DHS WORKING PAPERS

Demographic and Health Surveys

The Relationship of Service Availability to Contraceptive Use in Rural Guatemala

Thomas W. Pullum

Institute for Resource Development
Macro International, Inc.

DHS Working Papers

Number 2

The Relationship of Service Availability to Contraceptive Use in Rural Guatemala

Thomas W. Pullum

IRD/Macro International, Inc.
Columbia, Maryland USA

October 1991

Thomas W. Pullum is Training Director of the Population Research Center, University of Texas, Austin, Texas. Research for this paper was conducted with the support of the Demographic and Health Surveys Program. The author is grateful for comments and suggestions received from several DHS staff members.

The Demographic and Health Surveys Program is designed to collect data on fertility, family planning, and maternal and child health. DHS is funded by the United States Agency for International Development. DHS Working Papers are intended to disseminate early findings of in-depth analysis of DHS data. Comments about this working paper or requests for information about the DHS program should be sent to: DHS, IRD/Macro International, 8850 Stanford Boulevard, Suite 4000, Columbia, MD 21045, USA (Telephone 301-290-2800; Telex 198116 WPAS UT; Fax 301-290-2999).

Abstract

The Demographic and Health Surveys (DHS) program routinely collects information about the local availability of family planning services in a separate survey in which the sample clusters are the units of analysis. This paper proposes analytical and statistical models for using this information and illustrates the models with data from rural Guatemala. In most of the analysis, clusters are the units of analysis. The number of women in the cluster, the number of current users of modern methods, and the means and proportions of socioeconomic indicators such as literacy are attached to the cluster data. Then follows an assessment of the impact, upon contraceptive prevalence, of the presence of family planning facilities, the degree of access to these facilities, their quality, and the existence of a local distribution program. It is found that both distance and travel time to facilities are quite important and that a few indicators of quality are also significant, but specific coefficients are difficult to interpret because of missing data and the substitution of facilities. When the socioeconomic environment is taken into account, the effects of service availability are substantially reduced. Travel time and distance to facilities remain critically important, but the quality indicators lose their statistical significance.

Introduction

Research to identify the determinants of fertility levels, their differentials, and changes over time has been conducted with a wide range of perspectives and kinds of data. Some of this work has dealt with the consequences of exogenous influences; some has dealt with explicit policies and interventions to alter fertility levels. In recent years, as family planning programs have become widespread in developing countries, a variety of programmatic approaches have been taken to reduce the demand for children and to improve access to family planning. This paper will focus on one small piece of a larger puzzle: the importance of variation in access to modern contraception as a determinant of the current use of a modern method. In working with this small piece of the puzzle, the larger picture will be kept in mind, and, to the extent possible with the available data, non-programmatic influences on fertility will also be taken into account.

This specific type of evaluation research can be traced back to at least two kinds of origins. One of these is a broader interest in community or contextual influences on fertility: the local community, after all, provides a social and economic environment that affects the costs of children and of fertility regulation through a wide variety of mechanisms. Beginning in the mid 1970s, Ronald Freedman and colleagues at the University of Michigan designed and analyzed survey modules, first applied in Malaysia, that measured indicators of community modernity and development. These included the presence of health facilities and clinics, schools, cinemas, and other media influences; economic centrality; the extent of electrification; and the types and diversity of economic activities.

The second point of origin, also dating from the mid 1970s, came as part of the World Fertility Survey (WFS). In selected countries, women were asked about the presence of contraceptive supplies in the home. This information, which was relatively easy to collect, was taken as a reflection of the household's access to contraception.

It is now standard procedure for the Demographic and Health Surveys (DHS) to administer an additional questionnaire in each country that will be referred to here as the Service Availability Module (SA Module). The SA Module has been piggy-backed onto the basic cluster sample design of DHS surveys. In this design, the Primary Sampling Units (PSU) are clusters of adjacent households; a specified fraction, such as a third, of all eligible women within each cluster are interviewed. The SA Module is not part of the individual questionnaire. It is a separate inquiry into conditions prevailing throughout the cluster as a whole, is administered by other trained personnel, and usually is carried out at a somewhat later date. The information is collected by a combination of discussions with local informants and actual visits to local facilities. The purposes of the SA Module are (1) to describe service availability, particularly in rural areas and in terms of quality, and (2) to assess the impact of services on contraceptive use.

The cluster usually occupies a small enough geographical area that it is reasonable to suppose that all households in the cluster have the same service environment. However, in urban areas the cluster's environment is often much more diverse than can be measured by the SA Module. For that reason this study will be limited to rural areas, and its conclusions should not be extended to urban areas.

The SA Module collects information about the local standard of living, the community's integration with the rest of the country, and the local availability of health and family planning services, including contraceptive distribution programs. Compared to the earlier WFS community surveys, much less data is collected about local development, modernity, and standard of living, and much more data is collected about the range and quality of health and family planning services.

Although the collection of such data is now routine and the SA Module has recently been expanded for the second phase of surveys (DHS-II), there has been little actual analysis of the data. There are at least two reasons for this. First, there are some technical considerations in working with multi-level data, namely the linking of data files and the statistical nature of the data, that lie outside the realm of standard data processing and survey analysis. Second, multi-level data—particularly information about service availability—raise non-trivial conceptual problems. For example, there is the question of possible circularity. If modern contraception is not available anywhere, then there can be no users. If there are users, then obviously there must be some access. Moreover, local services are both a stimulus and a response to local demand. The family planning services in question were not established as part of an experimental design, and there is only the single cross-sectional measurement. Given these limitations and the likelihood of reciprocal effects, is it possible to describe the impact of services? What kind of causal model is appropriate? Can these data be used to project the consequences of expanding services or altering their quality?

The purpose of this paper is to demonstrate how the SA Module can give readily understandable estimates of the importance of service availability, in its various dimensions, to the current use of contraception. For this purpose, both an analytic strategy and specific statistical methods will be proposed. A range of specific problems and weaknesses in the data will be pointed out, and ways of confronting those problems will be illustrated. Finally, some recommendations for questionnaire design and data collection procedures will be made in an effort to reduce the seriousness of these problems in the future.

