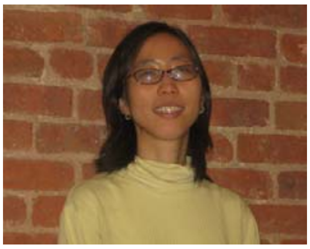
LOGISTIC REGRESSION TO DESCRIBE HOW THE RELATIONSHIP BETWEEN SOCIAL CONNECTION AND SELF-RATED HEALTH VARIES BY GENDER

Michael A. LEWIS[1,2]
Hunter College School of Social Work, New York, USA

E-mail: michael.a.lewis@hunter.cuny.edu

Eri NOGUCHI[3]
Columbia University School of Social Work, New York, USA

E-mail: en16@columbia.edu



Abstract: *Social Scientists and health researchers have allocated a great deal of time trying to understand the social correlates of health outcomes, some in the question of the association between social relationships and such outcomes. This paper builds on this work by exploring the association between how connected one feels to her or his neighborhood and the chance that one reported that she or he was in poor health. Using logistic regression on a data set of 1,899 females and 1,521 males from the United States we found that, for both males and females, those who reported that their neighbors give them a sense of community were less likely to report that they were in poor health, but this association was stronger for females than for males.*

Key words: *Logistic Regression; Social Connections; Self-rated Health; Gender*

1. Introduction

Social scientists and health researchers have done a great deal of quantitative work on the social correlates of health outcomes. Race (Gibbs, et al., 2006), gender (McDonough, Peggy and Vivienne, Walters, 2001), and socioeconomic status (Malat, Jennifer R.; van Ryn, Michelle; Purcell, David, 2006) are among these correlates. Another key correlate of health outcomes, one that will be the focus of this paper, is social connection. It has been found that the degree of people's connections to others is negatively associated with the chance of neonatal death (Gayen and Raeside, 2007), degree of mental health problems (Fiori, Antonucci, and Cortina 2006; Ueno, 2005; and Cannuscio, Colditz, and Rimm, 2004), impairment of physical functioning (Unger, McAvay, and Bruce 1999), and the risk of developing a disability (Mendes de Leon, Glass, and Beckett, 1999). The degree of people's connections to others has been found to be positively associated with the chance of recovering from a disability (Mendes de Leon, Glass, and Beckett, 1999) and level of self-rated health status (Helweg-Larsen, Kjoller, and Thoning 2003; Zunzunegui, Kone, and Johri, 2004; and Veenstra, Luginaah, and Wakefield, 2005). Very little of this research, however, has explored the question of whether the association between social connection and health outcomes varies by gender.

There are good reasons, however, for thinking that the association between social connection and health outcomes might differ by gender. First, such a difference may be due to gender based differences in the composition of groups to whom people are connected, and these compositional differences may result in females being influenced in ways that lead them to behave in ways different from males. Second, even if group composition is the same, females, on average, may relate to people differently than men, leading to differences in the influence of social connection on health.

This paper builds on previous work by exploring the association between social networks and health outcomes. More specifically, we address the question of the degree of association between whether one reported that one's neighbors give them a sense of community and the likelihood that one reported that she or he was in poor health. We also assess whether this association varies by gender. We will proceed as follows.

First, we will provide more details regarding the theoretical model we explored. Second we will discuss the data set and methods we used. Third, we will provide an overview of our results. And fourth, we will conclude with a brief discussion of these results.

Theoretical Model

Let i represent a given individual and assume that she or he lives in a given neighborhood. Let $COMMUNITY_i$ be a variable representing whether i's neighbors give him or her a sense of connection to others or a sense of community. If i feels that his or her neighbors give him or her such a sense, this variable takes the value 1. If not, it takes the value 0. Let $POOR_i$ be a variable representing whether i feels he or she is in poor health. If i feels he or she is in poor health, this variable takes the value 1. If i does not feel this way, it takes the value 0.

Having set out these preliminaries, we propose the following model of self-rated poor health:

$$P(POOR_i = 1) = f(COMMUNITY_i, \mathbf{V}_i) \qquad (1)$$

where $P(POOR_i = 1)$ refers to the probability that person i reported that she or he is in poor health, f refers to some mathematical function, and $\mathbf{V}_i$ represents a set of other factors that are thought to be associated with $P(POOR_i = 1)$.

