Models and Marginals: Using Survey Evidence to Study Social Networks

Miller McPherson
Duke University

Lynn Smith-Lovin
Duke University

Matthew E. Brashears
Cornell University

Fischer (2009) argues that our estimates of confidant network size in the 2004 General Social Survey (GSS), and therefore the trend in confidant network size from 1985 to 2004, are implausible because they are (1) inconsistent with other data and (2) contain internal anomalies that call the data into question. In this note, we assess the evidence for a decrease in confidant network size from 1985 to 2004 in the GSS data. We conclude that any plausible modeling of the data shows a decided trend downward in confidant network size from 1985 to 2004. The features that Fischer calls anomalies are exactly the characteristics described by our models (Table 5) in the original article.

We are grateful to Professor Fischer and the *ASR* editors for the opportunity to revisit our 2006 article on social isolation (McPherson, Smith-Lovin, and Brashears 2006). We see two central themes in his comment: (1) there are over-reports of social isolation in 2004 and (2) the General Social Survey (GSS) data do not support the claim that confidant networks changed significantly from 1985 to 2004. We strongly agree with Fischer's (2009) first claim. We make exactly that point in the abstract of our 2006 article.[1] We disagree, however, that these over-reports are confined to the 2004 data or that they are random. We use this opportunity to elaborate our analysis of the reports of social isolation.

We will show that Fischer's second (and most important) conclusion about the lack of a trend in social connectedness is extremely unlikely. We disagree with Fischer that the unweighted[2] cross-tabulation analyses that he presents are an appropriate approach to analysis, given the strong effects of cooperativeness and fatigue that we identified in our original Table 5. Such analyses are misleading because of omitted variable bias, among other problems. We review

We would like to thank Mark Chaves, Peter Marsden, and the Network Study Group at Duke University for useful suggestions. We also thank Jeff A. Smith for independently replicating our analyses. We, of course, remain responsible for our conclusions here. The data collection of network measures in the 2004 GSS was supported by National Science Foundation grant SES 0347699 to the first and second authors and a grant from CIRCLE to Tom W. Smith. The current analysis was supported by NSF grant BCS 0527671 to the first and second authors.

[1] "The data may overestimate the number of social isolates . . . " (McPherson et al. 2006:353).

[2] We do not emphasize the issue of weighting in this note because it does not affect the main substantive findings under dispute. We note, however, that without properly weighting for the complex sample design of the 2004 survey, Fischer's percentages refer only to the (two distinct) populations of respondents sampled in 2004 and are not representative of the non-institutionalized adult population of the United States (the population the GSS is meant to represent). All analyses of these data intended to reflect the general population must be weighted.

his findings to highlight these problems. We dispute his statement that the GSS trend data are inconsistent with other estimates and that no plausible social change could have produced a strong trend in networks. We conclude with thoughts about the perils of public sociology and the value of public data.

HISTORY AND THEORY

We began reviewing the 2004 data as soon as they were collected, because we initiated the NSF grant that supported the replication of the 1985 GSS network module. We immediately contacted the GSS to tell them there were too many reports of zero confidants in "Numgiven," the variable that codes the number of confidants in the 2004 data. Although we now know there were 41 miscoded cases (McPherson, Smith-Lovin, and Brashears 2008a), they found no problem at the time.

We modeled the over-reporting of zeros in the original article by controlling for artifacts of uncooperativeness and fatigue (see McPherson et al. 2006: Table 5). Because Fischer concentrates on social isolation—the reports of zero confidants—we use this opportunity to focus on the process through which zeros in particular might be over-reported in the data. We believe that there is actually a mixture of two processes in the data: a Poisson process of acquisition and loss of confidants as described below, and a binary mechanism (zero-inflation) that affects whether a respondent is coded into the zero category as a result of some independent process (e.g., respondent fatigue, interviewer effects, technical glitches in transcription, lack of rapport with the interviewer, or some substantive mechanism). We model these two processes explicitly with a zero-inflated Poisson analysis.[3]

How did we know immediately that there were too many zeros in Numgiven, and why do we now use an inflated Poisson model to model the data? Underlying any cross-sectional data like the GSS is a dynamic process, for which the cross-sectional measure is a snapshot at a single point in time. For the Numgiven variable, this process consists of discrete counts, which can be captured with a simple stochastic model. A person's confidants come from the very large number of potential partners in society, but one loses confidants from the relatively small set currently possessed. Under these conditions, the cross-sectional distribution of confidants will follow the Poisson distribution,[4] with cross-sectional mean equal to the ratio of the rate of gain to the rate of loss of confidants (for a derivation of this result, see McPherson 1981, forthcoming). When the rates of gain and loss depend on the social positions of the actors involved, we model this process as heterogeneous Poisson, which allows us to take the sources of variability in those rates into account.

No simple Poisson process will generate the number of reported social isolates in the 2004 data, given the shape of the rest of the distribution, so we knew that the 2004 Numgiven zeros were inflated. In the course of reinterviewing some of the 2004 respondents, we discovered more evidence strongly suggesting there were misreported zeros for Numgiven. An intensive search by NORC discovered the 41 miscoded cases shortly after we reported this fact to the GSS.

