Analyzing Multiple Informant Data from an Evaluation of the Health Disparities Collaboratives

A. James O'Malley, Bruce E. Landon, and Edward Guadagnoli

Objective. To use multivariate regression methods to analyze simultaneously data obtained from multiple respondents or data sources (informants) at health centers.

Data Source. Surveys of executive directors, medical directors, and providers from 65 community health centers (176 informants) who participated in an evaluation of the Health Disparities Collaboratives.

Study Design. Cross-sectional survey of staff at the health centers during 2003–2004.

Statistical Methods. In order to illustrate this method, we analyze the association between informants' assessments of the culture of the center and participation in the collaborative, and the association between computer availability and the effort made by management to improve the quality of the care and services at their center. Multivariate regression models are used to pool information across informants while accounting for informant-specific effects and retaining informants in the analysis even if the data from some of them are missing. The results are compared with those obtained by traditional methods that use data from a single informant or average over informants' ratings.

Findings. In both the Collaborative participation and quality improvement efforts analyses, the multivariate regression multiple informants' analysis found significant effects and differences between informants that traditional methods failed to find. Participating centers emphasized developmental (entrepreneurship, innovation, risk-taking) and rational culture. The effect of hierarchical culture (stability and bureaucracy) on participation depended on the informant; executive directors and medical providers were the most discrepant. In centers that participated in the Collaborative, the availability of computers was positively associated with the effort that management made toward improving quality.

Conclusions. The multiple informants model provided the most precise estimates and alerts users to differential effects across informants. Because different informants may have different insights or experiences, it is important that differences among informants be measured and ultimately understood by health services researchers.

Key Words. Multiple informants, multivariate regression, random effects, generalized estimating equations, health disparities, quality

Health services researchers often collect similar or even identical data from multiple sources (e.g., physicians, nurses, patients, teachers, parents) in order to increase the reliability of the measurements or to gain insights from several different perspectives and contexts. When multiple informants measure an independent variable, researchers are typically interested in how the variable or construct measured by each informant affects the outcome. When multiple informants assess an outcome variable, the focal point is often the overall effect of the independent variables on the outcome. In either case, more precise estimates are possible if effects can be pooled across informants.

Traditional approaches for analyzing data from multiple informants include conducting independent analyses for each informant, selecting one informant as the informant of interest and ignoring others (i.e., only using the data obtained from the chosen informant for the analysis), and reducing the data from the informants into an overall measure (e.g., the mean score across informants). However, these methods have several disadvantages. The effects of individual informants are often ignored, differences between or among informants are not testable, and the degree of correlation between informants is not estimated (Horton and Fitzmaurice 2004). In addition, summary measures (e.g., means) are not consistently defined when one or more informants' scores are missing, and will result in a loss of sample size and hence suboptimal analysis if only those units with a full set of responses are included.

The exclusion of data from some informants, or the reduction of data from multiple informants to a summary statistic, is one of the ways in which data are edited in practice. In general, investigators edit data in ad hoc, idiosyncratic ways and often do not disclose their unique strategies (Leahey, Entwisle, and Einaudi 2003; Leahey 2004). Because data-editing decisions can profoundly affect conclusions (Gould 1981; Dewald, Thursby, and Andersen 1986), it is important that these decisions are documented, or better still are avoided (Leahey 2004). For example, if the relationship between two variables varies in direction across informants the relationship between the means of the informants' ratings could be misleading. It is better to utilize multiple (even multiple discordant) views rather than reducing the data to a convenient form for analysis.

Address correspondence to A. James O'Malley, Ph.D., Assistant Professor of Statistics, Department of Health Care Policy, 180 Longwood Ave., Boston, MA 02115-5899. Bruce E. Landon, M.D., M.B.A., Associate Professor of Health Care Policy and Assistant Professor of Medicine, is with Harvard Medical School; and Practicing Internist, Beth-Israel Deaconess Medical Center; and the Department of Health Care Policy, Boston, MA. Edward Guadagnoli, Ph.D., Professor of Health Care Policy, is with the Department of Health Care Policy, Harvard Medical School, Boston, MA.

Individual responses to surveys can be influenced by several sources of variation including characteristics of individuals, which can affect an individual's interpretation of the construct being measured such as their training and experience, or position in the organization. For example, in sociology the social status of an informant has been shown to have a bearing on the response obtained (Leahey, Entwisle, and Einaudi 2003; Leahey 2004). These characteristics of the informant and context need to be accounted for in order to pool information sensibly across informants, to determine the best way of measuring a construct, and ultimately to make general inferences about the relationship of the variable being measured with other variables.

