Journal of Official Statistics, Vol. 21, No. 1, 2005, pp. 39–51

A Two-Step Nonparametric Sample Survey Approach for Testing the Association of Degree of Rurality with Health Services Utilization

John S. Preisser[1], *Cicely E. Mitchell*[2], *James M. Powers*[1], *Thomas A. Arcury*[3], *and Wilbert M. Gesler*[4]

Cross-sectional population surveys designed to identify factors associated with health services utilization may record data at multiple levels such as characteristics of individuals and geographical areas. In the Mountain Accessibility Project, a primary aim was to determine if a county-level categorical variable, degree of rurality, was associated with health services usage, as measured by the proportion of inhabitants in a county who reported a regular care visit to a health care practitioner in the previous year. A total of 1,059 adults from twelve counties in western North Carolina were interviewed and individual-level covariate data were collected. Exact tests for nonparametric statistics applied to county-level summaries provided superpopulation inference for the assessment of the association of degree of rurality with health services utilization. Motivated by hypothesis testing procedures used in randomized community trials, the two-step analysis approach employs covariate adjustment procedures using survey logistic regression for individual-level data.

Key words: Exact methods; logistic regression; Spearman correlation.

1. Introduction

Cross-sectional population surveys designed to identify factors associated with health services utilization often measure characteristics of both individuals and geographical areas. These studies often have multiple objectives. For example, the Mountain Accessibility Project (MAP), a study funded by the United States Agency for Healthcare Research and Quality, employed a sample survey in twelve western North Carolina counties in 1999 for the broad goal of identifying factors related to health services usage in a rural area. MAP employed a stratified three-stage cluster probability sample design. Within each of twelve counties or strata, U.S. Census area segments were sampled (primary sampling units), followed by households in the second stage, and adults in the third stage (observational units). The primary (i.e., confirmatory) hypothesis of MAP was to assess whether degree of rurality, a county-level variable, was associated with health services

[1] University of North Carolina, Department of Biostatistics, CB #7420, Chapel Hill, NC 27599-7420, U.S.A. Email: jpreisse@bios.unc.edu and jpowers@bios.unc.edu
[2] PPD, Inc., Research Triangle Park, NC 27560-7200, U.S.A. Email: cicely.mitchell@rtp.ppdi.com
[3] Wake Forest University School of Medicine, Department of Family and Community Medicine, Winston-Salem, NC, 27157-1084, U.S.A. Email: tarcury@wfubmc.edu
[4] University of North Carolina, Department of Geography, Chapel Hill, NC, 27599-3220, U.S.A. Email: wgesler@email.unc.edu
Acknowledgements: We thank Sally Nash, Jill Sherman, and John Spencer for their role in data quality assurance and Gary G. Koch for his valuable comments. This research was supported by grant R01 HS09624 from the Agency for Healthcare Research and Quality.

utilization. Using a two-step nonparametric sample survey analysis approach, inference is directed at a conceptual superpopulation of rural counties that the twelve sampled counties are considered to represent.

The two-step analysis approach follows from the superpopulation inferential perspective adopted, and is different in important ways than the analysis approach adopted for the secondary hypotheses of MAP (Woods et al. 2003; Arcury et al. 2005) that follow from classical sampling theory. These secondary or exploratory hypotheses address the effects of person-level geographical, sociodemographic, cultural and health status characteristics on health services utilization among adults in the twelve-county region. Taking a finite population inferential perspective, the twelve-county region is considered as the population of interest, county is managed as a stratification factor, and census segment is the primary sampling unit from which individual households are drawn. This contrasts with the superpopulation perspective of the primary hypothesis in which the county is considered as the primary sampling unit, even though the set of twelve counties are not actually sampled from a larger population of counties.

This article describes the statistical methods used to test the primary hypotheses involving the assessment of the effects of the degree of rurality, a county-level variable, on health services utilization. Since it is possible that differences in health services utilization between counties with different levels of rurality may be due to imbalances in individual-level covariates, covariate adjustment is employed. We use a general statistical strategy for producing both unadjusted and adjusted tests based upon an approach used in COMMIT, a community intervention trial for smoking cessation. The COMMIT Research Group sought to study the effects of a four-year community level intervention on smoking cessation rates (COMMIT 1995; Gail et al. 1996; Green et al. 1995). To compare intervention and control groups, they applied permutation tests to observed community-specific proportions of smoking cessation. Adjustment for individual-level covariates was achieved with a two-step procedure that applied similar tests to community statistics that were the averages of residuals from a multivariable regression of cessation on those covariates.

