Comparison of Population Estimates in the 2002 and 2005 New York City Housing and Vacancy Surveys
July 7, 2008

The New York City Housing and Vacancy Survey (NYCHVS) is a comprehensive survey that collects and produces data on the quality and quantity of housing in the City and the demographic, social, and economic characteristics of the people in those housing units. Public officials, private organizations, and individual researchers use the information from the survey to develop, analyze, and evaluate policies and programs.

The 2005 NYCHVS data are generally comparable to the 2002 NYCHVS data. However, included in the large amount of information from the survey are counts and characteristics of the population by race and ethnicity. Over the last several surveys, questions have been raised as to the consistency of the race and ethnicity estimates within each survey and from one survey to the next. To properly use and understand these data from the NYCHVS requires knowledge of the methodology and techniques used by the Census Bureau to collect, process, and present the data. That information is provided in detail in the form of questions and answers below beginning with options on how to best use the race and ethnicity data from the 2002 and 2005 NYCHVS.

1. What are the options for using population data by race and Hispanic origin from the NYCHVS and from other sources?

Response: (a) Population data from the 2002 and 2005 NYCHVS can be used to measure population levels for individual race groups and by Hispanic origin, as well as to make comparisons between groups for a particular survey year. (b) For comparisons of characteristics by race and Hispanic origin between survey years, users are encouraged to use percentages, means, and medians rather than absolute numbers. (c) To compare population levels by race and Hispanic origin yearly over time, users should consider the annual population estimates produced as part of the Census Bureau's Population Estimates Program found at www.census.gov/popest/estimates.php.

2. How was the 2002 and 2005 NYCHVS sample determined?

Response: The sample for the 2002 and 2005 surveys consisted of housing unit addresses selected from three different sources.

- Housing units included in Census 2000 selected from the Census 2000 address file.

- Housing units built since Census 2000 selected from New York City Certificates of Occupancy (C of O). For the 2002 NYCHVS, the selection was based on C of O issued between January 2000 and November 2001; and for 2005, those issued between December 2001 and October 2004. Housing unit addresses that were in both the Census and on the C of O lists were dropped from the latter.

- Housing units in structures owned by New York City as a result of real estate tax delinquency or failure to pay other charges or fees (known as in rem units). Since all units on the in rem list were also in the census or on the C of O list, the weighting of these units was adjusted to reflect the additional chance of selection.

The sample for the 2002 NYCHVS was 18,293 housing units of which 17,157 were completed interviews. For 2005, there were 18,516 sample units and 16,834 completed interviews.

3. How was the population data collected in the NYCHVS?

Response*:* Census Bureau field representatives visited every 2002 and 2005 NYCVHS sample address to collect information about each housing unit in the sample and, if the sample unit was occupied, about the people living in the unit. A household roster was developed and demographic characteristics (age, sex, race, Hispanic origin), as well as economic and social characteristics, were collected for each member of the household.

4. **How were population estimates developed for the NYCHVS?**

Response*:* To compute population estimates, sample weights were calculated for each person in the household based on a multi-stage process.

- Base weight – the base weight for each person was the reciprocal of the probability of selecting the housing unit for the NYCHVS sample. For example, Census 2000 counted 3,200,912 housing units in New York City. Since the sample size for the 2005 survey was 18,516, the probability of selection was approximately 1/173, and the base weight approximately 173.

- Nonresponse adjustment - The base weight was adjusted to account for household noninterviews resulting from respondent refusals, the inability to locate a knowledgeable respondent after repeated tries, or incomplete interviews.

- Ratio adjustments – Several ratio adjustments were applied in the weighting process including ratio adjusting to independently developed control estimates of population and housing units. For example, if the independent control estimate for the White population in Manhattan was 1,000,000, while the survey estimate was 995,000, the weight for selected people in the household would be adjusted by 1,000,000/995,000, or 1.0050.[1]

The final weight for each person equaled the product of the base weight and all adjustments.

[1] The NYCHVS controlled for age and sex as well as for selected race groups. This example does not include age or sex in order to provide a simple, straightforward example of how the factors are applied.

5. Why does the NYCHVS adjust population estimates to independently developed controls?

Response: The Census Bureau has used independently developed population controls as part of the weighting process for the NYCHVS since the 1975 survey (housing unit controls have been used since the 1991 NYCHVS). The Census Bureau develops these independent estimates as a byproduct of its Population Estimates Program that provides annual estimates of population and housing units for the United States, states, counties, and other geographic areas (this program is undertaken by the Bureau's Population Division). The population controls applied in the NYCHVS are by borough, age, race, and sex. They were originally developed beginning with Census 2000 population totals and then adding or subtracting the demographic components of population change such as births, deaths, and net domestic and international migration.[2] Controls are used in the NYCHVS, as they are for most other demographic surveys, for several reasons:

- They insure that survey estimates of total population equal a "known" total, and that there is a certain amount of consistency between different surveys for the estimates being controlled.