Multivariate regression models have recently been used to analyze data from multiple informants (Horton, Laird, and Zahner 1999; Goldwasser and Fitzmaurice 2001; Horton et al. 2001; Lash et al. 2003; Horton and Fitzmaurice 2004). The key characteristic of this approach is that individual data from all informants are modeled, enabling the evaluation of comparative inferences about informants (e.g., testing whether the informants have the same or different effects), and avoiding reducing or omitting the data for a unit of analysis because data from some informants were unavailable or missing.

Another model that might be considered for multiple informant data is the "method factor" model (Bollen and Paxton 1998). In the context of a method-factors model, the informant is the "method" of measurement. The analysis of a method-factors model is typically based on a structural equations model that treats the informants' ratings as subjective error-prone measurements of an unobserved trait. In contrast, the multivariate regression models referred to in the preceding paragraph allow for multiple underlying traits, and directly model informant specific effects as opposed to estimating the effect of a common underlying trait (although this could be accommodated within the multivariate regression framework). Because we want to model the effect of informant-specific effects on the outcome, rather than the effect of an underlying trait, herein we use the multivariate regression approach.

In this paper, we analyze data collected from surveys of personnel at community health centers (CHCs) that are part of an ongoing evaluation of a Bureau of Primary Health Care-sponsored quality improvement (QI) program (HRSA 2004). We compare standard approaches for analyzing multiple informant data, including picking one informant as representative of the whole and averaging over informants' ratings, with multivariate regression methods. We address the cases in which information on an independent or dependent variable of interest is available from multiple informants. Although multivariate regression models for analyses of multiple informants have

already been established, the advantages of using this approach over standard approaches have not been demonstrated previously using real data. The results will be useful to health services researchers as multiple informant data become increasingly common.

METHODS

Overview

As part of a controlled evaluation of a QI program at CHCs, we surveyed three types of health professionals at participating and nonparticipating health centers about personal and health center characteristics including culture, QI efforts, and availability of computers. Both culture and QI efforts are assessed by multiple informants while the availability of computers is assumed to be known. To illustrate the case where an independent variable is available from multiple informants, we examine whether staff members' assessments of the culture at a health center are associated with the center's participation in the QI program and whether associations vary across informants. To address the situation where a dependent variable is available from multiple informants, we examine whether characteristics of a health center, such as computer availability, size of center, and urban versus rural location, and informant characteristics such as gender and age are related to assessments of QI efforts at the center. For both situations, we compare results under the multiple informant (MIF) multivariate regression model with the results from standard regression models for a single informant (SIF) and the mean of the informants' ratings (MR) to assess the extent to which these often-used basic methods may lack precision or mask important effects. The analyses embody six distinct models corresponding to whether the independent or dependent variable is measured by multiple informants and the corresponding three modeling approaches (i.e., the SIF, MR, and MIF models) applied to each case. In addition, we use the MIF model to determine whether different effects are observed across different types of informants.

Setting

In 1998, the Bureau of Primary Health Care's Health Resources and Services Administration (HRSA) initiated the Health Disparities Collaboratives in response to mounting evidence of disparities in the health and quality of care of the underserved, uninsured, and underinsured served by CHCs in the United States. The goal of these disease-specific collaboratives was to

generate and document improved health outcomes in underserved populations with chronic medical conditions (HRSA 2004). The evaluation included a national sample of 44 health centers that participated in a collaborative and 21 control centers matched by region, setting (urban versus rural), number of sites, and the total number of patients. QI teams established at each center are encouraged to design interventions guided by the Chronic Care Model (Wagner, Austin, and Von Korff 1996).

Survey of Health Center Staff

At each center, we surveyed the executive director, deputy director, medical director, and a randomly selected provider (at up to three different sites for centers with multiple sites). We mailed 117 surveys to executive and deputy directors and 235 surveys to clinical personnel, for a total of three to nine surveys per center. Nonrespondents were subject to mail and telephone follow-up. A total of 298 surveys were returned (an overall response rate of approximately 85 percent).

For the analyses described herein, we use data from the executive director (or deputy if the director is not available), medical director, and one clinician (a physician or other clinician if a physician is not available) at each center. In centers with multiple clinician respondents, we selected a provider from the primary site (this was the site that the collaborative team leader was from for participating centers and the site with the greatest number of informants for control centers), resulting in at most one respondent of each type from each center (the standard scenario in multiple informants' analyses).

