Comment on McVeigh and Diaz, ASR, December 2009

Accurate Measurement of Same-Sex Couples: Empirical Implications

Rebecca DiBennardo[a] and Gary Gates[b]

Abstract

Although the Census represents the richest data resource measuring the LGBT population, scholars note serious measurement problems with their same-sex couple calculation. After the publication of Census 2010 data on same-sex couples, the Census Bureau acknowledged this measurement problem and suggested that as many as 40 percent of same-sex couples tabulated in Census 2000 and 28 percent of those tabulated in Census 2010 were likely misclassified different-sex couples (O'Connell and Felix, 2011). As a result, the Bureau released new"preferred" estimates for the number of same-sex couples in both 2000 and 2010. Using these data, we develop new estimates for Census 2000 county-level data on same-sex couples and replicate a 2009 study (McVeigh & Diaz) that included a Census 2000 same-sex couple measure. After replicating regression analyses, substituting our adjusted estimates of same-sex couples for the original tabulations, and comparing differences, our results show that using an adjusted same-sex couple measure increases the magnitude of the effect that the presence of same-sex couples has on county-level voting behavior. These results demonstrate how same-sex couple adjustments might impact findings from previous scholarship, as well as suggests that researchers should use the adjusted data moving forward in order to prevent further measurement error.

Keywords

demography, sexuality, quantitative methods, methodology, public policy

Acknowledgements

We would like to thank Rory McVeigh and Maria-Elena D. Diaz for generously sharing their data for this analysis.

[a] University of California, Los Angeles
[b] The Williams Institute, University of California, Los Angeles School of Law

Corresponding Author:
Rebecca DiBennardo, Doctoral Candidate, Department of Sociology, University of California, Los Angeles, 264 Haines Hall, 375 Portola Plaza, Los Angeles, CA 90095-1551
E-mail: redibennardo@ucla.edu; Phone: 845.519.9410; Fax: 310.206.9838

INTRODUCTION

Sexual orientation constitutes a social location that is important to central sociological questions about health, inequality, and political rights. However, a 2011 Institute of Medicine analysis of the state of research on the health and well-being of the lesbian, gay, bisexual, and transgender (LGBT) population notes the challenges associated with the limited data resources available to study this population. The lack of large population-based samples that include LGBT-identified individuals may explain why an increasing amount of research relies upon U.S. Census Bureau data on same-sex cohabiting couples (e.g., Klawitter and Flatt, 1998; Black, et al., 2000; Black et al., 2002; Jepsen and Jepsen, 2002; Gates and Ost, 2004; Rosenfeld and Kim, 2005; Jepsen and Jepsen, 2006; Black et al., 2007; Schwartz and Graf, 2009; McVeigh and Diaz, 2009; Rosenfeld, 2010; Carpenter and Gates, 2010; Klawitter, 2011). Identification of same-sex couples has been possible in all decennial censuses and the annual American Community Surveys since 1990, and represents the largest and most diverse data resource that measures at least a part of the LGBT population.

In conjunction with this proliferation of research, scholars have also noted serious measurement problems with the Census Bureau's same-sex couple data (Black, et al., 2007; Gates and Steinberger, 2009). These problems stem from a classic false positive scenario whereby a very small fraction of different-sex couples incorrectly designate the sex of one partner, making the couple appear to be same, rather than different, sex. In Census 2000, the Census Bureau began including any same-sex couples who designated a partner as either a "husband/wife" or "unmarried partner" in their tabulations of same-sex unmarried partners. The inclusion of spousal couples (couples who identify one partner as a "husband/wife") in the same-sex tabulations meant that even small errors in the sex designation among different-sex married

couples would create substantial measurement error within the same-sex couple tabulations, since the ratio of different-sex to same-sex couples is approximately a hundred to one.