- They correct for known coverage errors that are common to all household surveys. Research has shown that surveys tend to miss a substantial number of people within households.

- They eliminate the variance for the survey estimates being controlled, and reduce the variances of survey estimates that are correlated with the controlled estimates.

Items that are not controlled to independent estimates (such as income, educational attainment, etc.) are more subject to sampling variability. Using independent estimates for certain characteristics does not have any direct effect on other survey estimates.

6. Can using independently developed population controls have differential effects on population estimates by race and Hispanic origin?

Response: Yes, using population controls can have differential effects on estimates of race and Hispanic origin depending on which groups are being controlled as opposed to which groups are of interest. The 2002 and 2005 NYCHVS used population controls for the following race groups: White, Black, and a catchall All Other Races group.[3] No controls were available by Hispanic origin.[4]

---

[2] For more detailed information on the methodology used to produce population estimates go to www.census.gov/popest/estimates/php.

[3] In the text of this document, the terms White, Black, and Asian refer to all Whites, all Blacks, or all Asians regardless of their Hispanic origin. The terms White, non-Hispanic; Black, non-Hispanic; and Asian, non-Hispanic are used when Hispanics are not included in the group. People of more than one race are included in the All Other Race category.

[4] People of Hispanic origin may be any race.

An example may be illustrative. Assume Blacks, Hispanics, and All Other Races were undercounted at a higher rate than Whites in the 2005 NYCHVS. The ratio estimate adjustment for Blacks and All Other Races would be larger than the adjustment for Whites. Also assume that more Hispanics answered that they were White in the race question than answered that they were Black or any other race. Since there were no controls specifically for Hispanics, more Hispanics would receive the lower adjustment factor for Whites than the higher factor for Blacks or All Other Races. As a result, the adjustment factors applied to Hispanics would not adequately adjust for the actual undercount of Hispanics in the survey. Additionally, different sub-groups within the Hispanic group, for example, Puerto Ricans, might have been over or undercounted at different rates than other sub-groups. This would affect whether or not the adjustment factors applied to these groups were appropriate.

7. Can using independently developed population controls have differential effects on population estimates by race and Hispanic origin across surveys (2002 and 2005)?

Response: Yes, population controls can have differential effects on the estimates of race and Hispanic origin across surveys depending on the coverage rates for each of the groups being controlled as opposed to the groups of interest. Another example may be useful. Assume that Blacks, Hispanics, and All Other Races were undercounted at the same rate in both the 2002 and 2005 NYCHVS, and that more Hispanics answered White to the race question in both years. Now assume that overall Whites were overcounted in the 2002 survey (their adjustment factor would be less than 1.0) and undercounted in the 2005 survey (their adjustment factor would be greater than 1.0). Since there were no controls specifically for Hispanics, more Hispanics would receive the adjustment factor for Whites in both 2002 and 2005 than for the other groups. Once again the adjustment factors applied to Hispanics would not adequately reflect the situation as it actually existed.

8. What impact did the independent controls have on the 2002 and 2005 NYCHVS population estimates of race and ethnicity?

Response: It is difficult to separate out the effects of each of the steps in the weighting process, but the overall affect of the weighting can be observed.

The unweighted results of the 2002 and 2005 NYCHVS showed a count of people as follows:

	2002		2005	
	Number	Percent	Number	Percent
Total	38,950	100.0	37,740	100.0

White, alone[5]	22,156	56.9	21,018	55.7
Black, alone[5]	11,817	30.3	11,084	29.4
All other races[5]	4,977	12.8	5,638	14.9
Hispanic[5]	10,456	26.8	10,694	28.3
Puerto Rican	3,780	9.7	3,820	10.1
Not Hispanic	28,494	73.2	27,046	71.7

As noted earlier, although the sample size for the 2005 NYCHVS was larger than for the 2002 survey, the number of completed interviews was less. The table above shows that the smaller number of completed interviews resulted in a smaller number of people with complete data records. The table also shows a decrease in the number and percentage of respondents reporting White and Black and an increase in those reporting Hispanic origin and Puerto Rican. The proportional increase in Hispanics was greater than the increase in Puerto Ricans. An important point from this table is that more respondents in the survey reported themselves and other household members as Hispanic and as Puerto Rican in the 2005 survey than in the 2002 survey.