Dependent Variables

The binary variable indicating participation in the collaborative is the dependent variable for the case where the independent variable is measured by multiple informants (case 1). The five-item *leadership* scale (see Table 1 for description) of the measure of QI in the organization (Shortell et al. 1995) is the dependent variable for the case where the dependent variable is measured by multiple informants (case 2).

Independent Variables

The independent variable of main interest for the analysis of collaborative participation is culture. The culture of a center is assessed by each informant according to four ideal types of culture (group, developmental, hierarchical, and rational) based on the competing values model described by Zammuto

Table 1: Definition of Culture and Leadership Scales

<i>Measure</i>	<i>Scale</i>	<i>Interpretation of High Value</i>
Culture	Group	Emphasizes teamwork, participation, fairness, and loyalty
	Developmental	Emphasizes entrepreneurship, growth, risk-taking, innovation, and change
	Hierarchical	Emphasizes bureaucracy, rules, and tradition
	Rational	Emphasizes efficiency and achievement of goals
Quality improvement effort	Leadership	Senior executives articulate clearly the center's values relevant to quality and care, and have proven ability at making the changes needed to improve quality

and Krakower (1991). Scores indicate the extent to which that aspect of culture is emphasized at the expense of the others (see Table 1 for interpretation). The scores are positive quantities that sum to 100; hence, the types of culture compete for their share of the 100-point total.

The key independent variables in the QI (leadership) analysis are participation in the collaborative and computer availability. Other independent variables used in both models are fixed characteristics of the centers (computer availability, number of sites, number of patients treated, urban versus rural setting), and characteristics of the individual respondents (informant type [executive director, medical director, provider], gender [male versus female], ethnicity [Hispanic versus non-Hispanic], race [black versus nonblack], age, and length of employment).

Statistical Models for Multiple Informant Data

Although the models can all be described for general outcome variables (continuous, ordinal, categorical), we present them for the case of a binary outcome for the analysis of assessments of the independent variable by multiple informants and the case of a continuous outcome when multiple informants assess the dependent variable. We denote the individual informants within a center by $j = 1, \dots, J$ where J is the total number of informants, the representative informant in a single informant analysis by j_* , and the centers by $i = 1, \dots, n$. In practice, j_* often corresponds to the informant deemed to have the greatest knowledge, insight, or authority. However, j_* could be selected at random. The data from all other informants are not used in a single informant analysis.

Case 1: Independent Variable Assessed by Multiple Informants. In this case, the dependent variable (participation in the collaborative) is measured at the unit of analysis (the CHC) and thus does not vary across informants, whereas the key independent variable (informant-assessed culture of the CHC) is informant specific.

The SIF model regresses the unit-level dependent variable (Participation_{*i*}) on informant j_* 's rating of culture (Culture_{*ij_**}) and any other covariates either specific to the informant j_* (CovI_{*ij_**}) or measured at the unit level and thus common to all informants (CovU_{*i*}). The MR model takes the same form but uses the mean rating of culture across informants at the center in place of the single informant's rating. The SIF and MR models are standard logistic regression models.

In contrast to the SIF and MR models, the MIF model simultaneously fits separate logistic regression equations for each informant (Horton, Laird, and Zahner 1999; Pepe, Whitaker, and Seidel 1999). These are incorporated into a single model by defining the dependent variable separately for each informant even though it has the same value for all informants. In the case of a binary outcome such as participation in the collaborative, the model may be expressed as

$$\begin{aligned} \text{Logit}(\text{Pr}[\text{Participation}_{ij} = 1 | \text{Informant} = j]) \\ = a_j + b_j \text{Culture}_{ij} + c_j \text{CovI}_{ij} + d \text{CovU}_i \quad (1) \end{aligned}$$

where Logit denotes the logit link function that characterizes a logistic regression model, Pr denotes probability, (a_j , b_j , c_j) are regression parameters specific to informant j , and d is a regression parameter common to all informants.

The distinguishing feature between the MIF model in Equation (1) and fitting separate SIF models to each informant is that the former is a single model. Only the MIF is able to test whether the construct measured by each informant's rating of culture has the same or differential associations with the likelihood of participation in the collaborative; it does so by comparing the fit of the model in which $b_1 = \dots = b_J$ against the model that has unrestricted coefficients (i.e., allowing the effect to vary across informants). When the data indicate that the regression coefficient for a predictor variable is the same for all informants, the coefficient may be fixed across informants in the MIF model to obtain inferences with smaller standard errors. In addition, by modeling the data at the individual level, the MIF model avoids having to delete or impute data for centers with missing observations for some informants (as often is the case when using the MR model).