This approach will be illustrated by linking the individual and SA data collected by the DHS in Guatemala in 1987. Guatemala was selected in part because its SA Module has not yet been analyzed, except for some relatively simple comparative analysis (Tsui and Ochoa, 1989). Also, prevalence of modern methods is low in rural Guatemala, a condition that characterizes the African countries where major use of the extended SA module is now planned.

The final report for this survey was published in 1989 by the Instituto de Nutrición de Centro América y Panamá and the Institute for Resource Development (1989), and readers are referred to that report for a description of the survey and its results. Appendix A reproduces the SA questionnaire for Guatemala, while Appendix B gives a detailed description of the data processing steps and statistical procedures used in this analysis.

Statistical Issues

It is essential to understand the structure of the data at hand. As actually collected, multi-level data is far different from how it might be conceptualized in the abstract.

DHS surveys, like WFS and most other large surveys requiring personal interviews, are conducted in two stages. In the first stage, primary sampling units (PSUs) are drawn from a nationwide list of such units. (Of course, this process itself is typically broken down into stages or strata leading up to the final selection). The PSUs are referred to as segments or, the term preferred here, clusters. The quality of the sample hinges almost completely upon the representativeness of the clusters. All households in the clusters are visited, and the eligible women are identified. Then, during the second stage of the survey, a subsample of the women are interviewed. A cluster design has the advantage of far lower data collection costs per respondent than a simple random sample.

Most standard statistical procedures and software assume that individual respondents are part of a simple random sample and thus were selected independently of one another. A cluster design is less efficient than this ideal: the effective sample size is often about 30 percent less than it would be for a simple random sample, and the standard errors of estimates are correspondingly greater. (Standard errors are inversely proportional to the square root of the effective sample size.) The corrected standard errors for the principal estimates reported in the DHS surveys are calculated and published in the country reports (for Guatemala, see INCAP and IRD, 1989: Appendix III). Despite this fact, virtually all analysts use software that ignores this design effect.

In Guatemala, for example, the design effect for the proportion of women currently using a modern method of contraception is reported as 2.153, a particularly high value (INCAP and IRD, 1989: 81, Table III.1). The design effect can be interpreted as the ratio of the sampling variance in the actual survey to what it would have been in a simple random sample of the same size (Kish, 1965: 193). Alternatively, the reciprocal of the design effect can be interpreted as the ratio of the effective sample size to the actual sample size. Thus, the effective sample size for this variable in the Guatemalan survey was 1/2.153 or 46.4 percent of the actual total of 5,160 interviews. This paper will offer a behavioral interpretation of the similarity of women within a cluster on this variable, but regardless of its sources, the similarity invalidates the assumption of a simple random sample.

Consider now the information gathered by the SA Module. This information is not obtained independently for each woman; rather, it is collected at the cluster level and attributed to the woman as a description of her environment. Therefore, the number of observations equals the number of clusters, *not* the number of women. The cluster data may (improperly) be inflated by the number of women in the main survey, but this action does not enhance their accuracy. The importance of this fact for the statistical analysis cannot be over-emphasized.

For example, the SA Module in Guatemala asked, "Is this cluster covered by a program for community distribution of contraceptives?" Of the 116 usable rural clusters in the country, 33 (or a proportion p=.2845) were covered by such a program. The standard error of this proportion is the square root of p(1-p)/n, where n is 116; this is calculated to be .0419, with a 95 percent confidence interval of .2024 to .3666. That is, the 95 percent confidence interval for the proportion of all rural clusters in the population that are covered by a community distribution program ranges from 20.2 to 36.7 percent. To construct an interval for the percentage of women living in such areas, which would be of more interest, the clusters are weighted in proportion to size; the result is an interval ranging from 25.5 to 42.5 percent, centered at 33.4 percent. The interval is as broad as the one for the clusters, because there are only 116 data points.

Since this analysis is interested in individual women—they are the end point of the sampling scheme—it would seem natural to employ women as the units of analysis and attribute to them the characteristics of the cluster where they reside. Suppose that the variable under discussion, the presence of a community distribution program, was added to the data for individual women. The proportion of women who live in a rural cluster with a program would be p=.3342, and every standard computer package would calculate the standard error as the square root of p(1-p)/n, where n is the number of rural women: 3,043. That is, the standard error of the proportion would be .0086, and the 95 percent confidence interval would range only from 32 to 35 percent. Not only would the computer package ignore the design effect mentioned earlier; much more importantly, it would not recognize that there are only 116 data points, or pieces of information, distributed across the 3,043 women.

Estimates of means, correlations, and regression coefficients that are calculated by attaching cluster data to individual records are subject to this kind of misinterpretation. No one would consider conducting a national fertility survey with only one or two hundred women. It is all too easy, however, to succumb to the fallacy of attaching general information about the clusters to the specific data gathered on each of the women and then to proceed as if the data had the same statistical accuracy as several thousand independent observations.

To overcome this problem, the method used here partitions or allocates the available information into two levels: one for clusters and another for individuals. In contrast, other procedures all attach the macro (or cluster) data to the micro (or individual) data records and, from that point onwards, treat the macro predictor variables in the same way as the micro predictor variables. For example, in predicting contraceptive use, the usual statistical procedures do not differentiate between a macro level predictor, such as presence of a secondary school, and a micro variable, such as the woman's own level of education. (Empirical Bayes procedures take account of the error structure of the data—that is, of the fact that cluster-level variables do not vary within clusters—but they differ from our approach in important ways. See Appendix B for a description of alternative statistical approaches.)