Similar covariate adjustment methods may be used in survey data analysis when interest is in the association of an exposure or explanatory variable measured at the geographical unit level with a health usage outcome measured at the level of the individual. One feature of the analysis of MAP data not found in the COMMIT trial is the use of survey sampling weights. The first step of the analysis presented in this article applies survey logistic regression to a health services utilization outcome, incorporating household-level sampling weights and adjusting for individual sociodemographic and health status characteristics. The second step computes Wilcoxon Rank Sum or Spearman Correlation Statistics to the averages of residuals across individuals in a county to assess the relationship of degree of rurality and health services utilization. Because the number of counties is small, exact methods provide tests of significance for these associations.

2. Study Design and Variable Definitions

2.1. Study design

Study investigators gathered individual and household information in the Mountain Area Health Education Center (MAHEC) region, located in the Appalachian mountains of western North Carolina (Figure 1). For most residents of this region, distances to urban places of any size are relatively large, roads are winding and difficult for travel in inclement weather, and amenities are few. The study counties permit us to evaluate the following alternative hypotheses about health care utilization in some of the most isolated communities in the United States relative to their null counterparts: (i) individuals in more rural counties have less utilization of health services; (ii) individuals in more rural counties have less utilization of health services after adjusting for individual-level socio-demographic and health status variables.

A sample of 1,059 adults who resided in the twelve counties in the MAHEC study area were interviewed. A stratified three-stage cluster sampling design was used to choose a person for an interview. In the first-stage, small-size U.S. Census area segments containing approximately fifteen housing units were selected with probability proportional to size in order to achieve an equal probability of selection at the household-level. Area segments were stratified by county and ethnicity based upon classifying segments in each county as minority or nonminority. A high concentration of African American headed households (about 44.5%) comprised the minority-classified segments. There were six counties with minority segments. Twenty segments were allocated to each county. Five additional segments were allocated to those counties with segments classified as minority so that African American households tended to be sampled at a higher rate than nonminority households. In the second stage, four housing units were selected with equal probability of selection within each selected segment. In the final stage, one adult was selected from each selected household for an interview with equal probability of selection. For the purposes of this research, an adult is defined as a person age 18 or over. Of the 73% of adults that were eligible to interview, the response rate was 84%. The final sample size was 1,060 households; however, one household provided a child interview only. The 1,059 adult interviews analyzed each had a sampling weight that included components corresponding to each stage of sampling, a first-stage unit selection weight reflecting the probability of

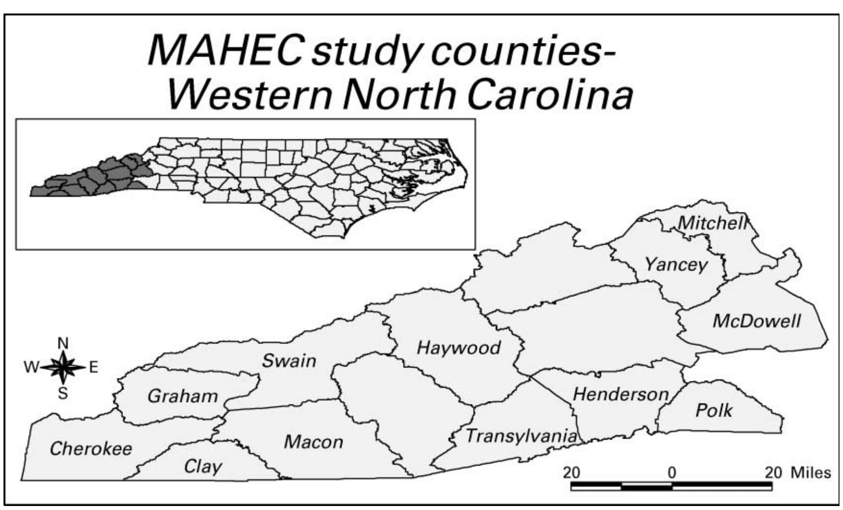

Fig. 1. Mountain area health education center study area

selecting the area segment within the given county, the conditional housing unit weight reflecting the probability of selecting a housing unit, and the conditional weight of the respondent as a multiplier based on the number of adults on the housing unit roster. Further adjustments to weights were made for several types of nonresponse. Sample selection, construction of analysis weights and data collection, via in-person interviews, were completed by Research Triangle Institute (2000).