In the MIF model, the robust “sandwich estimator” (White 1982; Liang and Zeger 1986) is used to correct the covariance matrix of the parameter estimates for the clustering of observations within centers.

Case 2: Dependent Variable Assessed by Multiple Informants. In this case, the dependent variable (the “leadership” component of the QI effort scale) varies between informants whereas the key independent variables of interest are often at the unit level (e.g., computer availability). A linear regression model for informant j_* regresses the rating for the dependent variable (Leadership $_{ij_*}$) on the informant-specific and unit-level covariates, again denoted by CovI $_{ij_*}$ and CovU $_b$ respectively. The MR model has the same form as the SIF model but with the center mean rating of leadership as the dependent variable; this is often referred to as an *ecological regression model* because both the dependent and the informant-level independent variables are aggregated to the unit level. A problem with the MR model is that the relationship between aggregated variables can be opposite to the relationship at the individual level (Simpson’s paradox). This can lead to misleading conclusions when models of aggregated data are treated as if they yield individual-level results (the ecological fallacy).

The MIF model is the generalization of the SIF model with the following general form:

$$E[\text{Leadership}_{ij} | \text{Informant} = j] = a_j + b_j \text{CovI}_{ij} + d \text{CovU}_i \quad (2)$$

In this model, a_j accounts for unexplained variation in Leadership $_{ij}$ specific to the j th informant, and the dependence of b_j on the informant enables a test of differences in the effects of CovI $_{ij}$ across informants to be conducted. If the values of b_j do not vary significantly they are assigned a common coefficient and the model simplifies. By using the data from all informants, more information is available to estimate the effects of the independent variables than for the SIF model.

To account for the correlation between the observations made on the same center, one can either treat the center as a random variable and use a hierarchical linear model (Raudenbush and Bryk 2003) with random effects for the centers, or fit a marginal regression model using generalized estimating equations (Liang and Zeger 1986; Zeger and Liang 1986). The hierarchical linear model (also known as a mixed effects model) accommodates a wider range of correlation structures but requires more assumptions, such as assuming a particular distribution for the random effects.

When the dependent variable is measured by multiple informants, the MIF model reduces to the MR model in the special case where there are no informant-specific effects, equal variances among informants, and equal correlations between informants (Goldwasser and Fitzmaurice 2001).

Statistical Analyses

As a precursor to performing a multiple informant analysis, we assess the level of correlation between the informants' ratings for each measure of culture and for leadership. Very high correlations indicate that a multiple informants' analysis might be redundant because the informants' ratings are strongly related. On the other hand, low correlations might indicate that averaging the informants' ratings may obscure important differences.

For the SIF and MR models, we report estimated effects and *p*-values for the key independent variables in the model (the informants' ratings of culture in the collaborative participation model and the characteristics of the centers and informants in the leadership model). In order to include the same centers in the analysis as for the MIF analysis, the mean of the informants' ratings was used in the MR analysis even if only a single informant responded. An alternative approach often used in practice is to omit centers that do not have ratings from each informant. Although unit deletion ensures that the mean is evaluated over the same set of informants at each center, there can be a substantial loss of sample size. There are a number of other methods for handling the missing observations in the MR model (e.g., mean imputation, regression imputation) but these are plagued by their own difficulties and are beyond the scope of this paper.

For the MIF model of participation, we report estimated effects and *p*-values of the ratings of culture for each informant and measure of culture (individual contrast test). We also test whether the effects of the predictor variables vary across informants or whether the same coefficient can be used for all informants (interaction test). If there is insufficient evidence to conclude that the effects of culture vary across informants, we refit the MIF model without the corresponding informant type-culture interaction to estimate a common (i.e., "pooled") effect. We include only the developmental, hierarchical, and rational components of culture as predictor variables because group culture is determined by these. The associated parameter estimates are interpreted as the effect of a one-unit increase in that component of culture and a one-unit decrease in group culture. The effect of a one-unit increase in group

Table 2B: Center Characteristics for Collaborative and Control Centers

<i>Center-Level Variables</i>	<i>Collaboratives (n = 44)</i>	<i>Controls (n = 21)</i>
Computer available	68%	76%
Number of sites	6.3 (6.7)	4.5 (3.1)
Total number of users	14,949 (10,119)	17,077 (12,407)
Urban	50%	57%

Note: The mean and standard deviation are given for continuous or interval-valued variables.

have computer availability (76 versus 68 percent), had higher annual patient loads (about 17,000 versus 15,000), and were slightly more likely to be located in an urban setting.