The method used here takes advantage of the hierarchical character of the data, namely the "nesting" of individuals within areas. It will allow conclusions about how important a given predictor variable is in explaining variation *between* clusters or in explaining variation *within* clusters; moreover, it recognizes that (except as indicated below) a predictor at the cluster level can *only* explain variation between clusters. The presence or absence of a pharmacy, for example, is a constant for everyone within a given cluster and cannot explain why one woman uses contraceptives while another woman in the same cluster does not. The method could be extended to utilize multi-level data in which clusters are nested within districts, districts are nested within regions, and predictors are available at each of these levels. The statistical literature would describe the procedure as hierarchical analysis of covariance with random effects (because the clusters are sampled at random).

The partitioning procedure is analogous in logic to the manner by which a "total" sum of squares is broken into a "between groups" sum of squares and a "within groups" sum of squares in the analysis of variance. In that familiar decomposition, the two sums of squares are statistically independent of one another and contain no overlapping information, because, within each group, the dependent variable is stated as a deviation from the group mean. In this analysis, there is a binary dependent variable (the current use of contraception) and logit regression is used instead of analysis of variance, but the reasoning is almost completely parallel.

Analytical Model

Most earlier analyses fail to provide a theoretical model to guide the statistical analysis, perhaps in part as a reflection of the inappropriateness of the statistical procedures. In this model, blocks of one or more variables are conceptualized and labelled as follows:

- Y is the mean of a dependent variable within a cluster;
- Y' is the within-cluster deviation of a dependent variable from the cluster mean;
- X is the mean of a predictor variable within a cluster;
- X' is the within-cluster deviation of a predictor variable from the cluster mean;
- W is a cluster-level socioeconomic variable; and
- Z is a cluster-level intervention variable.

In this study, for example, Y will be the proportion of women in a rural Guatemalan cluster who currently use modern contraception. Y' will measure the extent to which an individual woman departs from that proportion. Y is built up from individual responses and, in fact, is just the mean of a binary (0/1) response. In other analyses, Y could refer to the intention to use contraception, to current breastfeeding practices, or to other outcomes that might be affected by local services.

This analysis assumes that individuals within clusters have a common environment, including one another, and that there are many reasons why they should behave similarly. Therefore, information about these similarities is used to predict the mean outcome, and information about their differences is used to predict differences in outcomes. Patterns initially are identified using means at the highest level of aggregation; then the analysis works downward in steps to identify differences at progressively lower levels of aggregation; and finally individuals become the analytic units.

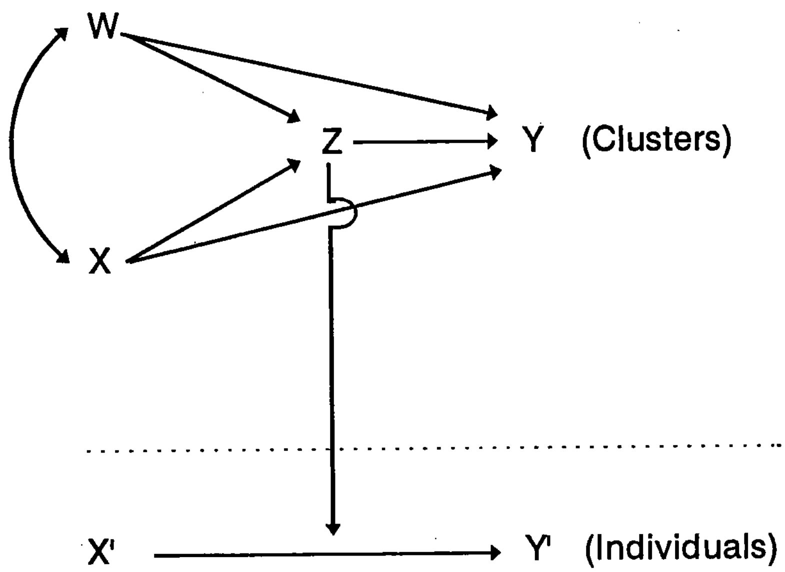
The variable, X, will describe the means or distributions of literacy, the source of household water, the number of children, and several other predictors drawn from the survey of individual women. The analysis first asks, for example, "How does the level of literacy in a cluster affect the level of contraceptive use?" The next question is, "How does an individual woman's deviation from the prevailing level of literacy affect her deviation from the prevailing level of contraceptive use?" In combination, the statistically independent answers to these two questions tell how an individual woman's literacy will affect her personal use of contraception, but the hierarchical partitioning allows us to determine just how internally homogeneous—or heterogeneous—the clusters are.

Variables labelled W are drawn from the SA Module and are intrinsically macro-level indicators that indicate the socioeconomic climate, modernity, or relative development of a community. Included are the presence of a post office, regional centrality, or major economic activity. However, they do not encompass either program or intervention variables; like the X variables, they must be controlled statistically to evaluate the impact of the intervention variables. The SA Module contains few variables of this type, so it is a particular advantage of the procedure used here that they can be supplemented by summaries of individual-level predictors.

Finally, Z refers to the specific variables in the SA Module that are postulated to affect the level of contraceptive use, such as distance or time to services and the quality of services. Specifically, it is hypothesized that the presence of services, especially those of higher quality, will increase contraceptive use even after other determinants have been controlled.

Figure 1 illustrates the hierarchical approach followed by this analysis. The dashed line indicates the division between the cluster-level and individual-level analyses—or, more precisely, the "between clusters" and "within clusters" analyses. Solid arrows indicate postulated effects between variables of different types. There are no arrows from Y to Y' or from X to X' because the cluster means (Y and X) are statistically independent of the within-cluster deviations (Y' and X', respectively). Effects from X and W to Z are postulated because it seems likely that services will be better developed and utilized in or near clusters that have higher literacy rates, are more central, and are more developed, than in clusters that are poorer, more remote, and less developed. This is an empirical question; one can imagine government policies that would intentionally place hospitals and health centers in the least modern locations.