2.2. *Variable definitions*

The twelve counties were all rural areas as each is characterized by one of four different levels of rurality, as measured by Beale codes (Butler and Beale 1994). The levels of Beale code are integers from 0 to 9 based on urban population size, adjacency to a metropolitan area, and degree of rurality. Beale 0 classifies a more urban county, and Beale 9 classifies an extremely rural county. MAP included only the most rural areas, those with Beale 6, 7, 8, or 9. Twelve of thirteen eligible counties in the region were selected to give a 4-2-2-4 design; four counties (Haywood, Henderson, McDowell, and Transylvania) were classified as Beale 6 (urban population of 2,500 to 19,999 residents, adjacent to a metropolitan area); two counties (Cherokee and Macon) were classified as Beale 7 (urban population of 2,500 to 19,999 residents, not adjacent to a metropolitan area); two counties (Polk and Yancey) were classified as Beale 8 (completely rural population or an urban population of less than 2,500 residents, adjacent to a metropolitan area); and four counties (Clay, Graham, Mitchell, and Swain) were classified as Beale 9 (completely rural population or an urban population of less than 2,500 residents, not adjacent to a metropolitan area). From the Beale code, three county-level independent variables are of interest: urban population size (high, Beale 6 or 7 versus low, Beale 8 or 9), adjacency (yes, Beale 6 or 8 versus no, Beale 7 or 9), and degree of rurality considered to be ordinal (low, Beale 6, versus moderate, Beale 7 or 8, versus high, Beale 9).

The variables based upon Beale code were prespecified during the design phase of the study as part of the *a priori* primary hypotheses. The dependent variable is a binary outcome indicating whether a resident made any visit for regular care (i.e., check-up visit) to a health care practitioner such as a doctor, nurse, physician's assistant, or nurse practitioner in the year preceding the interview. Other dependent variables, any visit for chronic care, and any acute care visit, also part of the primary hypotheses of MAP, are not discussed here. The individual-level covariates reported on in this article are age, gender, tobacco use (yes/no), the number of reported chronic conditions, SF-12 mental and physical health scores (Ware, Kosinski, and Keller 1996), and type of insurance (private, public, private and public, or none). For the SF-12 scales, higher scores indicate better health. In total, more than two-dozen individual-level variables were screened for inclusion in the final models.

3. Statistical Methods

There are different analyses that might have been chosen to test the confirmatory hypothesis in MAP. MAP investigators chose an approach according to a superpopulation perspective that was deemed appropriate for a confirmatory hypothesis that was central to the primary aim of the project. Before describing the two-step nonparametric approach that was employed for testing the primary hypothesis, a statistical analysis arising from the finite population perspective is described. Contrasting the approaches, each of which is valid in its own context, will help provide a rationale for the two-step approach of MAP.

3.1. A finite population perspective

The analysis that takes a finite population perspective employs survey logistic regression fitted to individual-level data for the binary outcome of any regular care visit. In this analysis approach, inference is directed at the twelve-county region defined as the finite population and represented by the individuals in the survey sample. Estimation of the survey regression model and construction of variance estimates follows from the stratified three-stage cluster sampling design. Using Taylor series linearization, variance estimates for regression coefficients are constructed using counties as strata and census segments as primary sampling units. Survey weights equal to the inverse of an individual's probability of selection into the sample are used in both regression model parameter estimation and variance estimation. A model selection procedure is employed to identify individual-level covariates for inclusion in the final model that contains county indicator variables. Contrasts among the counties' indicator variables are formed to test hypotheses concerning the county-level variable degree of rurality.