The mean and standard deviation of the ratings made by each type of informant and the characteristics of the informants are displayed in Table 2C. Executive directors gave the highest ratings of any informant for the QI leadership rating at both collaborative and control centers. The component of culture that received the largest average rating was *group* while *developmental* had the lowest, especially in control centers. Executive directors were older, and had been employed longer than the other informants. Informants at

Table 2C: Informant Ratings and Informant Characteristics for Collaborative and Control Centers

<i>Respondent-Level Variables</i>	<i>Collaboratives (n = 117)</i>			<i>Controls (n = 59)</i>		
	<i>Executive Director</i>	<i>Medical Director</i>	<i>Medical Provider</i>	<i>Executive Director</i>	<i>Medical Director</i>	<i>Medical Provider</i>
Quality improvement						
Leadership	4.5 (0.40)	4.2 (0.53)	3.7 (0.98)	4.4 (0.40)	3.8 (1.00)	3.7 (0.81)
Culture						
Group	36 (13)	37 (17)	32 (20)	43 (18)	41 (20)	38 (18)
Developmental	25 (9.7)	20 (7.9)	16 (7.8)	19 (11)	15 (9.6)	14 (6.4)
Hierarchical	14 (8.3)	19 (9.6)	29 (17)	18 (11)	26 (23)	26 (13)
Rational	25 (9.2)	23 (11)	24 (12)	20 (7.2)	18 (11)	22 (12)
Mainsite (%)	80	83	82	95	83	81
Male (%)	39	50	39	55	78	35
Hispanic (%)	11	11	5	20	17	5
Black (%)	16	6	13	15	11	10
Age	54 (6.9)	45 (8.2)	44 (9.6)	49 (7.7)	48 (6.9)	42 (11)
Length employed	14 (7.5)	8.3 (5.6)	5.1 (4.2)	12 (10.5)	9.3 (8.5)	4.8 (4.6)

Note: The mean and standard deviation are given for continuous or interval-valued variables.

control centers were more likely to be male and Hispanic than at collaborative centers. However, the mean age, employment at primary site, and length of employment were similar for collaborative and control centers.

Association between Informants

The correlations (Table 3) between executive and medical directors of the ratings of culture were modest (around 0.3), those for executive directors and medical providers were small (less than 0.12), and those for medical directors and medical providers were variable (range 0.13–0.47). The correlations between the ratings for leadership were smaller than for culture for all three pairings of informants. The general lack of correlation (discordance) of the informants’ ratings suggests that an MIF approach has the potential to be more informative than the SIF or MR approaches.

Collaborative Participation Analysis

Table 4 compares the results obtained from the SIF and MR analyses with those for the MIF analysis. In general, the analyses led to different conclusions about the association between the culture at a health center and participation in the collaborative.

Only the MIF model identifies a significant interaction for hierarchical culture between the informants ($p = .04$), suggesting that relying on the SIF or MR models would have led to misleading conclusions; executive and medical

Table 3: Pearson Correlation Coefficients between Pairs of Informants Based on Their Assessments of Culture and Their Assessments of the Leadership Component of the QI Effort Scale

Measure	Correlations between Informants		
	Executive Director Medical Director	Executive Director Medical Practitioner	Medical Director Medical Practitioner
Culture			
Group	0.28	− 0.04	0.47
Developmental	0.31	0.07	0.13
Hierarchical	0.30	0.12	0.16
Rational	0.33	0.03	0.31
Quality Improvement			
Leadership	0.18	0.02	0.15

QI, quality improvement.