With the release of Census 2010 data on same-sex couples, the Census Bureau officially acknowledged this measurement problem and suggested that, consistent with the findings of Black, et al. (2007) and Gates and Steinberger (2009), as many as 40 percent of same-sex couples tabulated in Census 2000 and 28 percent of those tabulated in Census 2010 were likely misclassified different-sex couples who had miscoded the sex of one spouse or partner (O'Connell and Felix, 2011). As a result, the Census Bureau released new state-level "preferred" estimates for the number of same-sex couples for both 2000 and 2010. They also released previously unavailable estimates for the error rate of sex misclassification among different-sex married and unmarried couples for each state in each year. These substantially better data provide information that enables researchers to adjust Census data at more detailed geographic levels allowing them to develop more accurate estimates for the number of same-sex couples in various jurisdictions.

This paper does just that: we develop new estimates for Census 2000 county-level data on same-sex couples and demonstrate how these adjustments might impact findings from scholarship that relies upon these data. We focus on a study published in the December 2009 issue of *American Sociological Review* by Rory McVeigh and Maria-Elena D. Diaz. They examine sources of variation in voting on state initiatives between the years 2000 and 2008 proposing to ban marriage rights for same-sex couples. McVeigh and Diaz looked at county-level voting results, theorizing that county-level data allowed them to examine with more precision how the structural features present in communities "affect[ed] the extent to which same-sex marriage [was] perceived to be threatening to (1) community residents' interests and

values and (2) the community as a whole" (2009: 893). This scholarship represents a perfect test of the most basic impact of the measurement issue, because it relies specifically on counts of same-sex couples, which we can now easily adjust.

As part of their analysis, McVeigh and Diaz examine the role that exposure to same-sex couples within communities (using Census 2000 county-level tabulation of same-sex couples) can play in shaping opinions about marriage. In general, they find that higher proportions of same-sex couples in a county predict decreased support for bans on marriage rights for same-sex couples. A variety of reasons could explain this finding: exposure to same-sex couples may increase support for their relationships, the presence of same-sex couples likely proxies a broader acceptance for non-traditional family norms, and the presence of same-sex couples who likely support getting the right to marry mathematically reduces the votes for same-sex marriage bans in a county. However, our analyses focus less on the explanations for the findings and more on showing how improved measurement of same-sex couples could attenuate these and other findings and produce perhaps more robust results.

To that end, we replicate the McVeigh and Diaz regression analyses substituting our adjusted estimates of same-sex couples for their original tabulations of this population based on Census 2000 data. We compare the differences and assess the degree to which the use of more accurate data might impact analyses. Our results show that using an adjusted same-sex couple measure increases the magnitude of the effect that the presence of same-sex couples has on voting behavior in a county.

SAME-SEX COUPLE DATA ADJUSTMENTS

In order to investigate the likelihood that a different-sex couple misclassified itself as a same-sex couple, O'Connell and Felix (2011) analyzed the records of every household in the 2000 and 2010 decennial census tabulations that included a person designated as the "husband/wife" or "unmarried partner" of Person 1, known as the householder. Using internal Census files that show the percent of Census respondents with a given first name that identify as male or female, they matched the first names of all individuals in couples to the likelihood that the name on the Census record was consistent with the sex recorded on the record. If there was less than a five percent chance that the recorded name was consistent with the recorded sex, the record was altered and the sex changed to the sex considered to be more consistent with the name. All couples were analyzed and reclassified in this fashion. In doing so, the analyses report the degree to which different-sex married and unmarried couples likely misclassified the sex of one partner in these Censuses.