Figure 1 Conceptual model of the analysis



The model shows an arrow coming from Z that intersects the arrow from X' to Y'. This represents possible interactions, such that the effect of X' upon Y' may vary according to the level of Z. It is not possible for Z to have a direct effect upon the within-cluster deviation Y' (because Z is a constant within each cluster), but it would be possible and of considerable interest for such interactions to be present. Other interactions could be postulated and represented in the same way (for example, the effect of X' upon Y' might also vary according to the level of W or X), but such effects would be less relevant. The curved arrow connecting the W and X blocks of variables simply indicates that a causal direction has not been assigned to their relationship.

There are possible reciprocal relationships. For example, it could be postulated that the level of services is partly a response to demand and that current use is a lagged indicator of earlier use, which stimulated the growth of services. This effect could be partly captured by an arrow directed from Y back to Z, although it could also be expressed by using information in the survey about desired family size and past contraceptive use. Such complexities will be overlooked. The model could be expressed more elaborately in terms of latent variables and multiple indicators (and perhaps estimated with LISREL), but such refinements will be deferred to future applications if they appear justified.

There are several distinct ways in which the effect of service availability on contraceptive use can be described within this framework. A complete analysis would have the following four phases, although this paper will concentrate on the first two:

(1) Fit or predict Y using Z alone, obtaining unadjusted effects of service availability on current use.

(2) Fit Y using W, X, and Z, estimating effects for Z that are adjusted for, or net of, the covariates W and X.

(3) Fit Z with W and X, and then estimate the indirect pathways from W to Y and X to Y via Z. The effects of W and X may be expressed through, or implemented by, service availability. For example, if better quality facilities tend to be located in communities where the standard of living is higher, then the facilities will be a mechanism by which a higher standard of living leads to the improved use of contraception.

(4) Finally, at the micro level, the fitting of Y' with X' may be improved by including interactions between X' and Z.

Baseline Model: Partitioning into Two Levels

Total unexplained variation in current contraceptive use can be divided into two independent components: that between clusters and that within clusters. With an interval-level dependent variable, the total unexplained variation would be the familiar sum of squared deviations about the mean. In the theory of generalized linear models (McCullagh and Nelder, 1983), the total sum of squares is just one element in a larger class of measures of fit, namely the element appropriate for interval-level variables with normally distributed errors. The general term for measures of fit is called the "deviance," always calculated as -2 times the natural logarithm of the likelihood function for a specific model. Deviances for logit models, like sums of squares, have chi-square sampling distributions with specified degrees of freedom.

Table 1 shows the deviances for all of Guatemala and for just the rural portions of the country. In addition, it shows the allocation or partitioning of the deviance into two components: between and within clusters. The SA Module was administered mainly to rural clusters in Guatemala, so this partitioning forms the starting point for the analysis.

Table 1 Partitioning of the deviance in current use of modern contraceptive methods in Guatemala (1=Yes, 0=No) into portions between clusters and within clusters

	Deviance Between Clusters	Deviance Within Clusters	Total Deviance
All of Guatemala	818.0, 241 df (20.1%)	3249.3, 4918 df (79.9%)	4067.3, 5160 df (100.0%)
Rural Guatemala	426.7, 121 df (23.1%)	1417.7, 3119 df (76.9%)	1844.4, 3240 df (100.0%)

As can be seen from Table 1, there are a maximum of 121 usable degrees of freedom for the clusters (there were 122 rural clusters in Guatemala; six will be omitted in the main analysis because the SA module was not administered to them). This is a rather small number of observations relative to the many variables in the SA Module. It is a mathematical inevitability that *any* 121 pieces of information about the 122 clusters would completely explain the variation between clusters, and it is likely that the bulk of the variation could be explained by a much smaller number, particularly if some of these variables are categorical. Just a few variables, whether they deal with service availability or anything else, can provide a fingerprint for a cluster. The saturation of the degrees of freedom would not be readily apparent under the conventional approach, which would attach the cluster data to the file of individuals—of whom there are 3,241 in the rural areas.

The second major observation to be drawn from Table 1 is that 23.1 percent of the total variation is between clusters rather than within clusters. This is the upper limit of what can be explained by characteristics that vary from one cluster to another. On the one hand, this can be described as just a fraction of the total; clearly, the great majority (76.9 percent) of the variation occurs within clusters and cannot be explained by service availability or any other cluster-level characteristics (except, as noted above, through the secondary effect of interactions with within-cluster deviations on individual-level characteristics). On the other hand, the variation between clusters is highly significant. Calculating the between-cluster deviance per degree of freedom (426.7/121) and dividing it by the within-cluster deviance per degree of freedom (1844.4/3240), an F ratio of 6.2 is obtained. If there were no tendency for women in the same cluster to be similar, the clusters would only account for 121/3240 or 3.7 percent of the deviance. Per degree of freedom, clusters are over six times as important as individuals as a source of variation.

A final important piece of introductory information is the level of current use of modern methods. Y is a binary variable with a binomial distribution; $\text{logit}(Y) = \log[\Pr(Y=1)/\Pr(Y=0)]$ is to be analyzed with linear additive models. When there are no covariates, the best-fitting model for data from all of Guatemala is $\text{logit}(Y) = -1.865$. The proportion currently using a modern method is then $\exp(-1.865)/[1+\exp(-1.865)] = .1341$ or 13.41 percent. This number can be verified in the final report (INCAP and IRD, 1989: 48, Table 5.5) by adding the percentages using the pill, IUDs, injectables, diaphragms, foam, condoms, and sterilization. Pills and female sterilization are the leading methods by far in Guatemala, accounting for 20 percent and 56 percent, respectively, of all current use.

In rural Guatemala, $\text{logit}(Y) = -2.410$, and the corresponding proportion of women currently using modern contraception is .0824 or 8.24 percent. This is a low level of use, making it more difficult to obtain significant effects than if the level of use were higher. Put simply, the focus of interest is on how 267 users, in a sample of 3,241 rural women, are distributed across 122 clusters. On average, each cluster consists of about 27 women, 2 of whom are current users.