Table 4: Parameter Estimates for the Culture and Informant Effects in the Models of Participation in the CHC Collaborative

Term	<i>Single Informants Model (n = 64 Centers)</i>		<i>Mean Regression Model (n = 65 Centers)</i>		<i>Multiple Informants Model (n = 65 Centers; 176 Informants)</i>				
					<i>Informant-Specific Effects</i>			<i>Pooled Effects</i>	
					<i>p-Value</i>				
	<i>Estimate</i>	<i>p-Value</i>	<i>Estimate</i>	<i>p-Value</i>	<i>Estimate</i>	<i>Contrast</i>	<i>Interaction</i>	<i>Estimate</i>	<i>p-Value</i>
Developmental									
Executive director	0.072	.050	0.151	.017	0.053	.052		0.061	.011
Medical director					0.050	.202	.787		
Medical provider					0.088	.059			
Hierarchical									
Executive director	−0.012	.780	−0.016	.698	−0.026	.395			
Medical director					−0.019	.337	.040		
Medical provider					0.035	.118			
Rational									
Executive director	0.043	.302	0.071	.149	0.063	.050		0.029	.107
Medical director					0.030	.265	.472		
Medical provider					0.012	.665			

In the MIF model, we first test whether the effects of culture are different among informants. If the p -value of the culture-informant interaction test is significant at the 0.05 level, we report separate effects for each informant. If the differences are not significant, we reestimate the model with the effect of that component of culture constrained to be the same for all informants (the resulting “pooled” test is displayed in the far right column). Although the models included other control variables (computer available, primary site, urban location, male, Hispanic, black, age, and length employed) these were not significant and so are not presented. Effects significant at the 0.05 level are in bold font.

CHC, community health centers.

directors have negative effects and providers have a positive effect. The SIF and MR models find nonsignificant results for hierarchical culture, the latter likely because the effects for different informants offset each other. The SIF model and MR models also fail to find significant results for rational culture, whereas the MIF model obtains a borderline significant result for executive directors ($p = .05$). However, the three methods obtain similar results for developmental culture, where, because the effects for developmental culture in the MIF model do not vary significantly across informants, they are pooled. The pooled effect under the MIF model finds the strongest evidence of an overall effect of developmental culture ($p = .011$), followed by the MR model ($p = .017$), and the SIF model ($p = .050$).

Table 5: Parameter Estimates for the Models Relating the Leadership Component of the QI Effort Scale to Informant and Center Characteristics

<i>Term</i>	<i>Single Informants Model (n = 64 Centers)</i>		<i>Mean Regression Model (n = 65 Centers)</i>		<i>Multiple Informants Model (n = 65 Centers; 176 Informants)</i>	
	<i>Estimate</i>	<i>p-Value</i>	<i>Estimate</i>	<i>p-Value</i>	<i>Estimate</i>	<i>p-Value</i>
Informant						
Executive director					0.540	.151
Medical director					0.412	.286
Medical provider					1.494	.008
Computer available	− 0.403	.058	− 0.362	.103	− 0.370	.033
Collaborative	− 0.254	.235	− 0.403	.073	− 0.275	.112
Collaborative × computer	0.373	.124	0.552	.031	0.429	.032
Primary site	0.047	.742	− 0.047	.744	0.072	.514
Urban	0.068	.527	0.073	.529	0.049	.573
Male	− 0.217	.029	− 0.070	.621	− 0.174	.025
Hispanic	− 0.080	.558	− 0.053	.804	− 0.045	.691
Black	0.010	.941	0.209	.220	− 0.011	.922
Age	0.000	.966	0.000	1.000	0.001	.865
Length employed	0.002	.767	0.007	.476	0.001	.853

The MIF model is the only model that estimates separate effects for each informant and so is the only model for which the main effect of each informant can be evaluated. The interaction effects between informant and the other control variables were not significant and so are not presented. Effects significant at the 0.05 level are in bold font.

QI, quality improvement.

QI Effort (Leadership) Analysis

In the MIF analysis, we found no significant interactions between any of the center characteristics (the variables of primary interest) and informant type. Therefore, all the effects for the MIF model in Table 5 are pooled effects. The MIF model finds that of centers participating in the collaborative, the leaders of those with computers make more effort to improve quality than the leaders of those that do not have computers ($p = .03$). The reverse is true of centers that do not participate in the collaborative. This effect is also detected by the MR model but not by the SIF model. The MIF model also finds that male informants tended to give lower ratings of leadership, an effect that was also detected by the SIF model but not by the MR model.