Data and Methods

The data we used to examine the above model came from the Social Capital Community Benchmark Survey, which was designed by the Saguaro Seminar at Harvard University Kennedy School of Government. A sample of respondents from across the United States was obtained through random-digit-dialing (RDD) procedures. The actual telephone interviews were conducted by Intersearch, an international survey firm, and each one took an average of 26 minutes to complete (Roper Center, 2005). Our analysis is based on 1,899 employed female and 1,521 employed male respondents. Some were employed full-time (>=35 hours/week) and some part-time (< 35 hours/week).

Most of the survey questions required respondents to choose, from among pre-determined answers, the one that, in their judgment, best reflected them. For example, someone who was primarily a full-time homemaker, but who had worked a little for pay the previous week, would decide for herself whether she identified as a "part-time worker" or "full-time homemaker." Data are available, from this survey, on a variable referred to as "health." It is based on respondents' self-reported health statuses. Respondents were asked to rate their health as poor (coded 0), fair (coded 1), good (coded 2), very good (coded 3), or excellent (coded 4). Those who stated that they didn't know their health status were coded 8 and those who refused to offer a response were coded 9. For the analyses described below, we recoded health to $POOR_i$. 0 was recoded to 1, 8 and 9 were recoded to system missing (that is, they were not considered in the analyses, and 1, 2, 3, and 4 were recoded to 0. We didn't consider codes 8 and 9 because we were only interested in those respondents who actually reported a category to describe their health status.

$COMMUNITY_i$ was also based on a recoding of an original variable. In the original survey, this variable is called belnei and came from the question whether respondents felt that their neighbors give them a sense of community. Those who answered "no" were coded 1, those who answered "it depends" 2, those who answered "yes" 3, those who answered "does not apply" 4, those who answered "don't know" 8, and those who refused to answer 9. We were not interested in those who didn't know the answer to the question or who refused to answer, so 8 and 9 respondents were eliminated from the analysis. 1, 2, and 4 were recoded as 0, and 3 was recoded as 1. In order to determine if our data were consistent with our proposed theoretical model, we analyzed the data using logistic regression analysis (LRA).

Analysis

As stated above, we assumed that variables in addition to $COMMUNITY_i$ might be associated with $POOR_i$. These variables are contained in $\mathbf{V}_i$ and were included in our LRA models as covariates. In many cases, these variables were also based on recodes of original variables, and respondents who answered with "don't knows" or who refused to answer questions were not further considered in our analyses. We discuss more about the relevant aspects of these variables in the appendix.

As stated in equation (1), our interest was in modeling the probability that $POOR_i$ = 1 as a function of $COMMUNITY_i$ and a set of covariates represented by $\mathbf{V}_i$. A standard statistical approach used to test such models, is the Linear Probability Model (LPM), a variant of Least Squares Regression Analysis. In our case, LPM could have been used to test the model specified in equation (1). A key problem associated with LPM, however, is that one can end up with predicted probabilities outside the interval [0,1]. To address this problem, we used the alternative approach of LRA. LRA models were run with SPSS 13.0, with separate models run for males and females to determine if the association between $COMMUNITY_i$ and $P(POOR_i = 1)$ varied by gender, controlling for the variables included in $\mathbf{V}_i$. Unlike many works in the social sciences, logistic regression was used as a descriptive as opposed to inferential technique. This requires some explanation.

Researchers who analyze survey data invariably have to deal with the problem of respondents refusing to answer questions, giving incomplete answers, or in some other way providing less than useful answers. This is the missing data problem. One way of handling it is through listwise deletion. But, as is well known (Berk, 2004), if one is interested in generalizing from a sample to a population, there is a key shortcoming of listwise deletion. Cases with complete sets of values may be systematically different from cases with missing values. Thus, even if one started out with a probability sample, once one drops cases with missing values from the analysis, there is a very good chance that the sample one ends up with is no longer a probability sample. This is a serious problem, statistically, because it can lead to biased estimates of population parameters (Gelman and Hill, 2007 and Berk, 2004). Given the fact that, after listwise deletion, we were left with only 1,899 out of the 15,299 females in the overall data set and 1,521 out of 10,526 males, we were quite concerned that the biased estimates problem might apply to our situation.

There are other techniques designed to address the missing data problem that are often utilized by social scientists. These techniques often involve the imputation of values for cases with missing data. But imputing values either has the same potential to result in biased estimates as listwise deletion does or requires one to have good information about the mechanisms that account for why cases end up with missing data (Berk, 2004). Following the advice of Berk (2004), since we didn't feel that we had such information, we chose not to utilize any of these imputation methods.