The weighted estimates from the two surveys were as follows:

	2002		2005	
	Number	Percent	Number	Percent
Total	7,944,577	100.0	8,011,655	100.0
White, alone	4,519,893	56.9	4,555,359	56.9
Black, alone	2,365,266	29.8	2,315,734	28.9
All other races	1,059,419	13.3	1,140,563	14.2
Hispanic	2,087,496	26.3	2,229,378	27.8
Puerto Rican	742,342	9.3	805,538	10.1
Nonhispanic	5,857,081	73.7	5,782,277	72.2

There are a number of different ways to look at the two tables above. One shows that the change in the proportion of Blacks and Hispanics between 2002 and 2005 was the same using weighted estimates as it was using unweighted data, while the change in the proportion of Puerto Ricans was 0.4 percentage points greater using the weighted estimates. Another shows that the weighting decreased the proportion of Hispanics in both 2002 and 2005 by 0.5 percentage points, while at the same time the proportion of

[5] The term ”alone” in this table refers to people of a single race. The category All Other Races includes people of more than one race. Hispanics may be any race.

Puerto Ricans decreased by 0.4 percentage points in 2002, but was not affected by the weighting in 2005.

9. Are there other factors that make it difficult to compare population estimates of race and Hispanic origin from the 2002 and 2005 NYCHVS?

Response: Yes**,** New York City challenged the Census Bureau's annual population estimates developed for the Population Estimates Program in 2003, 2004, and 2005, and in each case the City's challenge was accepted and the estimates revised. The revised estimates resulted in an additional 29,393 people in 2003; 64,259 in 2004; and 70,642 in 2005. In addition, each time a revision occurred, the Census Bureau recalculated earlier annual population estimates back to Census 2000.

The independently developed population controls used in weighting the 2005 NYCHVS reflected all of the challenges through 2005 as well as any other revisions that occurred between 2002 and 2005. The 2002 NYCHVS population results have not been reweighted to reflect any revisions to the annual independent population estimates that occurred after the release of the survey data.

10. Review the change in the White, non-Hispanic population and provide any additional data that may support the White, non-Hispanic population increase in the City, and particularly clarify the order of magnitude of the increase.

***Response*:** Table 1 shows results from the 2002 and 2005 NYCHVS. The estimates of the White, non-Hispanic population did not change between the two years (they are not statistically different at the 90-percent confidence level). Table 2 provides estimates from the Census Bureau's Population Estimates Program (prepared by the Bureau's Population Division), and shows a small increase in the White, non Hispanic population of 12,339 between 2002 and 2005, consistent with the NYCHVS results. The estimates from the Population Estimates Program are for the population in housing units and are comparable to the data collected in the NYCHVS.

The two tables also show that the total White population (both Hispanic and non-Hispanic) from the Population Estimates Program increased by 61,225 in the three-year period compared to 35,466 in the NYCHVS (the difference between the 2002 and 2005 total White population in the NYCHVS is not statistically significant).

It is difficult to compare race data with the American Community Survey (ACS) because those answering 'Some Other Race' in the ACS are not allocated to the major race categories as they are in the NYCHVS. Table 3 shows that in 2005 the ACS reported that 1,355,266 persons in New York City classified themselves as Some Other Race, compared to 938,665 in 2002. The estimate of the total White population decreased 174,717, from 3,673,929 to 3,499,212 from 2002 to 2005. Since this change is largely the result of the large increase in the number of Some Other Race responses, these data must be interpreted with caution.

11. Review the Black, non-Hispanic population decrease with data on the change in the number of Black, non-Hispanics from other data sources and determine the reasonableness of the change the 2005 NYCHVS shows.

Response: The NYCHVS results in Table 1 show a decrease of 102,722 in the number of Black, non-Hispanics between 2002 and 2005. Table 2, from the Population Estimates Program, shows a decrease in the Black-non Hispanic population of 14,896 and a decrease of 19,752 in the total Black population (including Hispanics). Table 2 provides the latest series of population estimates reflecting all challenges to the estimates. For example, when the Census Bureau accepted New York City's challenge to the 2005 population estimates, annual population estimates data from 2001 to 2004 were all revised.

From 2002 to 2005, the ACS estimate of the total Black population decreased by 110,526, from 2,122,488 to 2,011,962, while the NYCHVS showed an apparent decline of 49,532 in total Black population (Tables 1 and 3). The difference of 49,532 from the NYCHVS is not statistically significant.