The analysis will include rural women regardless of marital status, age, parity, pregnancy status, and the like. A case could be made for removing some non-users because they are not at risk of conceiving. However, there are some users in nearly every category of marital status, age, or parity. Some women are not currently in a union but were sterilized at an earlier time, and they reflect the determinants of use. Rather than make arbitrary exclusions, all the women will be retained and covariates will be employed to adjust for risk factors.

At this point the rural sample will be reduced by dropping the six clusters that were omitted from the SA survey. There is no evidence that the omitted clusters were unrepresentative. The reduced sample consists of 116 clusters, including 3,043 women, of whom 252 are current users (for an overall prevalence rate of 8.28 percent). The baseline logit regression coefficient is -2.405, and the baseline deviance between the clusters is 410.18.

Across the 116 clusters, the percentage of current users ranges from 0.0 to 37.5 percent. As Table 2 shows, this is a very skewed distribution, with no users at all in 51 clusters representing 36.9 percent of the women and a prevalence rate of 20 percent or more in only 15 clusters which contain 15.6 percent of the women.

The 116 clusters vary in size, with the number of women interviewed for the individual survey ranging from 7 to 55. At the first quartile, about 19 women per cluster were interviewed; the median number is 26; while at the third quartile, about 33 women per cluster were interviewed. The actual number of users was 0 in 51 clusters, but it ranged as high as 15 in a single cluster.

Table 2 Distribution of clusters according to prevalence rate and percentage of the sample in those clusters

Range in Prevalence (percent)[1]	Number of Clusters	Percentage of Women
0	51	36.9
0-4	16	16.0
5-9	18	16.0
10-14	10	9.2
15-19	6	6.3
20-24	6	6.0
25-29	4	3.8
30-34	3	4.1
35+	2	1.7

[1] The range "5-9 percent" includes exactly 5.0 percent and ranges up to, but does not include, exactly 10.0 percent. All other intervals are constructed similarly, except that "0-4 percent" does not include exactly 0.0 percent.

In this model, n is used to refer to the cluster size and x to refer to the number of users within the cluster; x can take any integer value from 0 to n. The statistical model postulates that x is binomially distributed around its expected value, with denominator or base n. The model attempts to fit x using n and other information that is provided in a linear logit format. If, for example, there are no covariates, the model would calculate the overall proportion of users across all clusters (the sum of x divided by the sum of n). Within a specific cluster, the predicted value of x would then be obtained by multiplying n by this proportion. The overall measure of fit is a summary of the difference between the observed x and its fitted or predicted value. An individual-level analysis is formally equivalent to this procedure except that n=1 for each individual and x can only take the values 0 or 1. More details are given in Appendix B.

Socioeconomic Indicators from the Survey of Women (X Variables)

As described earlier, three sets of variables may account for the between-cluster deviance of 410.18. The first set of indicators, called the X variables, are drawn from the survey of individual women and comprise the means and distributions of data collected at the individual level. These compositional variables will be interpreted as indicators of the demographic, social, and economic environment of the cluster. It is postulated that they affect the predisposition to use contraception, the availability of services, and the impact of those services.

Like most DHS surveys, the Guatemalan survey contains a rather limited number of such variables. For the purposes of this illustrative analysis, this number will be limited even further, to just five, based on a preliminary screening. These variables are listed in Table 3.

Table 3 Reductions in deviance by type X variables after screening

Variable	Reduction in Deviance[1]
Measure of Risk and Demographic Composition	
Marital status (6 categories)	96.30
Measures of Development	
Source of household water (4 categories)	98.70
Literacy (3 categories)	88.42
Has electricity? (No/Yes)	108.24
Has TV? (No/Yes)	173.13
All five variables combined	266.46

Note: The within-cluster proportion in each category of these variables is coded onto the record for each cluster and used as a predictor of the proportion of women in the cluster who are currently using modern contraception.
[1] The reductions in deviance are against a between-cluster baseline deviance of 410.18. All reductions are highly significant.

Table 3 shows the reduction in between-cluster deviance that can be attributed to each of the five variables. Marital status provides a control for the women's exposure to the risk of childbearing and thereby the need for contraception; it reduces the deviance by 96.30 points. The number of living children is important for explaining variation on an individual level—mainly, in Guatemala, in distinguishing between women with no children and women with at least one child. However, the parity distribution of a cluster is not statistically significant for predicting current contraceptive use when the other variables are present. Neither parity nor age shows large variation from one cluster to another once marital status is controlled.

Two of the four measures of the cluster's socioeconomic status stand out: literacy and the proportion of women (essentially the proportion of households) having a television set. The source of household water generally characterizes an entire cluster, not just a specific household, and reflects the standard of living of the community. The proportion with electricity indicates the extent of rural electrification; considerably more women have electricity than have television. It was thought that the region in which the cluster is located might reflect other, unmeasured status variables that tend to vary systematically from one region of the country to another, but region was not significant when the other measures were included.

An additive model including all five measures reduces the between-cluster deviance by 266.46 points at a cost of 12 degrees of freedom. The fit could be improved further by including other covariates from the survey of individual women or by including various interactions among the selected variables, but not by a large or statistically significant amount (although this conclusion is based on an admittedly incomplete screening). The residual deviance and degrees of freedom are 146.72 and 103, respectively. This block of variables reduces the baseline deviance by 65 percent, using 10 percent of the available degrees of freedom.

It is not necessary to present the coefficients or effects for these variables, because they are employed mainly as controls for assessing service availability. However, all the coefficients have the signs that would be expected, with current contraceptive use being more common in clusters with higher proportions of women who are currently married, have piped water, can read easily, have electricity, and have television. For example, if E is the proportion of women who have electricity, then the fitted equation is $\text{logit}(Y) = -3.156 + 1.857E$. Substituting E=0.0, E=0.5, and E=1.0 into this equation, and making the reverse transformation, the fitted percentage of current users is 4.1 percent, 9.7 percent, and 21.4 percent, respectively. Similarly, the fitted percentage of contraceptive users in a cluster where no women could read would be 1.4 percent; if all women could read, but with difficulty, it would be 9.2 percent; and if all women could read easily, it would be 47.0 percent. These are statistical simulations, of course, and do not imply that electrification and literacy would automatically and immediately induce higher levels of use.