O'Connell and Felix (2011) provided state-level figures for the number of reported different-sex married (mar) and unmarried partner (dsump) couples, along with estimates for the number of misclassified same-sex couples (ss_orig). They also provided separate figures for the number of same-sex couples who were likely misclassified as spouses and unmarried partners. For each state, we calculate the ratio of misclassified same-sex spouses to reported married couples (mar_err) and the ratio of misclassified same-sex unmarried partners to different-sex unmarried partners (dsump_err). As such, one can derive the Census "preferred" estimate of total same-sex couples (ss_adj) with the following equation:

Eq. 1: $ss\_adj = ss\_orig - ((mar*mar\_err)+(dsump*dsump\_err))$

To derive county level estimates, we apply Eq. 1 for each county, using the state-level error calculations (mar_err, dsump_err) applied to county-level figures for the number of married (mar) and unmarried (dsump) different-sex couples and the number of same-sex couples (ss_orig). If the resulting adjusted same-sex couple figure is negative, we set it to zero. So that the county figures ultimately sum to the official Census preferred estimates for the number of same-sex couples, we calculate the distribution of adjusted same-sex couples (ss_adj) across all counties and apply that distribution to the official Census preferred state estimate and use these figures as our final "preferred" estimate for the number of same-sex couples in a county.

McVeigh and Diaz's (2009) data set includes voting outcomes from 2,231 counties in 28 states that voted on ballot initiatives banning marriage for same-sex couples between 2000 and 2008, with the dependent variable of interest being percent of votes cast in each county in support of the marriage ban. McVeigh and Diaz include several independent variables related to their theoretical claims. Five of those variables measure traditional family and gender roles: percent of women not in labor force, occupational sex segregation, same-sex households, households married with children, and heterosexual couples cohabitating. They root the construction of these variables in the notion that more traditional communities will show lower labor force participation by women and higher rates of marriage (2009: 896). They also include a large number of control variables that account for possible economic, educational, racial, political, and other differences between counties.

Replication of Analyses

Using the original data, courtesy of McVeigh and Diaz, we were able to exactly replicate the 2009 study's results. We then replaced their county same-sex couple data with our adjusted figures and replicated each model included in their paper.

RESULTS

Table 1 shows comparisons between the original McVeigh and Diaz analysis and our analysis using adjusted same-sex couple data. The first two columns show original and adjusted models without state fixed effects, respectively, while the second two columns show original and adjusted results with state fixed effects. We focus specifically on the effect that the adjustment has on the impact that the percentage of same-sex couples (among households) in a county has on the likelihood that a county will support a ban on marriage for same-sex couples. In both the original and adjusted analysis, the independent variables explain over 70 percent of voting variation in models without fixed effects, and over 90 percent of variation in models with state fixed effects.

Table 1: Comparison Table, Original and Adjusted (Adj.) Same-Sex Couple Measures, OLS Regression of Percent Approving Same-Sex Marriage Ban with and without State Fixed Effects, U.S. Counties 2000-2008

Independent Variable	**Original Model 1**		**Adj. Model 1**		**Original Model 2†**		**Adj. Model 2†**	
Percent Same-Sex Households	**-2.038**		**-3.419**	**	**-5.428**	***	**-6.191**	***
Number of Observations	2231		2231		2231		2231	
R-Square	0.713		0.714		0.910		0.910	
*p<.05; **p<.01; ***p<.001, † State Fixed Effects Model								

The magnitude and significance of the same-sex couple coefficient changes in both models using the adjusted measure compared to the analysis using the original Census measure. In contrast to the Census measure in Model 1, the adjusted same-sex couple variable in Model 1 shows significance (at p< .001). In addition, the magnitude of the effect of the same-sex variable increases in both models, though much more modestly in the model using fixed effects. The

adjusted analysis also shows smaller robust standard errors in all models. Coefficients of other covariates are not shown in Table 1, but apart from the same-sex couple variable, the coefficients in the fixed and non-fixed effects models of the adjusted regression are almost identical in magnitude and significance aside from a few negligible differences.

If the adjusted measurement yields a more accurate estimate of same-sex couples in a county, and we believe that the presence of same-sex couples would likely reduce voting in support of marriage bans for same-sex couples, then we would expect the attenuation to yield larger coefficients on the adjusted same-sex couple variable in both models, which turns out to be the case. That the robust standard errors remain more consistent between fixed and non-fixed effects models in the adjusted regression suggests that they are likely more precise (Allison, 2005). In addition, the standard errors are lower overall than in the original analysis, which further supports the new model's precision.