While the survey logistic regression modeling approach is generally compatible with the survey design, treating county, or a county-level variable, as the primary exposure variable of interest in the model confuses the role of county since it is the stratification factor in the survey design. It is somewhat awkward to test hypotheses about a county-level variable when the primary sampling unit is defined at a lower level, here the census segment. The implication in MAP, for example, is that the statistical power of the test for the county-level variable urban population size, holding the number of individuals and the number of primary sampling units (census segments) fixed, is not much affected by the number of counties in the sample design. This may be appropriate when rurality is mainly viewed as a characteristic of the individuals in the counties that are being directly studied. The problem with the classical finite population approach for testing the primary hypothesis of MAP is that it is not clear if an observed difference between the group of low versus the group of high urban population counties is due to rurality or simply a chance occurrence due to inherent county-to-county variability. In order to assess the potential role of rurality it is necessary to assess the difference in county groups with respect to the county-to-county variability. This is achieved in analyses that treat county as the primary sampling unit. Because the large sample size requirements of survey regression methods are not met for this alternative inferential framework, a two-step analysis approach is adopted.

3.2. A superpopulation inferential perspective

To provide more substantive inference for the county-level variable of rurality, we go beyond the sampling framework employed in the MAP design and view the survey sample selected in MAP to be a representative sample from some superpopulation of rural counties. Due to the small number of counties, exact inference (Agresti, Mehta, and Patel

1990; Agresti 1992) is applied to nonparametric statistics for evaluation of the research hypothesis that greater rurality is associated with less health services usage. Rurality is measured by the county-level variables of population size, adjacency status, and degree of rurality. The association of each with the county proportion of health services utilization was summarized by the Spearman rank correlation statistic, as an estimate of the true superpopulation correlation ρ, using the SAS PROC FREQ procedure with the exact SCORR option, and its statistical significance determined from its exact distribution (Stokes, Davis, and Koch 2000; SAS Institute 1999). For the dichotomous factors (population size and adjacency), this statistic reduces to the Wilcoxon Rank Sum statistic. As the pre-planned alternative hypothesis was that greater rurality is associated with less health services utilization, a one-sided 0.05 significance level was used. The rurality measures were reverse scaled, giving the more urban counties a higher score, so that a positive correlation (and not a negative correlation) is consistent with the alternative or research hypotheses. The exact *p*-value is determined by summing probabilities of those tables in a reference set (given by conditioning on observed marginal totals) with a value of the Spearman test statistic at least as large as the value observed.

Three hypothesis tests corresponding to the three measures of rurality (urban population size, adjacency, and degree) were applied to the twelve-county summary statistics. The first unadjusted test, $H_0: \mu_{6,7} = \mu_{8,9}$, $H_1: \mu_{6,7} > \mu_{8,9}$, assesses whether the median proportion of health services utilization of counties with a low urban population size is smaller than that of counties with a high urban population. The second unadjusted test, $H_0: \mu_{6,8} = \mu_{7,9}$, $H_1: \mu_{6,8} > \mu_{7,9}$, assesses whether the median proportion of health services utilization of counties not adjacent to metropolitan areas is smaller than that of counties that are adjacent to metropolitan areas. The third unadjusted test, $H_0: \rho = 0$, $H_1: \rho > 0$, assesses whether counties with a higher degree of rurality have a smaller proportion of health services utilization.

Because the research study is observational and does not involve an assigned treatment, it was possible that an observed association between rurality and health services utilization was the result of differences among individuals across the twelve counties. Therefore, adjusted tests for rurality were performed, making adjustments for imbalances in individual-level covariates. The procedures for the adjusted tests were similar to those used in COMMIT; however, in MAP, rurality, instead of intervention, was the county-level variable of interest. Individual-level covariates were identified as significant by initially studying each possible covariate one at a time, with significance testing at the $\alpha = 0.10$ level based on a logistic model that used weights and that calculated Taylor series variance estimates to account for the stratified three-stage sample survey design. In this model, a single term was included for the covariate plus a separate intercept for each of the twelve counties. Next, the full model was fit using all of the significant covariates from the initial step. While county was again defined as a stratification factor for the purpose of variance estimation, the county intercepts were not included in the full model in order to predict the outcome under the null hypothesis of no rurality effect. County could not be included in the model since it is aliased with rurality and would therefore be accounting for a rurality effect. A backwards elimination model selection procedure was used to remove covariates not significantly associated when adjusted for other variables in the model. The variable's *p*-value had to be less than 0.05 for the variable to stay in the model. Once the final model was determined, the weighted average of predicted values for each county were obtained. Next, residuals were computed by subtracting the weighted average of predicted values for each county from the observed weighted proportion of heath services utilization. The same exact tests as for the unadjusted tests were performed, but the averaged residuals were used rather than the observed proportions.