DISCUSSION

In this paper, we compared three approaches for analyzing data from multiple informants. Two of the approaches make restrictive assumptions. The SIF model assumes that the chosen informant (in our case, the executive director) provides all relevant information. The MR model assumes that the informants' ratings are exchangeable (i.e., randomly distributed about some common value) and can thus be summarized by averaging. Methods that focus on a single informant or the mean of the informants' ratings provide no insight into whether combining information from all informants enables better understanding of the constructs being studied, or whether analyzing certain informants alone will lead to results that are substantially different from those for other informants. In contrast, the MIF models provide a general pathway for how to use information from multiple informants in analyses. MIF models allow the researcher to make a more informed decision on whether a single informant dominates (e.g., is by far the strongest predictor in the model) to the extent that results should be based on them alone and also allows for averaging data among informants when it is appropriate to do so (e.g., when the interaction effects with the informant are not significant). When the informants' effects are significantly different, there may be no "correct" response because the informants may be attuned to different characteristics of the center. However, recognizing this alerts the analyst to the possibility that a multidimensional trait may exist and provides the opportunity to seek informants that can provide information about each dimension. The SIF and MR models are reduced cases of the MIF model that are only appropriate in special circumstances.

Our multiple informants' analysis of CHC participation, found that the effect of hierarchical culture differed significantly across informants. The effects for executive directors (negative) and providers (positive) differed significantly, suggesting that these informants in particular may draw on different experiences when assessing culture. The medical provider had the largest effect of hierarchical culture on participation, implying that they might be the single most useful informant in terms of discriminating between centers (arguing against our use of the executive director as the informant of choice). The lack of significance for the SIF model is attributed to a lack of sample size when only the executive directors are analyzed, while the lack of significance for the MR model is attributed to the effects of the individual informants canceling each other. This example illustrates the danger of relying on a single informant or using MR models when data from multiple informants can be obtained or

are available. For rational culture, the MIF model reveals that although there is a significant result for executive directors, some caution should be exercised because results for the other informants are not significant. The SIF and MR models do not give any indication of the potential differences in effects across the informants.

In addition to obtaining results that are qualitatively more informative by revealing differences between informants, when it is appropriate to pool effects across informants the MIF model typically yields results that are more precise than traditional methods. This was illustrated by developmental culture in the collaborative participation analysis where only the MIF model found a significant result. Another example is the effect of computer availability on leadership for participating centers. Although all three models obtained similarly sized estimates, only the estimate based on the MIF model was significant. There are two reasons why the MIF model yields a more precise inference: (1) it is not restricted to a single informant and does not exclude centers for which some informants did not respond or were not available (approximately one-third of centers); and (2) it does not weight each informant equally as in the MR model but rather accounts for differences in the informants' ratings, yielding a more precise estimate.

A limitation of the analyses discussed herein is that we ignored the problem of missing data. Although, the MIF model takes into account all available observations, it implicitly assumes that missing data are missing completely at random. If this assumption is violated, the results have the potential to be biased or less precise. However, the MIF model can easily be extended to account for missing data (e.g., using multiple imputation [Rubin 1987]) under fairly general conditions. Although methods for missing data can also be applied to the SIF and MR models, the MIF model provides a more complete and informative framework.

A disadvantage of the MIF model relative to the SIF and MR models is that in order to incorporate the results from all informants, it makes additional assumptions about the distribution of the observations (e.g., their correlation structure, the distribution of any random effects) and thus is more reliant on the validity of the model than the SIF and MR models, which only involve one observation per center. However, the greater generality of the MIF analysis might outweigh the increased sensitivity to model assumptions in many situations.

Because the MIF model uses all the information in the data, it has the potential to lead to increased precision. From a design perspective, the MIF model encourages sampling more informants at each center. However, it is

important to note that sampling more informants does not automatically justify sampling fewer centers. The power of the MIF design depends on the poolability of the informants; a substantial increase in power relative to the single informant design occurs only when the effect of interest is the same or similar for all informants. Realistic and preferably conservative assumptions about the poolability of the effects across informants should be used when determining the appropriate sample sizes for a study involving multiple informants.

Finally, we note that an alternative way of modeling an independent variable for which there are data from multiple informants is to include covariates for each informant in a single regression equation. Although this is a fine approach for obtaining a predictive model of the outcome, including multiple measures of the same independent variable in one regression equation may obscure the association between the construct being measured and the dependent variable (Horton, Laird, and Zahner 1999) and may be limited by degrees of freedom/power issues.

CONCLUSION

The main objective of this paper is to increase awareness of the most appropriate statistical methods for analyzing data from multiple informants. We have shown that the MIF models produce a richer set of results than the standard (SIF and MR) models. First, we identified that the relationship of hierarchical culture with participation in a collaborative was respondent dependent. When the ratings of the informants are discordant, the SIF and MR models are not reliable. In cases where effects did not vary across informants (e.g., computer availability), the MIF model found significant effects not detected by the standard methods. It is important that health services researchers are aware of the pitfalls of standard methods for analyzing multiple informant data and that these can be overcome by using an MIF model.