In the original McVeigh and Diaz estimation, the fixed-effect model altered the magnitude of the coefficient on the same-sex couple variable substantially. It increased by a factor of 1.66. Using the adjusted data, the change in the coefficient between the two models was more modest, increasing by a factor of about 0.81. The more modest impact of adding the fixed effects could actually be the result of the original model essentially adjusting for cross-state variation in the measurement error of same-sex couples. Fixed effects account for time-invariant, state-level differences across the entire sample. Allison (2005) points out that, while in non-fixed effects models, unobserved variables are assumed to be uncorrelated with observed variables (thus forming the basis for OLS regression's "Best Linear Unbiased Estimate" notion), this is not the case for fixed effects models. He states, "In a fixed effects model, the unobserved variables are allowed to have any associations whatever with the observed variables…unless you allow for

such associations, you haven't really controlled for the effects of the unobserved variables. *This is what makes the fixed effects approach so attractive*" (2005: 3, our emphasis).

Thus, it is likely that the modest impact of the attenuation of the same-sex couple variable in the fixed-effect model reflects systematic measurement error in data collection regarding the percent of same-sex couples in these counties (Allison, 2005: 9). O'Connell and Felix (2011) demonstrate clear variation in the level of error among same-sex couples across states, with error rates in Census 2000 among different-sex married couples estimated to be as low as 0.28% in Mississippi and as high as 0.53% in Wisconsin. It appears as if the fixed effects in McVeigh and Diaz's model literally controlled for some of the measurement error in the Census same-sex couple variable. This measurement issue clearly affects the original fixed effects model: the same-sex coefficient increases from a 5.4 percent decrease in the percent supporting marriage bans for same-sex couples to a 6.2 percent decrease in the adjusted model. In other words, the estimated impact of same-sex couples in a county is 15 percent higher using the adjusted data.

In order to further test whether the adjusted model better "fits" the data, Table 2 presents the results from two Bayesian Information Criterion (*BIC*) tests comparing Model 1 (columns 1 and 2 of Table 1) and Model 2 (columns 3 and 4 of Table 1) from the original and adjusted analyses. *BIC* tests predict the likelihood of the best statistical model given the data at hand (Raftery, 1995; Hoeting et al., 1999). The *BIC* statistic for adjusted Model 1 (5.155) indicates positive evidence that this model is a better "fit" than the model using McVeigh and Diaz's original measure. The *BIC* statistic for adjusted Model 2 (8.431) indicates even stronger evidence that this model is a better "fit."

Table 2: Goodness of Fit Statistics Comparing Models of the Relationship Between Voting for Same-Sex Marriage Bans using Original Census Same-Sex Couple Variable versus Adjusted Census Same-Sex Couple Variable, US 2000-2008; N = 2231

	BIC	*R2*	F-test d.f.	Prob>F
Model without Fixed Effects				
Original Census Measure (M + D Analysis)	-2594.662	0.713		
Adjusted Census Measure	-2599.817	0.714		
Contrasts				
Original vs. Adjusted	-5.155	0.001	25; 297	0.000
Fixed Effects Model				
Original Census Measure (M + D Analysis)	-4958.483	0.909		
Adjusted Census Measure	-4966.914	0.910		
Contrasts				
Original vs. Adjusted	-8.431		52; 297	0.000

*See Table 1 for models used

DISCUSSION

In their article examining sources of variation in voting on marriage bans for same-sex couples, Rory McVeigh and Maria-Elena D. Diaz McVeigh argue that community context, specifically traditional gender roles and family structures within communities, played an integral role in shaping voting on such bans between the years 2000 and 2008. According to the authors, the presence of traditional sex roles reinforced both general homophobia and diminished contact with gays and lesbians, ultimately resulting in "intolerant" voting behavior (McVeigh and Diaz, 2009: 894).