One caveat of the adjustment procedure is that the designation of census segments as primary sampling units may lead to liberal inclusion of covariates, but this is an acceptable if not desirable feature for the purpose of identifying a pool of individual-level covariates for which to adjust. Choice of primary sampling unit in Step 1 does not affect the estimated coefficients or final result of the adjusted test given the choice of individual-level variables. This is because the nonparametric adjusted tests are based only upon the predicted values of utilization from the survey regression, and not their standard errors.

The two-step nonparametric sample survey approach can be viewed as a variation of nonparametric covariance analysis for sets of contingency tables (Preisser and Koch 1997). The more traditional application applies to independent data and provides a means of extending Mantel-Haenszel methods to adjust for additional covariates.

4. Results

The overall estimated proportion of western North Carolina residents who had a regular care visit in the year preceding the interview was 0.49, with a standard error of 0.03. Table 1 lists the estimated proportions of any regular care visit by county. These are weighted estimates and their design-based standard errors computed using Taylor expansion techniques with SUDAAN software (Research Triangle Institute 2001; Korn and Graubard 1999). Each adult had a survey-based weight, discussed in Section 2, representing a corresponding number of adults in the population. These results suggest that rurality may be related to health care utilization, since the two counties with the highest estimates of utilization were among the least rural (Beale = 6) counties.

SUDAAN software was used to obtain the results of the finite population-based survey logistic regression analysis (Table 2). The final model results reveal that higher resident

Table 1. Estimated population proportion (p) of residents who made a regular care visit

County	Beale	*n*	*P* (se)	rank	adjusted *p*	adjusted rank
Haywood	6	89	0.459 (0.079)	9	0.488	6
Henderson	6	115	0.475 (0.077)	7	0.447	9
McDowell	6	97	0.597 (0.073)	1	0.590	1
Transylvania	6	85	0.570 (0.061)	2	0.533	3
Cherokee	7	91	0.461 (0.057)	8	0.452	7
Macon	7	75	0.486 (0.070)	5	0.453	8
Polk	8	91	0.407 (0.053)	12	0.392	12
Yancey	8	82	0.564 (0.068)	3	0.541	2
Clay	9	81	0.408 (0.061)	11	0.429	11
Graham	9	90	0.424 (0.076)	10	0.426	10
Mitchell	9	81	0.485 (0.063)	6	0.498	4
Swain	9	82	0.512 (0.077)	4	0.495	5
Total		1059	0.49 (0.03)			

Table 2. Survey logistic regression parameter estimates (standard errors) for any regular care visit

Variable		Estimate	Standard error	p-value
Intercept		−4.06	0.89	<0.001
County (Beale)	Haywood (6)	0.24	0.44	0.58
	Henderson (6)	0.13	0.42	0.76
	McDowell (6)	0.76	0.41	0.06
	Transylvania (6)	0.65	0.37	0.08
	Cherokee (7)	0.16	0.37	0.68
	Macon (7)	0.12	0.39	0.77
	Polk (8)	−0.15	0.34	0.66
	Yancey (8)	0.47	0.39	0.23
	Graham (9)	−0.01	0.44	0.99
	Mitchell (9)	0.29	0.37	0.44
	Swain (9)	0.28	0.47	0.55
	(reference: Clay (9))			
Age	45–64	0.38	0.23	0.019
	65 and over	1.13	0.41	
	(reference 18–44)			
Female gender		0.51	0.22	0.023
Tobacco use		−0.44	0.21	0.033
Type of insurance	Private	0.93	0.30	0.002
	Public	0.81	0.41	0.049
	private and public	0.25	0.53	0.64
	(reference: none)			
Mental health status scale		0.27	0.13	0.036
Physical health status scale		0.18	0.10	0.091
Number of chronic conditions		0.15	0.06	0.012