Although the change is not statistically significant, we would like to see more consistency from NYCHVS to NYCHVS and between the NYCHVS and other surveys. One reason for inconsistencies can be attributed to the fact that the NYCHVS does not control estimates for Hispanic persons. The independently developed population controls used for weighting the NYCHVS are by three race categories only---White, Black, and All Other Races. However, even if the 2002 and 2005 data used controls based on both race and ethnicity, there still would be problems comparing data, since the Bureau's independent population estimates for New York City were challenged and revised three times between 2002 and 2005. Population data from the 2005 NYCHVS reflect these revisions while those from the 2002 survey do not. The 2002 NYCHVS data reflect the best estimates that we had at the time at the time of the survey.

12. Review the Puerto Rican population increase with data from other sources and explain the causes of such an apparent increase.

Response: The NYCHVS shows an increase of 63,196 in Puerto Ricans between 2002 and 2005, from 742,342 to 805,538 (Table 1). This compares to a decrease of 76,143, from 863,189 to 787,046 in the ACS (Table 3). There is no readily apparent explanation for this divergence. However, a review of unallocated and unweighted NYCHVS data from the 2002 and 2005 NYCHVS also shows an increase between the two years indicating that the increase in the NYCHVS reflects actual reporting and was not unduly affected by the weighting and editing of the data.

The Census Bureau's Population Estimates Program does not produce independent estimates specifically for Puerto Ricans. According to the latest annual independently produced population estimates for all Hispanics, their number increased by 45,660, from 2,221,255 in 2002 to 2,266,915 in 2005. This compares to an increase in Hispanics of

141,882 for the NYCHVS, from 2,087,496 in 2002 to 2,229,378 in 2005 (Tables 1 and 2). The ACS showed a decline of 16,295 Hispanic persons from 2002 to 2005.

As mentioned, we do not control NYCHVS results by Hispanic origin. Our population controls by race are for White, Black, and All Other Races. We believe that controlling survey estimates for Hispanics would lead to more consistent data from survey to survey and we will explore this possibility. However this would not totally solve the comparability problem, since data are frequently being challenged/revised, and since population controls for Puerto Ricans are not available.

13. Review the Asian, non-Hispanic population increase with data from other sources and determine the causes of such a small apparent increase in the Asian, non-Hispanic population from the 2002 and 2005 NYCHVS.

Response: The Population Estimates Program showed an increase of 61,118 non-Hispanic Asians, which is likely the best estimate of change between 2002 and 2005 (Table 2). The NYCHVS showed no statistical difference in the number of non-Hispanic Asians in the three-year period (Table 1). As discussed earlier, we do not control the NYCHVS survey estimates for the Asian population. The controls used for weighting are by White, Black, and All Other Races. Asians would be part of the All Other Race group. Lack of independent controls for Asians is probably part of the reason for the inconsistency between the NYCHVS and the Population Estimates Program. Variations of coverage from survey to survey will affect results, particularly for characteristics not controlled. Also, keep in mind that the latest annual population estimates reflect all challenges to date, while the survey results for 2002 do not reflect any challenge results. This may be another cause of the difference between the survey results and the population estimates. The ACS showed an increase of 32,175 the total number of Asians (Hispanic and Non Hispanic) from 2002 to 2005.

14. How are the ACS race and ethnicity data collected? Provide findings on the differences in the ACS and the NYCHVS race and ethnicity data collection methods.

Response: Race and ethnicity are self-identification items in both the ACS and the NYCHVS. However, there are significant differences in data collection between the ACS and the NYCHVS. Most significant is the mode of collection. The ACS uses three modes of data collection: mailout/mailback, computer assisted telephone interviewing (CATI), and computer assisted personal interviewing (CAPI). Research shows, that for some items respondents answers differ by mode of collection. In 2005, approximately 56 percent of the interviews for New York City were by mail, 14 percent were CATI, and 31 percent were CAPI. The NYCHVS is strictly a personal interview survey.

Although similar in wording, the race and ethnicity questions differ between the surveys. The ACS allowed for more racial distinctions in its race question (15) than the NYCHVS (12). The ACS allowed respondents to report "Some Other Race" and then provide a written description of that race. If possible, this description was then coded into one of the other race categories. Those responses that were not coded were left as Some Other

Race. In the 2005 ACS, 1,355,266 people in the City were classified in this category. The NYCHVS also allowed a response of Some Other Race, but all of these responses were allocated to one of the other major race categories.