ACKNOWLEDGMENTS

This research was supported by grant number 1 U01 HS13653 from the Agency for Healthcare Research and Quality and The Commonwealth Fund grant number 20030185. The authors thank Barbara J. McNeil, M.D., Ph.D., and Alan Zaslavsky, Ph.D., for helpful reviews of the manuscript and

suggestions, and Thomas Keegan, Ph.D., Yang Xu, and Mary Ly for help with data collection.

Disclosures: None.

Disclaimers: None.

REFERENCES

- Bollen, K. A., and P. Paxton. 1998. "Detection and Determinants of Bias in Subjective Measures." *American Sociological Review* 63: 465–78.
- Dewald, W. G., J. G. Thursby, and R. G. Andersen. 1986. "Replication in Empirical Economics: The Journal of Money, Credit and Banking Project." *American Economic Review* 76: 587–603.
- Goldwasser, M. A., and G. M. Fitzmaurice. 2001. "Multivariate Linear Regression Analysis of Childhood Psychopathology Using Multiple Informant Data." *International Journal of Methods in Psychiatric Research* 10: 1–10.
- Gould, S. J. 1981. *The Mismeasurement of Man*. New York: W. W. Norton.
- Horton, N. J., and G. M. Fitzmaurice. 2004. "Tutorial in Biostatistics: Regression Analysis of Multiple Source and Multiple Informant Data from Complex Survey Samples." *Statistics in Medicine* 23: 2911–33.
- Horton, N. J., N. M. Laird, J. M. Murphy, R. R. Monson, A. M. Sobol, and A. H. Leighton. 2001. "Multiple Informants: Mortality Associated with Psychiatric Disorders in the Stirling County Study." *American Journal of Epidemiology* 154 (7): 649–56.
- Horton, N. J., N. M. Laird, and G. P. Zahner. 1999. "Use of Multiple Informant Data as a Predictor in Psychiatric Epidemiology." *International Journal of Methods in Psychiatric Research* 8: 6–18.
- HRSA. 2004. "Health Disparities Collaboratives" [accessed on September 21, 2004]. Available at <http://www.healthdisparities.net/>
- Lash, T. L., S. S. Thwin, N. J. Horton, E. Guadagnoli, and R. A. Silliman. 2003. "Multiple Informants: A New Method to Assess Breast Cancer Patients' Comorbidity." *American Journal of Epidemiology* 157 (3): 249–57.
- Leahey, E. 2004. "The Role of Status in Evaluating Research: The Case of Data Editing." *Social Science Research* 33: 521–37.
- Leahey, E., B. Entwisle, and P. Einaudi. 2003. "Diversity in Everyday Research Practice: The Case of Data Editing." *Sociological Methods and Research* 32: 63–89.
- Liang, K.-Y., and S. L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73: 13–22.
- Pepe, M. S., R. C. Whitaker, and K. Seidel. 1999. "Estimating and Comparing Univariate Associations with Application to the Prediction of Adult Obesity." *Statistics in Medicine* 18: 163–73.
- Raudenbush, S. W., and A. S. Bryk. 2003. *Hierarchical Linear Models: Applications and Data Analysis Methods*. California: Sage Publications Inc.

- Rubin, D. B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.
- Shortell, S. M., J. L. O'Brien, J. M. Carman, R. W. Foster, E. F. X. Hughes, H. Boerstler, and E. J. O'Connor. 1995. "Assessing the Impact of Continuous Quality Improvement/Total Quality Management: Concepts versus Implementation." *Health Services Research* 30: 377-401.
- Wagner, E. H., B. T. Austin, and M. Von Korff. 1996. "Organizing Care for Patients with Chronic Illness." *Milbank Quarterly* 74: 511-44.
- White, H. 1982. "Maximum Likelihood Estimation of Miss-Specified Models." *Econometrika* 50: 1-26.
- Zammuto, R., and J. Krakower. 1991. "Quantitative and Qualitative Studies of Organizational Culture." In *Research in Organizational Change and Development*, edited by W. Woodman and W. Passmore, pp. 83-114. Greenwich, CT: JAI Press.
- Zeger, S. L., and K.-Y. Liang. 1986. "Longitudinal Data Analysis for Discrete and Continuous Outcomes." *Biometrics* 42: 121-30.