The interpretation of these variables at the cluster level may call for clarification beyond that given earlier. They have been described as indicators of the socioeconomic environment, but in fact they are aggregations of individual-level variables. These aggregations simply provide a structure for partitioning the influence of an inherently individual-level measure into two, independent components: one which is shared or homogeneous within the cluster and another which represents the heterogeneity of individuals within the cluster. Table 4 describes the balance of these components. The first row of this table refers to Y, the dependent variable, and shows how its total deviance, 1738.08, is allocated between and within clusters (the between-cluster deviance, 410.18, is 23.6 percent of the total). This partitioning is similar to that in Table 1 for rural Guatemala, except that Table 1 referred to the full 122 clusters, rather than the reduced set of 116. The remaining rows of Table 4 list the five predictor variables of type X and their own deviances. As would be expected, clusters display the most homogeneity for the presence of electricity in the household, the source of drinking water, and the presence of a television set (which, of course, depends on the presence of electricity). Literacy shows about the same level of heterogeneity as contraceptive use itself; marital status is the least homogeneous.

Table 4 Degree to which current use of modern contraception and the five individual-level predictors vary between and within clusters in rural Guatemala

		Percentage of Deviance	
	Total Deviance	Between Clusters	Within Clusters
Current Use	1738.08	23.6	76.4
Literacy	6162.17	21.2	78.8
Source of Water	7699.60	50.2	49.8
Electricity	3736.34	56.1	43.9
Television	2728.66	39.3	60.8
Marital Status	7851.92	13.7	86.3

Note: If these variables did not tend to be similar within clusters, then 3.8 percent of the total deviance would be expected to be between clusters, and 96.2 percent within clusters.

Because of the high degree of internal homogeneity of the clusters—equivalently, because of the substantial variability between clusters—it is reasonable to characterize the means and proportions on these variables as indicators of differences between clusters. Indeed, they are regarded in that light. However, it should be made clear that the statistical model does not reify them in that way but simply partitions the explained variation in the manner described above.

Socioeconomic Indicators from the SA Module (W Variables)

The next block of variables, referred to as W variables, comes from the SA Module itself. These variables are intrinsically defined at the aggregate level. Such measures are sometimes called "integral" to contrast them with "compositional" variables that are built up as summaries of measures taken at the individual level.

A relatively small proportion of the SA Module consists of a general description of the community. In searching for a set of indicators that provide a good profile of the cluster, in relation to current use of contraception, the information that was collected was not fully explored. Nevertheless, a concise set of indicators was readily identified. Four variables were significant. Table 5 lists the distribution of the women and the contraceptive prevalence rates for each category.

Table 5 Distribution and contraceptive prevalence rates for type of locality, presence of secondary school, presence of post office, and presence of theater or cinema

	Number of Clusters	Percentage of Women	Contraceptive Prevalence Rate (%)
Type of Locality		(Deviance reduction 92.74 with 5 d.f.)	
Finca	15	12.8	7.2
Aldea, nuclear	18	19.7	15.9
Aldea, dispersed	36	28.5	5.8
Caserio, nuclear	6	4.5	9.4
Caserio, dispersed	36	28.6	13.2
Pueblo	5	5.9	18.5
Secondary School		(Deviance reduction 36.71 with 1 d.f.)	
Present	22	22.4	14.3
Absent	94	77.6	6.6
Post Office		(Deviance reduction 1.37 with 1 d.f.)	
Present	30	27.4	9.3
Absent	86	72.6	7.9
Theater or Cinema		(Deviance reduction 62.59 with 1 d.f.)	
Present	21	21.4	16.5
Absent	95	78.6	6.1

The first variable, type of locality, is a proxy for the substantive differences between a finca, dispersed aldea, and nuclear caserio, on the one hand, and a nuclear aldea, dispersed caserio, and pueblo, on the other hand. Contraceptive prevalence in the second set of localities is roughly double that in the first set. The importance of this distinction rests on the unmeasured characteristics of those places.

There are interactions among the remaining three variables, presence of a secondary school, post office, or a theater or cinema. After experimenting with possible summaries or combinations of the three, an additive model proved to be most parsimonious. While the presence of a post office is not statistically significant by itself, in the presence of the other information it becomes quite important. The additive model including all four W variables reduces the baseline deviance by 116.09 points, at a cost of 8 degrees of freedom, leaving a residual deviance of 294.09 with 107 degrees of freedom. That is, the W variables explain 28.3 percent of the deviance using 7 percent of the available degrees of freedom.

Perhaps surprisingly, several other potentially interesting indicators were not significant. These included the size of the local community, the distance to the nearest community of 20,000 or more inhabitants, and the presence of a weekly market.

Program and Service Availability Indicators (Z Variables)

Next come the predictor variables of primary interest: the presence of a family planning program in the local area, some aspects of that program, the availability of services, and the quality of those services. Again, the selection of key indicators will be based on whether the responses significantly improve the ability to predict the number of current users in the cluster, given only the number of women in the cluster, with an additive logit regression model. To repeat, the baseline deviance is 410.18, with 115 degrees of freedom.

The SA Module in Guatemala asked, "Is this segment covered by a program for community distribution of contraceptives?" A "yes" response was given in 28.5 percent (33) of the 116 clusters, encompassing 33.4 percent of the women. The contraceptive prevalence rate in those clusters with a program was 14.6 percent; in the remainder, the rate was 3.5 percent. This variable alone reduced the deviance by 74.02 points, at a cost of only one degree of freedom. Additional questions about the program, such as the specific methods being promoted and the frequency of promotional visits to the local area, did not add significantly to the basic information that a program existed.