Fischer's Core Claim: No Trend in Social Connectedness

The GSS data, even under the most conservative assumption that *all* of the zeros in excess of the Poisson process are artifactual, are inconsistent with Fischer's core claim that there was no change in confidant networks from 1985 to 2004. The zero-inflated Poisson analysis in Table 1 shows the effects of our independent variables on both the heterogeneous Poisson model and the binary process of zero-inflation.[5] The zero-inflated model assumes there are two possible reasons for an observation to have a value of zero: (1) a Poisson count process, in

[3] The zero-inflated Poisson model is available under the zeromodel option in SAS, zip in STATA, and zeroinfl() in R.

[4] Readers familiar with the literature should note that this derivation of the cross-sectional distribution of Numgiven is very different from the standard Erdos-Renyi null model; we explicitly model the gain and loss of network ties as a stochastic process, while the Erdos-Renyi approach randomly assigns network ties in a static network.

[5] Our 2006 article used the negative binomial model, which adds an additional parameter for heterogeneity for the original analyses because of overdispersion in Numgiven. Analyses of the now-corrected data set show that the Poisson is the preferred model after the explanatory variables are taken into account.

which parameters govern all of the values of the variable and (2) a binomial process, in which parameters govern an additional probability of the zero category versus all else. The parameters of both processes are estimated simultaneously, so that each process acknowledges the effects of the other. The predictors for the binomial process model the probability that a case will be an "inflated" zero, taking into account how many zeros there "should" be according to the entire Poisson distribution of Numgiven, while the Poisson estimates take the zero inflation process into account.

Table 1 presents an analysis that adjusts for zero-inflation in both 1985 and 2004. The change in Numgiven from 1985 to 2004 is documented by the Wave coefficient in the second panel of Table 1, which models the dependence of mean Numgiven on our independent variables. The highly significant negative coefficient for Wave (1 = 2004) implies a mean decrease from 1985 to 2004 in Numgiven of around one confidant, on average, taking into account the excess zeros, the known threats to validity, and the substantive effects of sociodemographic variables. Figure 1 shows the estimated mean difference in Numgiven between 1985 and 2004 plotted across years of education.

This (unavoidably busy) figure contains a wealth of information. Each of the 2,957 respondents is indicated by either a larger square (1985) or a smaller dot (2004); the fitted mean Numgiven is the solid line for 1985 and the dashed line for 2004. The points have been jit-

Table 1. Zero-Inflated Poisson Model of the Number of Confidants (Numgiven) (using weighted data from the 1985 and 2004 General Social Surveys)

Independent Variable	Coefficient	Standard Error	Significance
Zero-Inflation Model Coefficients (binomial with logit link)			
Constant	–3.93	.67	.000
Wave (1 = 2004)	2.09	.37	.000
Cooperative [a]	.34	.41	.411
Restless/impatient [a]	1.94	.37	.000
Hostile[a]	2.27	1.62	.162
Number of missing	.50	.13	.000
Years of education	–.05	.03	.136
Female	–.19	.21	.371
Age	.01	.01	.003
Married	–.44	.22	.045
Black[b]	1.24	.24	.000
Other race [b]	.13	.50	.796
Count Model Coefficients (Poisson with log link)			
Constant	.59	.07	.000
Wave (1 = 2004)	–.30	.03	.000
Cooperative [a]	–.19	.04	.000
Restless/impatient [a]	–.21	.10	.042
Hostile [a]	–.48	.36	.182
Number of missing	–.11	.04	.006
Years of education	.05	.00	.000
Female	.05	.03	.045
Age	–.00	.00	.014
Married	–.02	.03	.494
Black [b]	–.11	.05	.024
Other race [b]	–.22	.06	.001
Log-likelihood: –5.37e+03 on 2 Degrees of Freedom			

Note: We suppress statistical interactions between Wave and Education and Wave and Age for clarity in this table, but we take them into account in detailed analyses where appropriate.

[a] Relative to respondent coded "friendly and interested" by the interviewer.

[b] Relative to white respondents.