Analysis of their findings using same-sex couple data that are adjusted for measurement error suggests that McVeigh and Diaz likely underestimated the impact of how possible interactions with same-sex couples affects voting behavior. Through the use of an improved same-sex couple measure, we show that the presence of same-sex couples within communities

leads to a significant decrease in voting for bans on marriage for same-sex couples in both fixed and non-fixed effects statistical models and has at least a 15 percent larger impact than McVeigh and Diaz's fixed effects model initially indicated. Our results demonstrate how same-sex couple adjustments might impact findings from previous scholarship, as well as the need for researchers to use the adjusted data moving forward in order to prevent further measurement error. In addition, we suggest that future survey design and research take steps to acknowledge both the complicated legal relationship schemas faced by same-sex families, as well as the diversity of LGBT family forms and relationship categories. This will require new and creative thinking about the best ways in which to understand evolving family structures.

REFERENCES

Allison, P. D. *Fixed Effects Regression Methods for Longitudinal Data Using SAS*. SAS Publishing, 2005.

Black, D. A, S. G Sanders, and L. J Taylor. "The Economics of Lesbian and Gay Families." *The Journal of Economic Perspectives* 21, no. 2 (2007): 53–70.

Black, D., G. Gates, S. Sanders, and L. Taylor. "Demographics of the Gay and Lesbian Population in the United States: Evidence from Available Systematic Data Sources." *Demography* 37, no. 2 (2000): 139–154.

———. "The Measurement of Same-sex Unmarried Partner Couples in the 2000 US Census" (2007).

———. "Why Do Gay Men Live in San Francisco?" *Journal of Urban Economics* 51, no. 1 (2002): 54–76.

Carpenter, C., and G. J Gates. "Gay and Lesbian Partnership: Evidence from California." *Demography* 45, no. 3 (2008): 573–590.

Gates, G. J, and J. Ost. *Gay and Lesbian Atlas*. Urban Inst Pr, 2004.

Gates, G. J, and M. Steinberger. “Same-Sex Unmarried Partner Couples in the American Community Survey: The Role of Misreporting, Miscoding and Misallocation.” In *Annual Meetings of the Population Association of America, Detroit, MI*, 2009.

Hoeting, J. A, D. Madigan, A. E Raftery, and C. T Volinsky. “Bayesian Model Averaging: a Tutorial.” *Statistical Science* (1999): 382–401.

Jepsen, C. A, and L. K Jepsen. “The Sexual Division of Labor Within Households: Comparisons of Couples to Roommates.” *Eastern Economic Journal* 32, no. 2 (2006): 299–312.

Jepsen, L. K, and C. A Jepsen. “An Empirical Analysis of the Matching Patterns of Same-sex and Opposite-sex Couples.” *Demography* 39, no. 3 (2002): 435–453.

Klawitter, M. M, and V. Flatt. “The Effects of State and Local Antidiscrimination Policies on Earnings for Gays and Lesbians.” *Journal of Policy Analysis and Management* 17, no. 4 (1998): 658–686.

Klawitter, M. “Multilevel Analysis of the Effects of Antidiscrimination Policies on Earnings by Sexual Orientation.” *Journal of Policy Analysis and Management* 30, no. 2 (2011): 334–358.

McVeigh, R., and D. D Maria-Elena. “Voting to Ban Same-sex Marriage: Interests, Values, and Communities.” *American Sociological Review* 74, no. 6 (2009): 891–915.

O’Connell, M., & Felix, S. (2011). *Same-sex Couple Household Statistics from the 2010 Census*: Fertility and Family Statistics Branch, Social, Economic and Housing Statistics Division, U.S. Bureau of the Census

Raftery, A. E. “Bayesian Model Selection in Social Research.” *Sociological Methodology* 25 (1995): 111–164.

Rosenfeld, M. J, and B. S Kim. “The Independence of Young Adults and the Rise of Interracial and Same-sex Unions.” *American Sociological Review* 70, no. 4 (2005): 541–562.

Rosenfeld, M. J. “Nontraditional Families and Childhood Progress Through School.” *Demography* 47, no. 3 (2010): 755–775.