This variable will be kept separate from the remaining information about the service environment. That is, the block of Z variables consists of two distinct components: the powerful binary indicator of whether there is a program or not and the detailed description of just what facilities are locally present, their accessibility, and their quality.

The SA Module included a page of detailed questions about health services and another about family planning services; this analysis is restricted to the latter. The questions covered five specific sources of family planning services: hospitals, clinics, health centers, pharmacies, and private physicians. As a rule, hospitals and health centers are government facilities, and the others are private. The following data was collected:

- **Presence or absence, in the locality, of a source of family planning services** - Various interviewers seem to have interpreted the term "locality" ("la localidad") in different ways, so that the stated distance to the nearest source was used instead. Presence was recoded into three categories: no facility named; named, but 30 km or more distant; named and less than 30 km distant. The first of these is regarded as a reference category with which the other two categories are to be compared. It was initially assumed that if no facility was named, then none was located anywhere in the vicinity. However, there is evidence that this category is a mix of "none present" and "no response". The results were substantially easier to interpret when it was kept separate rather than pooled with the category of greatest distance.

- **Distance to this source, in kilometers** - Distance was recoded into the following six categories: not named, 30 or more km, 20-29 km, 10-19 km, 5-9 km, and 0-4 km. Once again, "not named" is the reference category. Subsequent questions were only answered if the source was within 30 km.

- **Type of transportation** - Since there was little variation in the data, this variable was not used. Its main impact is to alter travel time, which is itself estimated on the questionnaire for the most common method of transportation. Therefore, travel time is used instead.

- **Travel time** - This was recoded into the following six categories: not named, 120 or more minutes, 60-119 minutes, 30-59 minutes, 15-29 minutes, and 0-14 minutes.

- **Contraceptive methods available** - The questionnaire recorded whether each of these methods was provided at the facility: the pill, IUD, injection, condom, female sterilization, male sterilization, and other. For this analysis, an extra code for the number of methods offered was also constructed.

- **Cost of each method** - This information was not used.

- **Year in which each method was first available at this source** - This information was not used.

- **Number of doctors involved in family planning** - This and the remaining questions were only asked for hospitals, clinics, and health centers.

- **Number of nurses involved in family planning**

- **Days of the week when open for family planning** - This information was compressed into the number of days per week when open at all.

- **Number of hours open each day** - This detailed information was not used and should be recoded and perhaps edited if used by other researchers. Hospitals are often listed as open 24 hours a day, 7 days a week, but it is unlikely that family planning advice or supplies are available during all of this time. The current version of the SA Module has revised these questions.

Detailed though this information may seem to be, it does not give a complete picture of the family planning environment. The most serious flaw is that it does not describe the existence or density of multiple sources of family planning services of the same type. This makes it difficult to examine possible substitution effects. Suppose, for example, that a respondent wishes to purchase condoms from a pharmacy. The nearest pharmacy may not sell them, but another one a short distance away might sell them. Information about the second pharmacy is not included in the questionnaire. Similarly, if a woman wishes to be sterilized, it is possible that none of the nearest facilities in each category offers sterilization, but there could be a second facility within the 30 km limit which does provide sterilization. Such a place will not be recorded. (It is also possible that gradations of nearness are not very relevant for a one-time behavior such as sterilization.)

Despite these shortcomings, it is possible to account for a great deal of the variation in prevalence using the information in the module.

1. Diversity of facilities. How many of the five types of service points are available within 30 km? This diversity index reduces the deviance by 56.96 points at a cost of five degrees of freedom. It would be expected that prevalence is low in communities with no access and higher in communities with greater access, while beyond some threshold, additional facilities might not make any difference. In fact, as shown in Table 6, no such pattern is observed. The main function of this variable is to identify five clusters (with identification numbers 100, 183, 209, 215, and 218) that are missing large amounts of data, three of which happen to have high proportions of users. It is suspected that these clusters are adjacent to an urban area with so many choices of facilities that the interviewers opted to omit the information. This variable will be retained and referred to as a measure of service diversity, because it is thought such a measure should be included. However, this indicator does not fully measure diversity; in part, it also serves to control for non-response.

Table 6 shows that almost all women (all but 3.8 percent) live within 30 km of at least one facility. The modal number is three different kinds of facilities out of the five possible categories.

It may be mentioned at this point that the SA data for Guatemala show several inconsistencies that suggest incomplete fieldwork and editing. It is troubling that so much data is missing and that, as this table shows, it is not missing at random.

Table 6 Contraceptive prevalence rates of clusters according to the number of different kinds of facilities (hospital, clinic, health center, pharmacy, and private physician) stated to be within 30 km of the cluster

Number of Kinds of Facilities	Number of Clusters	Percentage of Women	Contraceptive Prevalence Rate (%)
0	5	3.8	23.1
1	15	14.1	5.4
2	27	20.5	8.7
3	40	35.0	5.7
4	16	14.8	14.0
5	13	11.8	6.7

Note: Reduction in baseline deviance is 56.96 for 5 degrees of freedom.

2. ***Presence of specific types of facilities.*** Which of the five types of service points are available within 30 km? This variable distinguishes whether or not any source was named, because, as already noted, the failure to name a service point is likely to indicate incomplete data rather than the genuine absence of the type of facility in question. Interpreting this information is problematic. Table 7 gives the prevalence rates for clusters in which a facility of each type was not named, was stated to be more than 30 km from the cluster, or was stated to be within 30 km of the cluster.

One hospital or another was named for 50 clusters, although it was located within 30 km of only 23 of them. Examination of the data showed that these 50 clusters shared 20 different hospitals. This heavy overlap suggests (but does not require) that the clusters are not independent of one another.

For each type of facility except health centers, contraceptive prevalence is higher if the nearest facility is stated to be within 30 km than if it is stated to be farther away. However, there are very few clusters in which the nearest health center, pharmacy, or private physician is stated as being more than 30 km away. The reduction in deviance is not significant for hospitals and clinics. They will be retained, however, because certain characteristics of the hospitals and clinics, when these facilities are named, will prove to be significant.