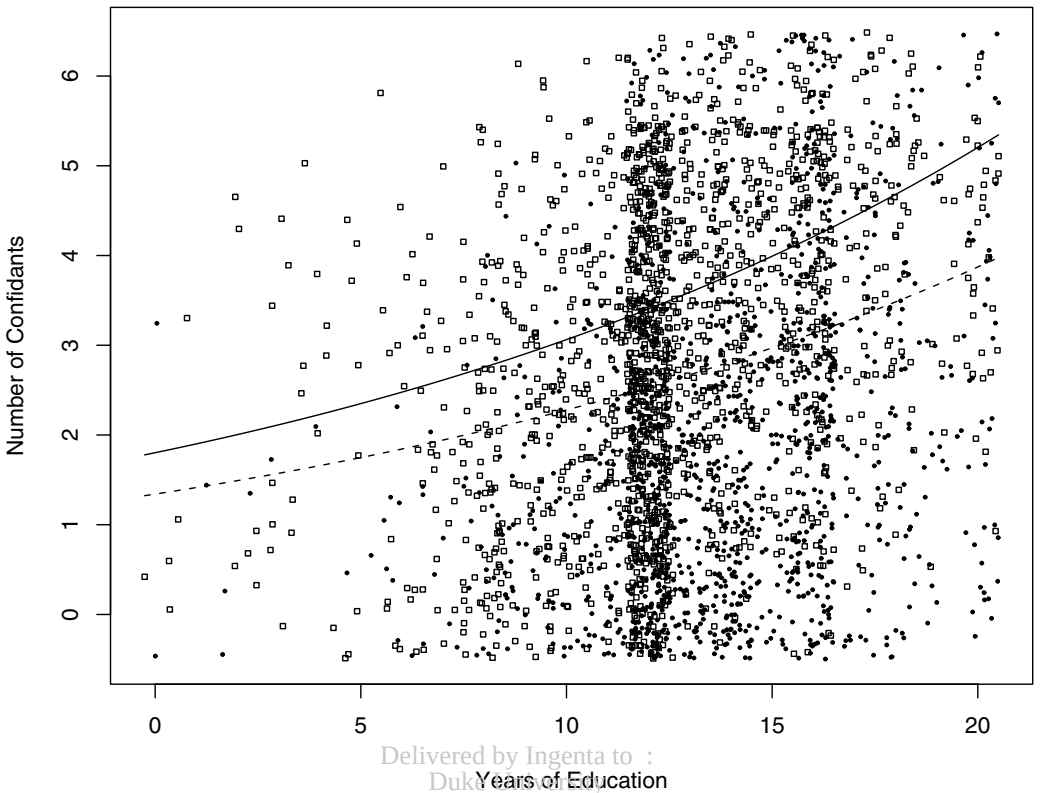

Figure 1. The Relationship between Education and Number of Confidants, 1985 and 2004

Legend: Squares indicate 1985 respondents, dots indicate 2004 respondents. The solid line is fitted mean for 1985, the dashed line is fitted mean for 2004.

Source: General Social Survey.

Note: Fitted means from zero-inflated Poisson regression model controls for presence of inflated zeros, fatigue, cooperation, age, gender, marital status, and race, as in Table 1.

tered to reveal their density in the scattergram. We first point out the need for humility in interpreting our model (and by implication, any summary of the data in the form of simple cross-tabulations); clearly, there is a great deal of variation in the data not explained by the model. However, close inspection reveals an unambiguous tendency for the 2004 values of Numgiven to inhabit the lower range, while the 1985 values are higher, on average. Note, for instance, the sparseness of 2004 observations in the upper-left quadrant in comparison with the lower-right quadrant. This impression is aptly summarized by the fitted curves for mean Numgiven, which are strongly different for 1985 and 2004 ($p < .000001$). It is important to remember that the fitted curves are adjusted for the fact that there are more inflated zeros in 2004 than in 1985 (see below), and for all the variables described in Table 1.

Our best model for Numgiven, controlling for the threats to validity posed by fatigue and cooperation; the known sources of variation due to years of education, age, marital status, and race; and taking into account the inflated number of zeros in the 1985 and 2004 data, shows a substantively and statistically significant difference between 1985 and 2004 in the GSS data.

As the table and figure conclusively show, the differences between 2004 and 1985 are extremely unlikely (less than one chance in a trillion) to have been due to sampling error, since the fitted means of network size are very different, taking into account all the available control variables. The trend is significant, even under the most conservative assumption that none of

the reported social isolation in excess of the Poisson process of gain and loss is substantively meaningful.

Our analysis rules out Fischer's (2009) simulated random mechanism, since the excess zeros in 2004 are not sufficient to have produced the 1985 to 2004 difference. In fact, one can easily demonstrate that there are major differences between 1985 and 2004 in Numgiven by simply throwing out all the zeros for both years. (One can either do this literally, or model it with zero-truncated count analysis. Both approaches lead one to conclude that there are substantial differences in Numgiven between 1985 and 2004 whether or not there are too many zeros.) If the 1985 to 2004 change were due to a process in which some positive counts were randomly coded as zero, as Fischer suggests, then ignoring the zeros would destroy the mean difference in number of confidants from 1985 to 2004. Since that mean difference remains great after eliminating the zeros, it is clear that the data show strong evidence for change.

As we repeatedly point out in the original article, we are unable to find any combination of variables that destroy the difference between 1985 and 2004 in the number of confidants (McPherson et al. 2006:367–71). We are forced to conclude that the data show a decline in social connectedness from 1985 to 2004. As Fischer's many cross-tabulations and our original Table 5 reveal, this decline does not depend on any variable that either he or we have been able to discover in the available data.

Fischer suggests (as does our original abstract) that there are too many zeros in the 2004 data. We show below that this is almost certainly true. He ignores the possibility of inflated zeros in 1985; we correct this omission. He posits a purely random mechanism for the inflation of zeros, but we show that his random mechanism can easily be ruled out.

TOO MANY ZEROS?

The coefficient for Wave, the dummy variable representing the change in inflated zeros from 1985 to 2004, appears in the second row of the first panel of Table 1. This significant positive coefficient implies that there are more inflated zeros in 2004 than in 1985. Because this coefficient is a logistic regression estimate, the value of 2.09 implies that the odds of an inflated zero in 2004 (i.e., a zero in excess of the Poisson-predicted zeros) are more than seven times the odds of such an event in 1985. This estimate takes into account both the main effects of the independent variables from our original Table 5 on the Poisson process, and the effects of those variables on the probability of zero inflation.