A way to address our dilemma, which some statisticians regard as quite useful but underrated (Berk, 2004), is to use regression analysis to describe patterns in a data set, instead of as a method of inference. In our case, logistic regression could be used descriptively to determine if the relationship between $COMMUNITY_i$ and $P(POOR_i = 1)$ varied by gender for the set of cases we had after listwise deletion (Berk, 2004). This use of regression has much in common with similar methods, such as classification algorithms, found in computer science (Bramer, 2007). Since the approach dispenses with the notion of using the sample to test hypotheses about population parameters, the concern about biased estimates is no longer relevant. A limitation of this approach is that one isn't in a position to know the extent to which one's findings hold in other populations. But since the alternative was, in our view, a good chance of ending up with biased estimates, we were willing to accept this limitation. We were also willing to accept it because we realize that it can be addressed by other researchers going out and trying to replicate our findings in different populations, something we regard as one of the hallmarks of science.

Results

Table 1 contains summary statistics for the variables included in our Logistic regression models, separately for females and males.

Table 1. Summary Statistics for Variables Included in Logistic Regression Models by Gender

Variables	**Females N= 1899**		**Males N = 1521**	
	Mean/Mode	Standard Deviation/Range	Mean/Mode	Standard Deviation/Range
Income[m,r]	2	0-5	2	0-5
Religious[m,r]	1	1-5	1	1-5
Soctrst	.03	.67	.02	.68
Tvhrs	2.65	2.50	2.63	2.44
Age	36.74	9.04	37.62	9.51
Kids_5	.55	.75	.67	.90
Commute	.40	.39	.49	.49
PARTTIME[b]	.23	.42	.07	.25
FULLTIME[b]	.77	.42	.93	.26
BLACK[b]	.20	.40	.13	.34
ASIAN[b]	.02	.15	.03	.16
NATIVE[b]	.01	.12	.02	.13
OTHER[b]	.08	.27	.09	.29
WHITE[b,ref1]	.69	.46	.73	.44
SATISFIED[b]	.95	.21	.95	.21
CITIZEN[b]	.96	.19	.93	.26
HOME[b]	.73	.45	.76	.43
POOR[b]	.01	.11	.01	.10
NOTPOOR[b,ref2]	.99	.10	.99	.10
COMMUNITY[b]	.82	.38	.83	.38

[m,r] Modes and ranges were reported since these were technically ordinal variables.
[b] These were assumed to be Bernoulli variables and means and variances were calculated accordingly (see Wasserman, 2005).
[ref1] The reference category for the race/ethnicity variables above it.
[ref2] The reference category for the variable directly above it.

Table 2 contains our key results of interest, the female and male adjusted odds ratios for $COMMUNITY_i$. These ratios tells us how the odds of reporting that one is in poor health for those assigned a 1 for $COMMUNITY_i$ compares with the odds for those assigned a 0.

Table 2. Female and Male Odds Ratios

Variables	**Female Odds Ratio**	**Male Odds Ratio**
Income	.81	.89
Religious	.98	1.23
Soctrst	1.67	2.45
Tvhrs	1.12	1.09
Age	1.03	1.02
Kids_5	1.86	.84
Commute	1.41	1.92
PARTTIME	1.52	2.44
BLACK	.11	.52

ASIAN	3.21	4.27
NATIVE	1.85	4.33
OTHER	.00	1.45
SATISFIED	.36	.25
CITIZEN	$3*10^7$	4.54
HOME	.87	.65
COMMUNITY	.39	.69
Constant	.00	.00

For females, Table 2 shows that the adjusted odds ratio was about .39. Thus, the adjusted odds of reporting poor health, if their neighbors give them a sense of community, were about .4 of the adjusted odds for those whose neighbors do not give them such a sense. That is, for women, neighborhood networks seemed to be associated with a decrease in the chance of reporting poor health. For males, Table 1 indicates that the adjusted odds ratio was about .69. Thus, the adjusted odds of reporting that one is in poor health, if their neighbors give them a sense of community, were .7 of the adjusted odds for those whose neighbors do not give them such a sense. Thus, for males also, a sense of community is associated with a decrease in the chance of reporting poor health, although this decrease is smaller than was the case for females.