The ACS Hispanic origin question allowed for five response options including a Not Hispanic option and an Other Hispanic write-in option. The Other Hispanic category was coded if possible to one of the other categories. The NYCHVS allowed for seven response options including Not Hispanic and Other Hispanic. The Other Hispanic category allowed for a write-in response, but this response was not coded.

15. Provide a comparison of the following NYCHVS population changes with changes in the Census Bureau's annual population estimates for New York City and the causes of the discrepancies in the changes between the two data sources for:

a. The level of the White, non-Hispanic population increase shown by the 2005 NYCHVS

b. The level of the Black, non-Hispanic population decrease shown by the 2005 NYCHVS

c. The level of the Puerto Rican population increase shown by the 2005 NYCHVS

d. The level of the Asian, non-Hispanic population increase shown by the 2005 HVS

Response: See Tables 1, 2, and 3 for a variety of comparisons (note - we did not include a table on estimates from the Current Population Survey (CPS), because comparisons of race from 2002 to 2005 are nearly impossible, since the CPS did not allow for multiple race entries until 2003, while the 2002 NYCHVS allowed the reporting of multiple races. In addition, the 2002 CPS estimates are weighted based on the 1990 census, while the 2005 estimates are weighted based on the 2000 decennial census.)

One reason for the difference in the estimates is that the population controls used for the NYCHVS for weighting are by race only—White, Black, and All Other Races. There are no controls for Hispanics. Variations of coverage from survey to survey will affect results, since different groups of the population have different coverage rates. The independent population estimates from the Population Estimates Program are not the result of any survey, but are based on a variety of data sources (see link below for a description of the methodology). Also contributing to the inconsistency is the fact the 2002 NYCHVS results do not reflect any challenge results since that date.

16. Are the data collection methods the Census Bureau applied in collecting race and ethnicity data for the annual population estimates and the NYCHVS for the City the same or very similar? If not, please explain the differences.

Response: No, the annual independent population estimates from the Population Estimates Program are not the result of a survey. They are estimates prepared using a

variety of data sources. To produce borough and city totals, the NYCHVS results are controlled to these independent estimates. All demographic surveys are controlled to independent population and/or housing unit estimates. For a description of the methodology and sources of information used to develop the annual population estimates, go to http://www.census.gov/popest/estimates/php.

17. What changes in procedures could the Census Bureau implement that would preclude the need for and the impact of recurring challenges and revisions to the annual population estimates?

Response: We are not familiar enough with the details of the Population Estimates Program to offer suggestions on how challenges can be prevented. However, this question will be passed on to the appropriate staff in the Population Division.

18. The fact that there are no independent population controls for Hispanics is a substantial problem, especially applied to the NYC environment. Can the Census Bureau begin planning to include independent population controls for Hispanics, with a sub-category for Puerto Ricans for the 2011 NYCHVS?

Response: One possible improvement in the way we use the independent population controls, is to control the NYCHVS data for Hispanics. This would likely make the overall estimate for this group more consistent from survey to survey. However, this would not directly help with our estimates of Puerto Ricans or other sub-groups. There are no controls currently available for Puerto Ricans thus it will not be possible to control for this sub-category in 2011.

19. Would reweighting the 2002 data make the race and ethnicity data more comparable to the 2005 NYCHVS data?

Response: Reweighting the 2002 data could slightly improve the consistency between the 2002 and 2005 data, but would likely have no significant overall effect. In addition, reweighting the 2002 survey data for population would likely start a chain reaction resulting in the reweighting of housing unit data for 2002 and possibly population and housing unit data for 2005. In addition, it would set a precedent for calls to reweight future NYCHVS data whenever challenges to the annual independent population estimates on which NYCHVS controls are based were accepted. Finally, if we did reweight 2002 NYCHVS data, we would have two sets of population and housing unit estimates in the public domain. This would cause unnecessary confusion in the user community, and undoubtedly cause some to call into question the validity of NYCHVS results.

20. Is there anything that can be done to make estimates of race and ethnicity more comparable between 2005 and 2008 NYCHVS?

Response: As mentioned, one possible improvement for comparing estimates of race and ethnicity between 2008 and future surveys would be to control the population by

Hispanic origin in addition to race, beginning with the 2008 NYCHVS. This would likely lead to more consistent estimates for this group between surveys. However, there are no controls available for Hispanic sub-groups, such as Puerto Ricans. Therefore this group will be subject to more sampling variability than other groups where controls are available for the foreseeable future. We can also begin exploring the possibility of controlling for other large race groups such as Asians.