Using the results of Table 1, we can go much further than asserting that there are inflated zeros in 2004. We can (1) estimate the effects of known threats to the validity of the items due to fatigue and non-cooperation on the probability of inflated zeros, (2) estimate the effects of known substantively relevant variables such as age, marriage, and education on the probability of inflated zeros, (3) estimate the number and proportion of such zeros in both 1985 and 2004, taking into account the above effects, and (4) describe the characteristics of individuals who are likely to be coded as inflated zeros. We take on the last two of these tasks first.

Although there are several ways to estimate the number of inflated zeros in each year, we use an informal Bayesian approach to the posterior distribution of Numgiven (cf. Gelman and Hill 2006). We estimate that there are 42 excess zeros in 1985 (with a 95 percent credible interval from 17 to 76), and 208 excess zeros in 2004 (95 percent interval from 171 to 244). Our preferred model thus projects that the 2004 data have roughly 166 more inappropriate zeros than the 1985 data, with a very high degree of confidence that 2004 has more than 1985. Put qualitatively, we are pretty sure that there are inflated zeros in both 1985 and 2004, and we are pretty sure that there are more in 2004 than in 1985, but the best estimate for the number of such zeros has a substantial amount of uncertainty.

The awareness of this uncertainty is one reason to be skeptical of Fischer's claim that no change has occurred, based on his strong assumption that there are exactly 200 randomly generated excess zeros in 2004 and exactly none in 1985. Taking the most conservative estimates of non-inflated zeros at face value, we would still be left with roughly a 70 percent increase in social isolation from 1985 to 2004. In 1985 there are 136 reported zeros, of which we estimate 42 are inflated, leaving 92 out of 1,531 cases for a proportion isolated of .06. In

2004 there are 356 reported zeros, of which we estimate 208 are inflated, leaving 148 isolates out of 1,428 respondents, for a proportion isolated of .10. The ratio of .10 to .6 is 1.7, suggesting a 70 percent increase in non-inflated zeros. Once again, we need to emphasize that we are removing the influence of *all* inflated zeros from this comparison, leading to an extremely conservative estimate of the increase in social isolation.[6]

There are many other reasons to be skeptical of Fischer's analysis. Turning again to Table 1, we see that the excess zeros, rather than being randomly distributed across social categories, as Fischer asserts, are systematically related to our measures of cooperativeness, fatigue, age, marital status, and race. Black respondents appear to have more inflated zero responses than do whites, as do older respondents, those with missing items preceding the Numgiven item, and those rated less cooperative by the interviewer. Married respondents are marginally less likely to give an inflated answer of zero. While it is beyond the scope of this note, it is possible to use the predicted probability of zero inflation for each respondent to search for coding problems or other patterns of inflation (e.g., subtle uncooperativeness or satisficing) in the data.

To sharpen the issues of agreement and disagreement with Fischer to this point, we agree that there are too many zeros in the 2004 data (as our original abstract says), we disagree that we should assume there were no such cases in 1985, and we disagree that a simulation assuming purely random error is the way to approach this question. We argue for a model-based approach to assess the substantive and artifactual variables that influence both social connectedness and potentially inflated reports of social isolation. Fischer's cross-tabulations ignore variables that we know influence reports of Numgiven, resulting in omitted variable bias. We now turn to these problems in more detail.

ARTIFACTS IN THE DATA

The coefficients for our fatigue variable (Number of Missing Values) and the cooperativeness dummy variables of Cooperative, Restless/Impatient, and Hostile (compared with Friendly/Interested) in our original Table 5 display major effects of these threats to validity that must be taken into account. These effects' very large size means that any analysis that excludes them will lead to biased and inconsistent results. Cross-tabulations such as those by Fischer, which produce estimates of the percent isolated not taking these artifacts into account, will be misleading because omitted variables will confound the analysis. The Missing Values variable (the count of the number of missing items on the 10 questions preceding the Numgiven variable) has a highly significant coefficient of .372 in our original Table 5, which means that the odds of reporting social isolation increase roughly 50 percent for each additional missing item in the preceding 10 items (exp(.372) = 1.5). Because the observed range of this variable is 0 to 10, one only has to exponentiate 3.72 (10 times .372) to see that the odds for reporting social isolation for someone with 10 missing items are more than 40 times the odds for someone with no missing items. Of course, there are few cases with many missing preceding items, but an approach that relies on simple cross-tabulation will not be able to tell where those extreme cases will be in the high dimensional multivariate space created by considering many independent variables simultaneously.[7] Because many of Fischer's results involve a very small number of cases (see our discussion below), large percentage shifts will occur with small changes in the independent variables.

Another way of illustrating the artifacts uncovered in our original article is to compare the fitted probabilities of social isolation for a typical case with no missing items to one with 10 missing items, as in Figure 2. The bottom two curves show the results of our original Table 5 for a representative individual with no missing items (2004 and 1985); the top two curves show the results for such a person with 10 missing

[6] A roughly comparable comparison derived from our original Table 5, column 3, produces .04 isolated in 1985 and .13 in 2004.