age, female gender, not using tobacco, having health insurance, better mental health and poorer physical health (as indicated by a greater number of chronic conditions) are positively associated with any regular care visit. The overall Wald test for county based upon 11 degrees of freedom is not statistically significant ($p = 0.43$). However, given the adoption of the finite population approach, the planned analysis uses Wald tests for pre-specified contrasts among the county coefficients. Two-sided tests reveal that the proportion of any regular care visit in the preceding twelve months did not significantly vary among county groups classified according to urban population size ($p = 0.27$), adjacency ($p = 0.23$), nor degree or rurality ($p = 0.18$).

Estimated proportions by each measure of rurality, unadjusted for individual-level factors, are listed in Table 3. When the counties were grouped by urban population size, counties with low urban population sizes had a lower estimated proportion of any regular care visit than counties with high urban population sizes. When the counties were grouped by adjacency, counties not adjacent to a metropolitan area had a lower estimated proportion of any regular care visit than counties adjacent to a metropolitan area. Finally,

Table 3. Estimated adjusted (r_a) and unadjusted (r_u) Spearman correlations and population proportions (p) of any regular care visit by population size, adjacency, and degree of rurality

	n	p (se)	r_u (p-value*)	r_a (p-value*)
Urban population				
High	552	0.50 (0.03)	0.339 (0.144)	0.241 (0.242)
Low	507	0.48 (0.03)		
Adjacency				
Yes	559	0.50 (0.04)	0.291 (0.186)	0.290 (0.197)
No	500	0.47 (0.03)		
Degree				
Low	386	0.50 (0.04)	0.386 (0.110)	0.325 (0.160)
Moderate	339	0.48 (0.03)		
High	334	0.46 (0.04)		

*one-sided p-value

when the counties were grouped by the degree of rurality, there was a trend for less health care utilization with greater degree of rurality.

In order to conduct superpopulation inference for the primary hypothesis of MAP, counties are treated as if they were the primary sampling unit. To apply the two-step approach, a model like the one in Table 1 is fit, but without the county-level indicator variables. Parameter estimates for the individual-level factors (not shown) are not very different from those in Table 2. Averaging residuals from the model across all individuals within a county gives the sets of twelve-county summary statistics to which exact tests are applied. To illustrate the effect of covariate adjustment, the overall mean of 0.49 is added to each county residual and presented as the adjusted proportion in Table 1. There is some difference in the adjusted ranks (ranks of the model adjusted proportions) as compared to the unadjusted ranks (ranks of the observed county proportions). These changes coincide with differences seen in Table 3 between the unadjusted Spearman correlation of the county summaries with the rurality measures as compared to the Spearman correlations based upon the adjusted summary scores. For all tests, unadjusted and adjusted, the null hypothesis of no rurality effect versus the alternative hypothesis that less rural counties have more health services utilization was not rejected. Thus, there is insufficient evidence to support the research hypothesis that individuals in more rural counties utilize less health services.

5. Discussion

In the MAP study, a combination of sample survey estimation of summary county measures and exact testing procedures applied to Spearman Rank statistics provided superpopulation inferential assessment of the relationship of degree of rurality with the health services utilization in western North Carolina. In this point of view, the sample of counties is considered to be like a random sample of counties from a larger, or super, population. The driving principle behind the analysis and inferential perspective is the belief that rurality, a county-level variable, should be assessed with respect to county-to-county variability. The classical finite population survey regression approach that treats county as fixed effects fails to disentangle the between-county variation from the effect of degree of rurality. It has been noted that in nested designs, the problem of inflation of Type I error occurs when dummy variables are used to represent cluster-to-cluster variation (Zucker 1990). In an analogous way, the classical finite sample survey approach may lead to the conclusion of a difference even in the absence of a degree-of-rurality effect. As discussed in Section 3.1, such differences found with this approach lead to more limited interpretations than differences found with a superpopulation approach.