We should also say a bit about how well our models fit the data. There are many ways to communicate such fit, but we think the most intuitive is to report the accuracy of predictions using our models. One way to think of LRA is as a method of trying to predict membership in a given set. In our case, we were trying to predict whether people were in the set of those who regard themselves to be in poor health or in the set of those who don’t. For both our male only and female only models, 99% of our predictions were correct, indicating excellent fit to our data.

Discussion

This paper has been concerned with the association between whether one gets a sense of community from her or his neighbors and the chance that one reports he or she is in poor health and whether this association varied by gender. Using logistic regression analysis to describe patterns in our data, we found an association between a sense of community and the chance of rating oneself in poor health, for both males and females, with the association for females being stronger than that for males. To our knowledge, this is the first paper to explore how the association between a sense of community and self-rated health varies by gender. This finding is important because it may lead to other research that explores how the association between social connection and other measures of social interest may depend on gender. Further research on this issue may deepen our understanding of the roles played by social connection and gender in social life.

References

1. Berk, R.A. **Regression Analysis: A Constructive Critique**, Sage Publications, California, 2004
2. Bramer, M. **Principles of Data Mining**, Springer, London, 2007
3. Cannuscio, C. C., Colditz, G. A. and Rimm, E. B. **Employment Status, Social Ties, and Caregivers' Mental Health**, Social Science and Medicine, vol. 58, no. 7, April 2004, pp. 1247-1256
4. Fiori, K., Antonucci, T. C. and Cortina, K. S. **Social Network Typologies and Mental Health Among Older Adults**, Journal of Gerontology, Series B: Psychological Sciences and Social Sciences, vol. 61B, no. 1, January 2006, pp. 25-32
5. Freedman, D., Pisani, R. and Purves, R. **Statistics,** W.W. Norton & Company, New York, 2007
6. Gayen, K. and Raeside, R. **Social Networks, Normative Influence and health Delivery in Rural Bangladesh,** Social Science and Medicine, vol. 65, no. 5, September 2007, pp. 900-914
7. Gelman, A. and Hill, J. **Data Analysis Using Regression and Multilevel/Hierarchical Models,** Cambridge University Press, New York, 2007
8. Gibbs, B. K., Nsiah-Jefferson, L., McHugh, M. D. and Prothrow-Stith, D. **Reducing Racial and Ethnic Health Disparities: Exploring an Outcome-Oriented Agenda for Research and Policy,** Journal of Health Politics, Policy and Law, vol. 31, no. 1, pp. 185-218
9. Helweg-Larsen, M. Kjoller, M. and Thoning, H. **Do Age and Social Relations Moderate the Relationship Between Self-Rated Health and Mortality Among Adult Danes?,** Social Science and Medicine, vol. 57, no. 7, October 2003, pp. 1237-1247
10. Kleinbaum, D.G. **Logistic Regression: A Self Learning Text,** Springer, New York, 1994
11. Malat, J. R., van Ryan, M. and Purcell, D. **Race, Socioeconomic Status, and the Perceived Importance of Positive Self-Preservation in Health Care,** Social Science and Medicine, vol. 62, no. 10, 2006, pp. 2479-2488
12. McDonough, P. and Walters, V. **Gender and health: Reassessing Patterns and Explorations,** Social Science and Medicine, vol. 52, no. 4, 2001, pp. 547-559
13. Mendes de Leon, C. F., Glass, T. A., Beckett, L. A. **Social Networks and Disability Transitions Across Eight Intervals of Yearly Data in the New Haven EPESE,** Journal of Gerontology, Series B: Psychological Sciences and Social Sciences, vol. 54 B, no. 3, May 1999, pp. S162-S172
14. Ueno, K. **The Effects of Friendship Networks on Adolescent Depressive Symptoms,** Social Science Research, vol. 34, no. 3, September 2005, pp. 484-510
15. Unger, J. B., McAvay, G. and Bruce, M. L. **Variation in the Impact of Social Network Characteristics on Physical Functioning in Elderly Persons: MacArthur Studies of Successful Aging,** Journal of Gerontology, Series B: Psychological Sciences and Social Sciences, vol. 54 B, no. 5, September 1999, pp. S245-S251
16. Veenstra, G., Luginaah, I. and Wakefield, S. **Who You Know, Where You Live: Social Capital, Neighborhood and Health,** Social Science and Medicine, vol. 60, no. 12, June 2005, pp. 2799-2818
17. Wasserman, L. **All of Statistics: A Concise Course in Statistical Inference,** Springer, New York, 2005
18. Zunzunegui, M. V., Kone, A. and Johri, M. **Social Networks and Self-Rated Health in Two French-Speaking Canadian Community Dwelling Populations Over 65,** Social Science and Medicine, vol. 58, no. 10, May 2004, pp. 2069-2081
19. * * * **Social Capital Community Benchmark Survey,** Roper Center, www.ropercenter.uconn.edu/scc_bench.html, retrieved on August 25, 2005