[7] About 10 percent of the respondents have one or more missing items. The means of the Missing Values variable are .077 for 1985 and .223 for 2004.

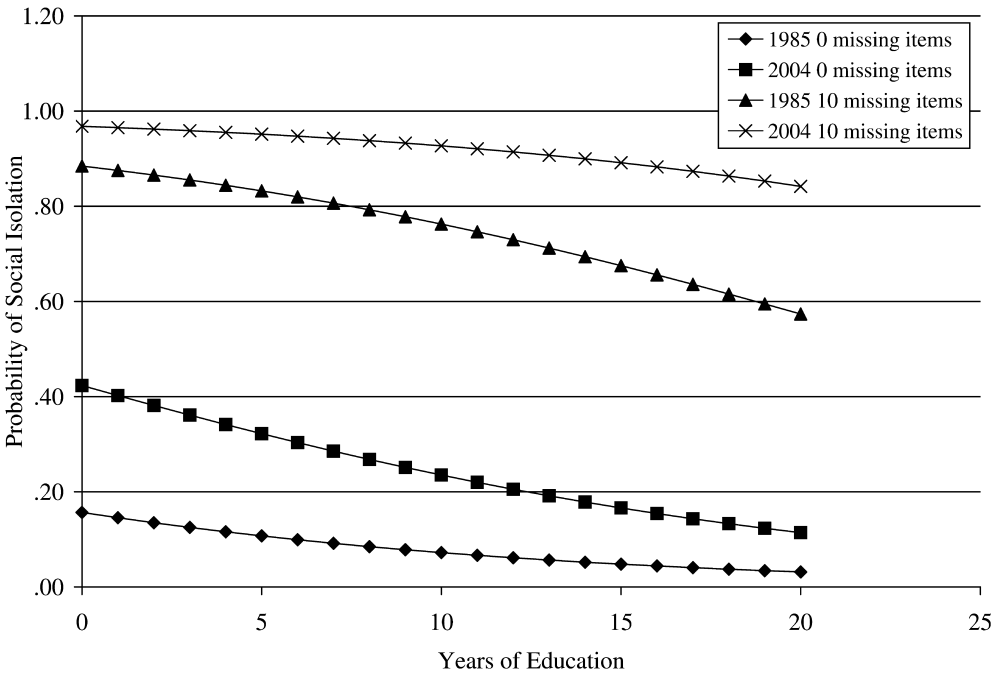

Figure 2. Fitted Probability of Social Isolation from Model IV (2006 Article)

items. The effects of the fatigue variable alone could have changed the apparent amount of social isolation in some of Fischer's cross-tabulations by over 80 percent. Our original analysis takes this striking artifact into account, while Fischer's does not. (The Missing Values variable we used in the 2006 article to represent fatigue was highly predictive of the subsequently discovered 41 miscoded cases. This fact is why our corrected tables published in the December issue of *ASR* [McPherson et al. 2008a] show so little change in our parameter estimates of the 1985 to 2004 change.)

In summary, Fischer's claim that we do not identify artifacts in the GSS data is based on a profound misunderstanding of the results in our original Table 5. The essence of our disagreement is whether or not to take into account the omitted variable bias due to the artifacts that we demonstrated in Table 5. As Figure 2 demonstrates, these effects alone could easily have caused the observed differences in reported social isolation from 1985 to 2004. But as the bottom panel of Table 1 shows, *they did not.*

FISCHER'S "ANOMALIES"

Relationships to Other Sociation Measures

Fischer's first table, showing the proportion of people who report zero confidants in 1985, 1987, and 2004, seems to show that 1987 is not consistent with the trend downward in confidants. We avoid extensive analysis of the 1987 data because scholars generally agree that these data are not comparable: the 1987 question was somewhat different, and it is clear that the fact that further information on only three alters was collected influenced interviewers' motivation to probe for more alters.[8] The mean values of Numgiven in 1985, 1987, and 2004 (3.03, 2.56, and 2.12, respectively), however, clearly illustrate our general point that one must pay attention to all of the data, not just a single cell or contrast. While the proportion responding zero in 1987 is small, the mean value of Numgiven for 1987 is even closer to the 2004 value than a simple linear trend would suggest.

[8] An extensive analysis by Bruce Straits (personal communication) confirms this fact.

In an analysis of convergent validity, Fischer relates Numgiven to other social contact variables. He finds instances in which Numgiven and other social contact variables have patterns that seem implausible to him. We point out that these other measures are of very different types of social contact. We know that the GSS question about discussing important matters in the past six months does not capture all kinds of social contacts. Instead, it tends to elicit extremely close ties (Bailey and Marsden 1999). One may discuss important matters with a 4-year-old child or an investment banker, but such contacts are seldom included in response to the Numgiven question. Instead, GSS respondents tend to mention confidants with whom they interact frequently, whom they have known for long periods of time, with whom they disclose much, and whose opinion they respect.

The other measures of sociation that Fischer uses to establish (lack of) convergent validity are quite different from Numgiven. The question about social contact (Numcntnt: "Not counting people at work or family at home, about how many other friends or relatives do you keep in contact with at least once a year?") occurred in the context of a long module about the use of Internet technology that would prime a cognitive search for people with whom one had contact over that technology. The period of "at least once a year" is broader than "in the last six months" (used in Numgiven). Numcntnt has a range of up to 500, a mean of 27.9, a standard deviation of 43.3, and significant clumping at 10, 20, 50, and the 100's. Clearly, respondents are estimating rather than thinking of specific alters (as they are required to do in Numgiven).

