing for households with \$0 adjusted income.

Source: U.S. Department of Housing and Urban Development Public Use Microdata Sample 2009

For PH and PBS8 households in income category j, the proportion of voucher households in that category,  $p_j$ , can be used as a nonparametric estimate of the propensity score of having the same income as a voucher household. For instance, an estimated 15.8 percent of HCVP households had incomes between \$0 and \$5,000 in 2009. For the 20.2 percent of PH households and 19.9 percent of PBS8 households in the \$0-to-\$5,000 income category, 0.158 is used as an estimate of the probability that these households had the same income as a voucher household.

Each PH and PBS8 household falls in 1 income category, and the sum of possible propensity scores across categories equals 1. Thus the propensity scores specify an alternative probability mass function for the 13 possible income categories for PH or PBS8 households.

For propensity score weighting, an adjusted weight  $w^*_{j,k}$  can be used equal to  $p_j W_k / N_{j,k}$ , where  $W_k$  is the total sum of original sampling weights for program k, and  $N_{j,k}$  is the sample size in income category j for program k.

Exhibit 4 reports the propensity score weights by income category for PH and PBS8. For example, an estimated 22.6 percent of voucher households had adjusted incomes between \$10,001 and

Exhibit 4

Income Percentage Frequencies and Propensity Score Weights

# **Housing Choice Voucher Program**

Adjusted Annual Income (\$)	N	Weighted Percent (p)	Standard Error
0-5,000	15,188	15.827	0.113
5,001-10,000	34,535	35.638	0.150
10,001-15,000	22,950	22.610	0.130
15,001-20,000	13,133	12.478	0.101
20,001-25,000	7,047	6.486	0.074
25,001-30,000	3,772	3.400	0.054
30,001-35,000	2,000	1.757	0.038
35,001-40,000	1,089	0.919	0.027
40,001-45,000	542	0.452	0.019
45,001-50,000	264	0.212	0.013
50,001-55,000	139	0.111	0.009
55,001-60,000	51	0.040	0.005
60,001–90,000	87	0.069	0.007

# **Public Housing**

Adjusted Annual Income (\$)	N	Weighted Percent	Standard Error	Propensity Score Weight (w*)
0-5,000	10,862	20.246	0.161	16.573
5,001-10,000	21,633	36.873	0.200	18.738
10,001-15,000	11,603	18.776	0.162	22.164
15,001-20,000	6,654	9.955	0.120	21.330
20,001-25,000	3,737	5.242	0.085	19.741
25,001-30,000	2,447	3.203	0.064	15.805
30,001-35,000	1,613	1.998	0.048	12.392
35,001-40,000	1,089	1.306	0.038	9.595
40,001-45,000	744	0.878	0.031	6.912
45,001-50,000	532	0.618	0.026	4.525
50,001-55,000	313	0.365	0.020	4.049
55,001-60,000	224	0.258	0.017	2.021
60,001-90,000	236	0.283	0.018	3.328

### **Project-Based Section 8**

Adjusted Annual Income (\$)	N	Weighted Percent	Standard Error	Propensity Score Weight (w*)
0-5,000	14,230	19.936	0.146	15.109
5,001-10,000	26,390	39.671	0.185	18.345
10,001-15,000	16,389	23.839	0.161	18.741
15,001-20,000	7,471	9.940	0.109	22.689
20,001-25,000	3,015	3.686	0.066	29.223
25,001-30,000	1,433	1.623	0.042	32.233
30,001-35,000	650	0.703	0.027	36.728
35,001-40,000	298	0.311	0.018	41.877
40,001-45,000	157	0.159	0.012	39.119
45,001-50,000	70	0.067	0.008	41.071
50,001-55,000	35	0.033	0.005	43.252
55,001-60,000	14	0.013	0.003	38.614
60,001-90,000	20	0.018	0.004	46.898

Source: U.S. Department of Housing and Urban Development Public Use Microdata Sample 2009

\$15,000, PUMS contains 11,603 PH households in this category, and the total sum of weights for PH was 1,137,413 (exhibit 1). The propensity score weight for PH households in the \$10,001-to-\$15,000 income category is 0.226 times 1,137,413 divided by 11,603, which equals 22.2.

For both PH and PBS8, the total sum of propensity score weights equals the sum of original sampling weights. Weighting PH and PBS8 households by the propensity score weights, the percentage of PH and PBS8 households in each income category is identical to the HCVP percentage. The only differences in the income distributions across programs after propensity score weighting are due to differences within income categories.