A popular method for the analysis of health data that accounts for cluster-to-cluster variability is given by hierarchical models, and for binary outcomes, in particular, the logistic-normal mixed model (Rice and Leyland 1996). In this approach, cluster-level fixed effects, such as degree-of-rurality, are assessed with respect to the cluster-to-cluster variation accounted for by the specification of counties as random effects, and through a statistical distribution (eg., normal) assumed for the random effects. The strength of this approach is its flexibility to handle both individual and cluster-level covariates, and its greater statistical power relative to summary statistic approaches. However, the gain in statistical power may be at the cost of robustness of results to assumptions made about the distribution of random effects when the number of clusters is small (Turner, Omar, and Thompson 2001). Furthermore, when applied to a complex sample survey setting, it discards information about the survey design, namely, the survey weights. Our two-step nonparametric approach, which may be less powerful, incorporates the survey weights into the analysis and makes somewhat fewer statistical assumptions.

Thus, a desire to incorporate the sample weights, and the particularities of the regional MAP sample survey, mainly the small number of counties sampled, led to the two-step analysis procedure and superpopulation outlook. A larger, possibly national, survey would be better suited to employ a sampling design that defined counties as primary sampling units and regions as strata. Then, a finite population perspective would follow naturally as the driving principle – assessment with respect to county-to-county variability – could be achieved with a one-step survey regression approach that incorporated these features of the survey design.

The survey design of MAP was not ideally suited for testing the primary hypothesis. Certainly, limited resources sometimes make it difficult to design a survey that can meet multiple objectives of a study with equal aptitude. Given budget considerations and the regional nature of the project, MAP study investigators decided during the planning phase that twelve counties afforded a reasonable opportunity to detect statistically significant associations between county-level rurality measures and health services utilization. This is because the one-sided exact Wilcoxon Rank Sum Test for a dichotomous factor like urban population size gives a statistically significant result, at the .05 level, when the sum of ranks for the six-member group with the smaller sum is between 21 and 28. The county-level analysis was chosen because of the uncertainty of whether any difference in county groups could otherwise (eg., with the approach of Section 3.1) be attributable to rurality and not to the chance occurrence related to inherent county-to-county variability. Because rurality is a county-level variable, its assessment in MAP was made with respect to county-to-county variability, and not with respect to the variability between census segments. Given the size of the MAP survey, the analysis fell under the purview of superpopulation inference. The analysis of data from the MAP study suggests that superpopulation inference may be an attractive sampling framework in other sample surveys where interest lies in a primary exposure variable that is at the same level as the stratification factor.

Both the superpopulation analysis and the finite population analysis discussed as alternative approaches for MAP's primary hypotheses have limitations. First, because MAP is an observational study, omission of important individual-level covariates in either approach may result in inference that fails to clarify the nature of observed differences in county proportions of utilization. Similarly, either approach may give misleading results if differences in health services utilization are attributable to unmeasured county characteristics. The finite population analysis offers a limited scope of inference for the relationship of degree of rurality and health care utilization. While the superpopulation-based two-step analysis approach allows for broad inferential statements, it requires the assumption that the twelve counties studied are representative of a larger population of similar rural counties. In effect, random sampling of counties from a superpopulation is assumed while not actually implemented. Finally, if standard errors in Table 1 had greatly varied, then perhaps Wald tests (LaVange et al. 1994) that would have taken this into account could have been used instead.

The issue of whether to treat counties as strata or primary sampling units in a survey regression model, assuming sufficient sample size would permit the latter, is analogous to the issue of whether to treat centers in a multi-center clinical trial as fixed or random effects. When treated as strata (fixed effects), centers are regarded as characteristic of individuals in the trial and inference regarding centers applies only to individuals in those centers. When treated as random effects, the centers in a clinical trial are regarded as being representative of a superpopulation of all possible centers conceptually similar to those in the study. Typically, clinical trials use a sample of convenience based upon eligibility criteria that define the superpopulation of interest. Like the MAP sample survey, clinical trials are not based upon an actual sample from the superpopulation. They are concerned with entities (eg., individuals within clusters such as patients within centers or residents within counties) thought to represent, like a random cluster sample, their population counterparts.