Another measure of sociation Fischer uses is the sum of yes/no answers to a list of 16 types of voluntary associations (Memnum). This is even less similar to Numgiven: people report memberships in groups that never meet face-to-face or in which they have no close ties.[9] Contrary to Fischer's assertions, the other sociation measures have very similar relationships with Numgiven in 1985/1987 and 2004. The Pearson's correlation of contacts (Numcntnt, logged to reduce the effects of outliers) and confidants (Numgiven) in 2004 is .17 ($p < .001$). The correlation between number of voluntary association membership types (Memnum) and confidants (Numgiven) is .23 in 2004, almost identical to the .22 relationship in 1987 (both $p < .001$).

Obviously, firmer statements about measurement of sociation and networks require more data. We can be fairly certain, however, that (1) all of these variables have measurement error in them, (2) they measure different types of social activity, and (3) the number of confidants in 1985 and 2004 is significantly related to other measures of sociation. There is no indication that 2004 Numgiven is less related to other measures of sociation than are relationships among other similar variables in other years or other data sets.

MARRIAGE AND EDUCATION: HOW CAN THESE PEOPLE BE ISOLATED?

Fischer spends significant effort discussing the relationships between social isolation, marital status, and education (his Tables 5, 6, and 7). We believe that his analysis of education is a good illustration of the pitfalls of tabular analysis of bivariate relationships. Fischer focuses on one cell of his Table 5, the respondents with postgraduate degrees. We apply our zero-inflated Poisson model (Table 1) to predict zero inflation in that cell. Given a married, friendly/interested, male respondent with 18 years of education and no missing values on preceding variables, our model generates predicted values of 2 percent inflation of zero responses in 1985 and 29 percent inflation of values in 2004. Fischer is surprised that there are so many zeros in this latter cell. Our model not only explains why there are so many zeros here but adjusts for this effect in the other estimates in the model.

To illustrate this fact, consider that the fitted mean number of confidants for these highly educated respondents, after the fatigue artifacts and the overinflation of zeros are taken into account, is 4.4 in 1985 and 3.2 in 2004. There is thus a strong trend downward in the number of confidants for the individuals in those cells, even though the marginal numbers studied by Fischer seem confusing to him. Our models

[9] Some dynamic evidence for the relationship between Numgiven and Memnum variables appears in McPherson, Popielarz, and Drobnic (1992), which finds that weaker confidant ties are more predictive of changes in memberships than are the very strongest ties.

and Fischer's tables both reinforce the main point of our 2006 article: the trend across these two time points (1985 and 2004) is significant and applies to the major subgroups that we use to divide the population.

Fischer's analysis of marital status (his Tables 6 and 7) shows the same problems of focusing on single cells of a table, rather than controlling for other features that might have changed during the period, and using statistical tests to tell whether the patterns are chance variations or statistically significant. Fischer (2009:663) claims that "the differences among marital categories essentially wash out in 2004." He is arguing here that there is a statistical interaction such that marital status has an effect in 1985 but not in 2004. In fact, a zero-inflated Poisson analysis with just the two independent variables Marital Status and Wave *does* show an interaction effect between these two variables in their effect on Numgiven. When we control for the other variables of our original Table 5, however, this statistical interaction is not significant. Once again, a model that takes multivariate effects into account leads to substantively different conclusions.

We also note that some of the distinctions Fischer stresses in his comment are based on a very small number of cases, a fact that is largely hidden because he reports percentages without reporting the corresponding numbers of respondents. For example, the 15-fold increase in the postgraduate respondents who gave no names in Fischer's Table 5 is the result of a difference in 22 cases between 1985 and 2004, out of the 2,967 respondents who answered the question in those two years.[10]

We end our discussion of Fischer's analysis of marital status and confidants with two more subtle points. First, while the average number of confidants has decreased, even taking into account the inflated zeros, the relationships among variables have remained remarkably consistent. A careful examination of Fischer's Table 7 shows that among married respondents who named at least one confidant, the proportion who do *not* list a spouse is relatively stable between 1985 and 2004 (27.8 and 25.3 percent, respectively). This pattern strongly suggests (as we demonstrate above in the zero-inflated Poisson model) that the inflation of zeros that we mention in our original abstract is the primary measurement issue in the data; the rest of the data structure remains similar. We make a similar point in our discussion of Tables 3 and 4 in our original article.

In an attempt to call all the 2004 data into question, Fischer (2009:664) argues that married respondents are "a category of people who were living with a confidant." The reader should note that all four of the authors here (including Fischer) are members of academic couples who work in the same department as their spouses. It is difficult for people like us to imagine not reporting a spouse as someone with whom we "discuss important matters." But notice that in Fischer's own Table 7, roughly a fourth of all married people who gave an answer other than zero in response to the Numgiven question did not name a spouse. Furthermore, 198 respondents in 1985 and 100 respondents in 2004 named their spouse only *after* first naming someone else as a confidant. Their spouse was not the first person who came to mind. One of the valuable things about representative survey data is that they help us transcend our egocentric view of the social structure created by strong homophily in social relations (McPherson, Smith-Lovin, and Cook 2001).