Appendix

This appendix focuses on the covariates in the logistic regression models we discussed in the body of the paper. This is so that researchers who'd like to replicate our findings will know precisely what we did. The table below contains the names of the original variables found in the survey, the original coding, the recoded variable names, and the recodes.

Original Variables/Coding and Recoded Variables/Recodes

Original Variable Name	Original Codes	Recoded Variable Name	Recodes
Wrktime	other = -9; don't know = -7 and 98; refused = -6 and 99; no answer = -5; blank = -4; number of hours, on average, one works in week	PARTTIME	average hours worked/week < 35hrs/week = 1; average hours >= 35 hrs/week = 0; don't know, refused, blank, and no answer not considered
happy	0 = not happy with one's life; 1 = not very happy with one's life; 2 = happy with one's life; 3 = very happy with one's life; 8 = don't know; 9 = refused	SATISFIED	0 and 1 recoded to 0; 2 and 3 recoded to 1; 8 and 9 not considered
Citizen	non-citizen = .00; citizen = 1.00	CITIZEN	non-citizen = 0; citizen = 1
Own	rent home = 0; own home = 1; don't know = 8; refused = 9	HOME	rent home = 0; own home = 1; don't know and refused not considered
race_all	white = 1; black = 2; Asian/Pacific Islander = 3; Alaskan Native/Native American = 4; Other = 5; don't know = 8; refused = 9	[1]BLACK; ASIAN; NATIVE; OTHER; WHITE was the reference group	black = 1 non-black = 0; Asian/Pacific Islander = 1 non-Asian/Pacific Islander = 0; Alaskan Native/Native American = 1 non-Alaskan Native/Native American = 0; Other =1 non-other = 0
income	income in [$0, $20K); in ($20K, $30K); in ($30K, $50K); in ($50K, $75K); in ($75K, $100K); income >= $100K	INCOME	income in [$0, $20K) = 0; in ($20K, $30K) = 1; in ($30K, $50K) = 2; in ($50K, $75K) = 3; in ($75K, $100K) = 4; income >= $100K = 5
Relaten2	< yearly attendance at religious service = .00; a few times/year = 1.00; 1-2 times/year = 2.00; almost weekly = 3.00; weekly or more = 4	RELIGIOUS	< yearly attendance at religious service = 1; a few times/year = 2; 1-2 times/year = 3; almost weekly = 4; weekly or more = 5
[2]soctrst	Index created by survey designers		
Age	measured in years		
kids_5	Number of kids under five for whom one is primary caretaker		
Commute	number of hours it takes to commute to work		
Tvhrs	number of hours of television watched on the average weekday		

[1]These were dummy variables; the coding scheme is 1 for a person who falls in a given category (such as the category of black persons and 0 for those who don't fall in the given category). White persons served as the reference category in our analysis.

[2]Where no variable appears in the third column of the table, this indicates that the original coding was used in our analysis.

---

[1] **Correspondence:** Michael A. Lewis, 129 East 79th Street, New York, NY 10075

[2]Michael Lewis holds a masters degree in social work From Columbia University and a doctorate in Sociology from the City University of New York Graduate Center. He teaches courses in social welfare policy and political economy at the Hunter College School of Social Work. Lewis' research interests are in poverty, social welfare policy, and quantitative methods. His work, some co-authored with Eri Noguchi, has appeared in a number of peer-review journals.

[3]Eri Noguchi holds a masters degree in social work, a masters degree in public administration, and a doctorate in Sociology, all from Columbia University, and currently teaches courses on human behavior, program evaluation, and administration at Columbia University and Hunter College Schools of Social Work. Her research interests include labor markets, poverty, social welfare policy, and civic engagement/participation. She also serves as the Chief Program Officer of the Association to Benefit Children - a New York-based non-profit organization dedicated to providing early childhood, youth, and wrap-around family support services to children and families.