Additional limitations relate to the design of the MAP study itself. County-level measures of rurality other than Beale code that might have established a link between rurality and health services utilization for the most rural counties were not collected. The sample size of only twelve counties afforded only limited power for the primary hypothesis. Restricting the survey to twelve counties limits the amount of county-level information that could be included in the analyses, even if it had been available. Results of the study suggest a much larger survey involving more counties is needed to clarify the nature of the relationship between county-level rurality measures and health care utilization. A survey with more counties would also enable multiple county-level factors to be collected and included in the statistical analysis, thus overcoming a notable limitation of the MAP project.

6. References

Agresti, A. (1992). A Survey of Exact Inference for Contingency Tables. Statistical Science, 7, 131–177.

Agresti, A., Mehta, C.R., and Patel, N.R. (1990). Exact Inference for Contingency Tables With Ordered Categories. Journal of the American Statistical Association, 85, 453–458.

Arcury, T.A., Preisser, J.S., Gesler, W.M., and Powers, J.M. (2005). Access to Transportation and Health Care Utilization in a Rural Region. Journal of Rural Health, 21, 31–38.

Butler, M.A. and Beale, C.A. (1994). Rural-urban Continuum Codes for Metropolitan and Nonmetropolitan Counties, 1993. Washington, D.C.: Agriculture and Rural Economy Division, Economic Research Service, U.S. Department of Agriculture, AGES 9425.

The COMMIT Research Group. (1995). Community Intervention Trial for Smoking Cessation (COMMIT): I. Cohort Results for a Four-Year Community Intervention. American Journal of Public Health, 84, 183–192.

Gail, M.H., Mark, S.D., Carroll, R.J., Green, S.B., and Pee, D. (1996). On Design Considerations and Randomization-Based Inference for Community Intervention Trials. Statistics in Medicine, 15, 1069–1092.

Green, S.B., Corle, D.K., Gail, M.H., Mark, S.D., Pee, D., Freedman, L.S., Graubard, B.I., and Lynn, W.R. (1995). Interplay Between Design and Analysis for Behavioral Intervention Trials with Community as the Unit of Randomization. American Journal of Epidemiology, 142, 587–593.

Korn, E.L. and Graubard, B.I. (1999). Analysis of Health Surveys. New York: John Wiley and Sons.

Lavange, L.M., Keyes, L.L., Koch, G.G., and Margolis, P.A. (1994). Application of Sample Survey Methods for Modeling Ratios to Incidence Densities. Statistics in Medicine, 13, 343–355.

Preisser, J.S. and Koch, G.G. (1997). Categorical Data Analysis in Public Health. Annual Review of Public Health, 18, 51–82.

Research Triangle Institute. (2000). Data Collection Report. Geographic Accessibility of Health Care in Rural Areas Project RTI Project 7438" Research Triangle Park, North Carolina: Research Triangle Institute.

Research Triangle Institute. (2001). SUDAAN User's Manual, Release 8.0. Research Triangle Park, NC: Research Triangle Institute.

Rice, N. and Leyland, A. (1996). Multilevel Models: Applications to Health Data. Journal of Health Services Research and Policy, 3, 154–164.

SAS Institute, Inc. (1999). SAS/STAT User's Guide. Cary, NC: SAS Institute Inc., Version 8, 3884 pp.

Stokes, M.E., Davis, C.S., and Koch, G.G. (2000). Categorical Data Analysis Using the SAS System. Second edition. Cary, NC: SAS Institute, Inc.

Turner, R.M., Omar, R.Z., and Thompson, S.G. (2001). Bayesian Methods of Analysis For Cluster Randomized Trials With Binary Outcome data. Statistics in Medicine, 20, 453–472.

Ware, J.E., Kosinski, M., and Keller, S.D. (1996). A 12-item Short-form Health Survey: Construction of Scales and Preliminary Tests of Reliability and Validity. Medical Care, 34, 220–233.

Woods, C.R., Arcury, T.A., Powers, J.M., Preisser, J.S., and Gesler, W.M. (2003). Determinants of Health Care Use by Children in Rural Western North Carolina: Results From the Mountain Accessibility Project Survey. Pediatrics, 112, e143–e152.

Zucker, D.M. (1990). An Analysis of Variance Pitfall: The Fixed Effects Analysis in a Nested Design. Educational and Psychological Measurement, 50, 731–738.

Received November 2002
Revised October 2003