"The Utilization of First Names to Evaluate Reports of Gender and Its Effect on the Distribution of Married and Unmarried Couple Households"

Authors: Martin O'Connell and Gretchen Gooding
Affiliation: US Census Bureau
Address: US Census Bureau
Fertility and Family Statistics Branch
Population Division, 2351-3
4700 Silver Hill Road
Suitland, MD 20746
Phone: 301-763-2416
Email: Martin.T.Oconnell@census.gov (Martin O'Connell)
Gretchen.E.Gooding@census.gov (Gretchen Gooding)

ABSTRACT

In Census Bureau classifications, married-couple households consist of opposite-sex couples, while unmarried partner households may consist of either opposite or same-sex couples. This classification relies not only on the accuracy of the household relationship responses—either as a spouse or unmarried partner—but also on those of gender. Although gender is usually the most accurately reported item on a survey, an analysis of the names of people may occasionally be at odds with their reports on gender. Because the number of unmarried-partner households is relatively small, minor errors in gender could have a substantial impact on estimates. Using the 2004 Test Census of New York, we analyze 60,244 "coupled households" in this survey—55,026 married couples, 4,112 opposite-sex unmarried partners, and 1,106 same-sex unmarried partners—to see how the utilization of male-female sex probabilities attached to each person's name would alter this distribution.

"The Utilization of First Names to Evaluate Reports of Gender and the Effect of this Process on the Distribution of Married and Unmarried Couples"

By Martin O'Connell and Gretchen Gooding

Population Division, US Census Bureau

One of the most widely discussed household and family tabulations from Census 2000 concerned that of unmarried partner households. Of the 5.5 million unmarried partner households in 2000, 4.9 million were opposite-sex partners while another 0.6 million were same-sex partners. When added to the 54.5 million married-couple households (consisting only of spouses of the opposite sex), there were a total of 60 million households containing married or unmarried couples (coupled households).[1]

Crucial to the classification of coupled households into one of these three groups is the joint combination of responses to the relationship of the partner to the householder (a spouse or an unmarried partner) and the gender of the two people. In the editing process of Census 2000, unlike the 1990 Census, if a household consisted of a married couple, with both spouses reporting the same sex—and where no imputations were made for either person for either their relationship or sex due to incomplete data or non-response—then the partner who reported being a spouse of the householder was changed to being an unmarried partner of the householder. In 1990, the *relationship* category would have remained the same (spouse) but the *sex* of the partner would have usually been changed.

---

[1] Tavia Simmons and Martin O'Connell, *Married Couple and Unmarried Partner Households: 2000,* Census Special Reports, CENSR-5 (US Census Bureau: Washington, DC, 2003).

This decision was made for two basic reasons. First, individual reports of sex on surveys have usually the best consistency responses when re-interviews are made. In addition, the gender item on the Census 2000 form had the lowest imputation rate of all items asked on the 100 percent Census form (1.1 percent in Census 2000). If any item could be counted on as being correctly answered, it likely would be the item on gender.[2] Second, since the 1996 Federal Defense of Marriage Act instructed all Federal agencies only to recognize opposite-sex marriages for the purposes of enacting any agency programs, reports of same-sex spouses were assigned as unmarried partners instead of randomly assigning them relationship codes through normal allocation routines.

Using the 2004 Test Census of New York conducted in the county of Queens, this paper will examine the gains or losses to different types of coupled households if respondents' names were utilized to verify their report of their sex. This sample survey contained 60,244 "coupled households"—55,026 married couples, 4,112 opposite-sex unmarried partners, and 1,106 same-sex unmarried partners. Unique to this survey is a probability index attached to each person's name that identifies the "maleness" of the name. This index was developed from the existing 2000 Census file of New York state and is constructed by taking the ratio (from 0 to 1000) of the number of times this name was recorded by a male to the total number of times this name was recorded by either a male or female.[3] For example, a ratio of 950 indicates that when this name appeared in the Census of New York in 2000, 950 times out of 1000 that person was a man. A ratio

[2] Not counting the 105 million householders—who were not asked their relationship—about 3.4 percent of people had their relationship imputed.

[3] This file was used since was generated using millions of observations collected in New York in Census 2000.

of 20 would indicate that only 20 times out of 1000 the person was a man or conversely, 980 times out of 1000 that name was identified as being reported by a woman. A decision, then, could be made as to whether to accept the respondent's reply of their sex on the likelihood of this gender indicator or to nullify their response and change it to the sex with the higher probability level.

By setting different "acceptance levels" for this indicator, one could see the effect of using an alternative piece of information—a person's name—in the review or editing of data files.[4] This paper will examine how many same-sex unmarried partner households have partner names that could imply a mistake in the marking of their gender—that they are likely to be opposite-sex couples—hence, an *overestimate* of same-sex partners has resulted in the data. But this paper also addresses the following issue: How many opposite-sex couples (married or unmarried), when using the same verification procedure, would generate an alteration of one partner's sex, thus *adding* to the count of same-sex unmarried partners?

We will look at the gross flows of gains and losses produced by a consistent application of a name verification routine, regardless of the relationship status of the couple, and see what the net addition or loss to the different coupled universes would be under different levels of confidence. There may be other issues in the application of names to be used in editing a person's response to the gender item—for example,

[4] Some Census 2000 editing routines did use a person's name to assign a male/female value for the gender item when that question was left blank on the form.

language issues, accuracy in optical scanning of handwritten names—but those cannot be addressed in this paper.

Different levels of name "acceptance" produce changes in estimates of household types. We created three index ranges (0-10, 0-50, 0-100) to edit sex responses. An index of 0-10 means that 99% of people with that name were of the opposite sex in Census 2000; 0-50 means that 95% were of the opposite sex; and 0-100 means that 90% were of the opposite sex. Thus, the lower the index level of a respondent's first name, the more frequently that name was associated with the opposite sex. For example, 98 percent of the people with the name of "Mary" in the 2004 American Community reported that they were female, compared with 86 percent of people with the name of "Robin" and 75 percent of people with the name of "Leslie." Some respondents with these names may have mis-marked their response in the sex item as male while others may, in reality, be male and not female. The question is, how confident is one in accepting a "name" response over a "sex" response?

Distributions are indeed altered when the sex probabilities attached to each person's name as used to overrule the response in the sex item. Using the most conservative index level (0-10), there is a net increase from 1,106 to 1,449 couples for same-sex couple estimates. At the 0-100 index level, the estimate for same-sex couples increases to 1,819. Looking at the components of change at the 0-100 index level, there is a loss of 311 same-sex couples at the 0-100 index level because they are reassigned to the opposite-sex couple category. However, this procedure transfers 956 married couples

and 68 opposite-sex couples to the same-sex couple category, thus offsetting this loss. Due to the magnitude in the number of married couples, even transfers of only 1 or 2 percent of these couples produce a relatively large addition to the same-sex population. Overall, while first names offer the potential to edit/verify reports of sex on questionnaires, using first names in an impartial and systematic way to invalidate reported sex responses will yield more same-sex couples than originally reported.