SIMULATIONS AND IMPUTATIONS

Rather than accept the 2004 data as evidence for change in network size, Fischer tries two more techniques to argue that his best estimate is that no change occurred in confidant networks between 1985 and 2004. First, he conducts simulations that randomly assign 20 percent of the 1985 cases to zero. Our zero-inflated Poisson analyses show that the inflated zeros are clearly not random ($p < .00001$).

Fischer's second method of imputing data to explore the change in network size is more interesting (although reported only briefly in his footnote 16). He uses demographic variables and a set of questions about sociability (getting together with relatives, friends, and neighbors or going to a bar) to predict network size in 1985. He then substitutes the 2004 mean values on those variables into the prediction

[10] One graduate-educated person answered zero in 1985, and 23 did so in 2004 (weighted analysis). Two of those 23 respondents in 2004 were rated "restless and impatient" by their interviewers.

equation to get an estimate of network size in 2004. The predicted mean for 2004 Numgiven using his imputation procedure is 3.2, which is actually higher than the observed 1985 mean of 3.06. Fischer's imputation looks plausible until one realizes that virtually all of the variance explained in his imputation equation comes from the demographic variables. Adding the sociability variables to the demographic variables in the 1985 equation only increases the explained variance by .01. Fischer's imputation equation essentially assumes no change in network size, other than that which can be explained by demographic shifts. His finding of no change in the imputations is therefore not surprising.

OTHER EVIDENCE ON NETWORK TRENDS AND SOCIAL CHANGE

Fischer ends his comment with the suggestion that our 2006 results are implausible because they are inconsistent with other studies and cannot be explained sociologically with other major social changes during the same time period. We address these claims very briefly.

Evidence from Other Data

The major work positing a trend in social connectedness is, of course, Putnam's *Bowling Alone* (2000). In reviewing that book, Fischer (2005:158) asked, "Have many forms of sociality declined since about 1970?" He concludes that "given the wealth of data in *Bowling Alone,* the burden of proof is on the critics [who claim there is no decline]." Putnam (2000) analyzed political, civic, and religious participation, as well as informal social connections both in and out of the workplace. He argues for downward movement in all of them. Several of his figures show the steepest declines after 1985: the league bowling of the title (p. 112), daily informal socializing activity (p. 108), going to friends' homes (p. 99), active organizational involvement (p. 60), and attending a public meeting on town or school affairs (p. 43).

Other Social Change, 1985 to 2004

Fischer (2009:659) argues that the social change our models describe is implausible because "no social factors that might even plausibly cause such isolation . . . changed to any comparable degree in the same period." First, we remind the reader that our GSS variable measures only the closest of social ties. A subtle shift in the social structure toward a more extensive set of weaker ties could lead to a decline in closest confidants. Our people who report zero confidants are not totally isolated; they just lack these very strong ties. We briefly note below several social changes that occurred during this period that might have led to such a restructuring.

Since 1985, the Internet has come into vogue and been adopted (to some extent) by roughly two thirds of the U.S. population (Pew Research Center 2009). There is little reason to suppose that individual usage has strong effects on socializing (Robinson and Martin 2009), but one can imagine macro-level shifts in communication patterns as a result of such a sweeping technological change. Weaker ties might be fostered and maintained at a higher rate while strong ties are diffused, a pattern that Mayhew and Levinger (1976) suggested would occur with increasing system size. In a sense, the inexpensive ease with which we can now contact others without regard to physical distance has expanded the size of our personal social systems, but possibly at the cost of intimacy.

Evidence of other major social changes from 1985 to 2004 can be found in Fischer and Hout's *Century of Difference* (2006). They document growing inequality during this period, especially based on educational differences (Figure 6.4, p. 146). Family work hours rose as a result of women's employment (Figure 5.13, p. 125); college graduates are working longer hours now than in the mid-1980s (Figure 5.12, p. 123). The overall diversity of our society by race, religion, ethnicity, and nativity has increased. People are more likely to live alone (Figure 4.10, p. 84), with the change in the past two decades especially notable among the middle-aged. We would not specifically argue for the causal impact of any one of these factors. We do think, however, that many important features of social life that are not well documented in the GSS changed between 1985 and 2004.

PUBLIC SOCIOLOGY AND PUBLIC DATA

We have argued in this reply that parameter estimates (including percentages) that fail to model data appropriately will produce mis-

leading results. Public sociology is particularly susceptible to this pitfall. In hindsight, it was not a good strategy to emphasize the raw mean number of confidants and the marginal proportions of social isolates in our original article. The 1985 to 2004 differences estimated by our models in Table 5 are much more meaningful numbers, although not as vivid.

Our 2006 article received a great deal of press attention when it was first published. It received more than 12,000 hits on Technorati (evidence of significant discussion on the Web) and the second author did hundreds of interviews with print and broadcast media. The media buzz seems to have focused the public's (and Fischer's) attention on the marginals rather than the model.