Exhibit 3 reports income-adjusted percentages of PH and PBS8 households by rent burden category, with standard errors. Propensity score weighted estimates indicate that, when comparing households that have similar incomes, rent burden differences between the HCVP and the other two programs are even more pronounced.

In 2009, an estimated 59.5 percent of HCVP households had rent burdens of less than 32 percent. Before accounting for income differences, an estimated 85.3 percent of PH households had rent burdens of less than 32 percent, as did an estimated 85.8 percent of PBS8 households. After propensity score weighting, an estimated 88.1 percent of PH households and an estimated 88.5 percent of PBS8 households had rent burdens of less than 32 percent.

Not adjusting for income differences, the mean estimated difference in percentages of PH and HCVP households in the 0- to 31-percent rent burden category was 25.8 percentage points, with a 95-percent confidence interval of (25.4, 26.2). Adjusting for income differences, the mean estimated difference was 28.6 percentage points, with a 95-percent confidence interval of (28.3, 29.0).

Ignoring income differences, the mean estimated difference in percentages of PBS8 and HCVP households in the 0- to 31-percent rent burden category was 26.3 percentage points, with a 95-percent confidence interval of (25.9, 26.7). Matching on income, the mean estimated difference was 29.1 percentage points, with a 95-percent confidence interval of (28.7, 29.4).

No S202 and S811 households were in the 2009 PUMS sample for some income categories. To compare rent burdens in S202 households against rent burdens in the other programs, the propensity scores can be based on the proportion of S202 households in each income category, thereby giving households in the other programs zero weight for income categories with no S202 households. To adjust for the possibility that the S202 population contains households in more income categories than is reported in PUMS, a Bayesian method can be used to smooth the income distribution.

# **Conclusion**

To promote government openness and transparency, in February 2012, the U.S. Department of Housing and Urban Development launched a new Public Use Microdata Sample, or PUMS, database, which contains household-level data for 5 percent of households that are assisted through the Housing Choice Voucher Program, public housing, Section 8 Project-Based Rental Assistance, the Section 202 program that subsidizes supportive housing for the elderly, and the Section 811 program that subsidizes supportive housing for people with disabilities.

PUMS includes household characteristics as well as geographic information, enabling researchers to perform analyses not possible using other datasets.

This article presented a data analysis example using PUMS data comparing rent burdens across programs, adjusting for income differences using Barskey et al.'s (2002) nonparametric technique. Before accounting for income differences, rent burdens are much higher in the voucher program compared with HUD's other rental assistance programs. Results indicate that when comparing households that have similar incomes, rent burdens of HCVP households are even higher compared with those of other assisted households.

# Acknowledgments

The author thanks Dav Vandenbroucke for helpful comments.

# Author

Brent D. Mast is a social science analyst at the U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Program Monitoring and Research Division.

# References

Barskey, Robert, John Bound, Kerwin Kofi Charles, and Joseph P. Lupton. 2002. "Accounting for the Black-White Wealth Gap: A Nonparametric Approach," *Journal of the American Statistical Association* 97 (459): 663–673.

Casey, Connie H. 1992. *Characteristics of HUD-Assisted Renters and Their Units in 1989*. Report prepared for the U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Washington, DC: U.S. Government Printing Office.

DiNardo, John, Nicole Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach," *Econometrica* 64 (5): 1001–1044.

Lohr, Sharon. 1999. Sampling: Design and Analysis. New York: Duxberry Press.

Mast, Brent D. 2012. "Comparing Public Housing and Housing Voucher Tenants With Bayesian Propensity Scores," *Cityscape: A Journal of Policy Development and Research* 14 (1): 55–72. Also available at http://www.huduser.org/portal/periodicals/cityscpe/vol14num1/Cityscape\_Mar2012\_comp\_publsg.pdf.

Rucinski, Dianne, and Leslie Athey. 1996. "Identifying Recipients of Housing Assistance Through Survey Questions." Presented at the 51st Annual Conference of the American Association for Public Opinion Research, Salt Lake City, UT, May 16–19.

Shroder, Mark. 2002. "Does Housing Assistance Perversely Affect Self-Sufficiency? A Review Essay," *Journal of Housing Economics* 11: 381–417.



# **Additional Reading**

Mast, Brent D. Forthcoming. "Markov Chain Model of Rent Burden in the Housing Choice Voucher Program," *Journal of Housing Research*.

McClure, Kirk. 2005. "Rent Burden in the Housing Choice Voucher Program," *Cityscape: A Journal of Policy Development and Research* 8 (2): 5–20. Also available at http://www.huduser.org/Periodicals/cityscape/vol8num2/ch1.html.