We compounded the issue by accepting an invitation from the editors of *Contexts* to furnish an abridged version of our findings for that journal, aimed at a larger audience. In that version, we presented several charts that emphasized easy-to-understand marginals (McPherson, Smith-Lovin, and Brashears 2008b). While the last two figures presented our model-adjusted estimates, our text described descriptive statistics that could have led readers to an exaggerated conclusion. We would like to alert others to the perils of trying to (over)simplify complex phenomena; public sociology has both pitfalls and promise when isolated phrases from research articles in the *ASR* may appear on the front page of *USA Today.*

While the public attention to our results has given rise to the current controversy, the public nature of our data represents a clear advance in the scientific enterprise. By the time our paper was under review at the *American Sociological Review,* the data were already publicly available. Any reviewer or reader could download the data for free on the National Opinion Research Center's GSS Web site (http://www.norc.org/GSS+Website/) or do quick and easy analyses at the Survey Documentation & Analysis (SDA) Web site at Fischer's own institution (http://sda.berkeley.edu/archive.htm). Given the surprising nature of our findings, many researchers did analyze aspects of the data and began a conversation with us about the findings. Scholars can now debate evidence in real time while manuscripts are actually under review. We hope that the support of these large infrastructure data sources continues, because it fosters both the continuity of design that allows us to observe social trends and the open use of data to argue about evidence for those trends.

CONCLUSIONS

Fischer (2009:668) concludes his comment by saying that "the best estimate is that the 'true' percentage of 2004 respondents who were isolated was roughly the same—or perhaps less—than the percentage in 1985/1987, somewhere under 10 percent." We categorically disagree that the data show no change. Neither we nor Fischer have been able to destroy the 1985 to 2004 difference without assuming it away. Even accounting for the inflated number of zeros in 2004, there is a major decline in Numgiven in the data. If the 1985 to 2004 difference is illusory, it is due to the effect of variables that we have not been able to discover in those data. We are working on a survey experiment in the GSS to study the effects of fatigue and context on the network item in 2010. We expect that the next round of data on Numgiven will offer some new answers, and some new puzzles.

Miller McPherson *is Professor of Sociology at Duke University and Professor Emeritus at the University of Arizona. His evolutionary model of affiliation, originally published in this journal, has been extended to the study of a broad range of social and cultural phenomena, including occupations, musical genres, and churches. Current projects include a test of his ecological theory with nationally representative data from the Niches and Networks project, funded by the Human and Social Dynamics Initiative at the National Science Foundation.*

Lynn Smith-Lovin *is Robert L. Wilson Professor of Arts and Sciences at Duke University. Her research examines the relationships among social association, identity, action, and emotion. Her current projects involve an NSF-funded experimental study of justice, identity, and emotion; NSF-funded research with Miller McPherson on an ecological theory of identity; and an Office of Naval Research funded study of affective impression formation in the Arabic language.*

Matthew E. Brashears *is an Assistant Professor of Sociology at Cornell University. He is interested in social networks, information transmission and transformation, social psychology, and gender. He has been published in the* American Sociological Review, Social Psychology Quarterly, *and* Social Science Research.

REFERENCES

Bailey, S. M. and Peter V. Marsden. 1999. "Interpretation and Interview Context: Examining the General Social Survey Name Generator Using Cognitive Methods." *Social Networks* 21(3):287–309.

Fischer, Claude S. 2005. "Bowling Alone? What's the Score." *Social Networks* 27:155–67.

———. 2009. "The 2004 GSS Finding of Shrunken Social Networks: An Artifact?" *American Sociological Review* 74(4):657–69.

Fischer, Claude S. and Michael Hout. 2006. *Century of Difference: How America Changed in the Last One Hundred Years*. New York: Russell Sage Foundation.

Gelman, Andrew and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Mayhew, Bruce H. and Roger L. Levinger. 1976. "Size and the Density of Human Interaction in Social Aggregates." *American Journal of Sociology* 82(1):86–110.

McPherson, J. Miller. 1981. "A Dynamic Model of Voluntary Affiliation." *Social Forces* 59(3):705–728.

———. Forthcoming. "A Dynamic Baseline Model of Ego Networks." *American Behavioral Scientist*.

McPherson, J. Miller, Pamela A. Popielarz, and Sonia Drobnic. 1992. "Social Networks and Organizational Dynamics." *American Sociological Review* 57(2):153–70.

McPherson, Miller, Lynn Smith-Lovin, and Matthew E. Brashears. 2006. "Social Isolation in America: Changes in Core Discussion Networks over Two Decades." *American Sociological Review* 71:353–75.

———. 2008a. "Erratum: Social Isolation in America: Changes in Core Discussion Networks over Two Decades." *American Sociological Review* 73:1022.

———. 2008b. "The Ties that Bind are Fraying." *Contexts* 7(3):32–36.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:415–44.

Pew Research Center. 2009. *Pew Internet and American Life Surveys, March 2000–December 2008*. Retrieved March 12, 2009 (http://www.pewinternet.org/trends/Internet_Adoption_Jan_2009.pdf).

Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community.* New York: Simon and Schuster.

Robinson, John P. and Steven Martin. 2009. "IT Use and Declining Social Capital? More Cold Water from the General Social Survey (GSS) and the American Time-Use Survey (ATUS)." Unpublished paper.