When current use is fitted with the single diversity indicator and the five presence indicators, the baseline deviance is reduced by a total of 78.19 points, at a cost of 14 degrees of freedom (rather than 5+5x2=15 degrees of freedom, because of an empty cell), giving a residual deviance of 331.99 with 101 degrees of freedom.

Table 7 Contraceptive prevalence rate of clusters according to the presence of a hospital, clinic, health center, pharmacy, or private physician

Type of Facility	Number of Clusters	Percentage of Women	Contraceptive Prevalence Rate (%)
Hospital		(Deviance reduction 4.17 with 2 d.f.)	
Present, <30 km	23	22.0	11.1
Present, >30 km	27	20.2	6.1
Not Named	66	57.8	8.4
Clinic		(Deviance reduction 3.89 with 2 d.f.)	
Present, <30 km	33	28.3	9.2
Present, >30 km	33	24.3	6.6
Not Named	50	47.4	8.6
Health Center		(Deviance reduction 17.33 with 2 d.f.)	
Present, <30 km	108	93.7	7.7
Present, >30 km	3	1.3	10.0
Not Named	5	5.0	18.5
Pharmacy		(Deviance reduction 12.92 with 2 d.f.)	
Present, <30 km	97	82.9	8.1
Present, >30 km	8	5.9	3.4
Not Named	11	11.2	12.0
Private Physician		(Deviance reduction 10.78 with 2 d.f.)	
Present, <30 km	57	51.2	7.6
Present, >30 km	7	6.1	3.8
Not Named	52	42.7	9.8

Note: Each indicator is significant; reduction in baseline deviance for all five variables is 40.23 for 10 degrees of freedom.

3. ***Access to facilities***. Next are added a minimal set of indicators of accessibility, based on distance and travel time, for each of the five service points. In this case, distances below 30 km are subdivided into four categories: <5, 5-10,10-19, and 20-29 km. Because distance and travel time are refinements of the presence indicators just described, they are used in place of those indicators rather than in addition to them. After experimenting with the use of time alone, distance alone, some other composite of the two (such as their sum), or a different categorization of the indicators, the most parsimonious way to include them proved to be a simple additive model. While a couple of individual indicators are not statistically significant here, they become significant in later models, so all of them are retained.

The additive model using the diversity index, the five distance indicators, and the five travel-time indicators has a residual deviance of 174.82 with 67 degrees of freedom. Clearly, adding details of time and distance substantially improves the prediction of the local prevalence rate.

Both the distance and the time indicators are important; each adds considerably to the other. Distance and time are most important for the nearest health center, pharmacy, and private physician; they are much less important for the hospital and clinic. Perhaps this is because hospitals and clinics are used more for methods that require only one visit (e.g., sterilization) or an annual visit (e.g., the IUD), while the other facilities are used more for methods that require re-supply. Distance and time might be less important when visits are infrequent.

The effects of time and distance are not linear and not even monotonic. With the exception of only one type of facility (details will not be given here), prevalence is higher if the facility is 0-4 km from the cluster than if it is 5-9 km, and higher if it is 5-9 km than if it is 10-14 km distant. Beyond that distance, however, there is considerable irregularity, with prevalence sometimes increasing at the greater distances. The coefficients for the different travel time intervals are even more erratic.

It is suspected that the irregularity of the coefficients and several other apparent inconsistencies are due to substitution effects or trade-offs between facilities. For example, a health center may be a source of condoms, but a pharmacy might be preferred if available, with clients willing to go the extra distance to get to the pharmacy. Such a scenario is only hypothetical; it would be impossible with the limited degrees of freedom and small number of users in this study to demonstrate substitution effects conclusively, because they would require large numbers of interaction terms.

4. Quality of facilities. What characteristics of the local facilities, other than access, are associated with local prevalence? Here the focus is on which of six specific methods are offered, the total number of methods offered, the number of doctors, the number of nurses, and the number of days open (ignoring number of hours per day). Each of the following indicators was statistically significant in the final equation, and no others were significant:

- the number of doctors in the nearest hospital,
- the number of nurses in the nearest clinic,
- the number of nurses in the nearest health center,
- whether or not the nearest hospital offers female sterilization, and
- whether or not the nearest private physician offers pills.

When pharmacies are examined separately, it is important to know whether or not they sell condoms. However, this piece of information is not significant in the presence of all the other indicators of the family planning environment.

Together, the five variables listed above reduce the deviance by a further 78.49 points, at a cost of 5 degrees of freedom. This is a substantial improvement in fit, especially considering that access has already been taken into account. These five characteristics should be regarded simply as indicators. In particular, the number of staff in a facility is probably a proxy for the overall size, diversity of services, visibility, and client base.

As additional evidence of substitution effects, only two of these variables actually show positive associations with local prevalence in the combined equation: the number of doctors in the nearest hospital and whether the nearest private physician offers pills. The other three coefficients are very significant—and negative. These patterns imply non-additivity and the need for interaction terms. Rather than adding interaction terms, however, the model will be stated in terms of the variables it includes and the specific parameter estimates will be de-emphasized.

The model that attempts to fit current prevalence in the clusters using the measures of diversity, presence of specific facilities, access, and quality has a residual deviance of 96.33 with 62 degrees of freedom. These variables have reduced the baseline deviance by 77 percent while using 46 percent of the available degrees of freedom.

The relationships described here are "zero-order" because they do not control for any of the other characteristics of the clusters and do not elaborate the possible pathways of effects. The next section will extend the analysis to include the other major blocks of variables.

Statisticians generally advise against models that use such a high proportion of the total degrees of freedom. When all blocks of variables are brought together the model will use over half of the degrees of freedom. The biggest consumer of degrees of freedom is the set of measures of distance and time, and they are so central to this study that they must be retained.

Adjusting the Importance of Service Availability for Other Influences

It is clear that service availability is strongly related to prevalence, although it is difficult to interpret coefficients for reasons that can probably be traced to (1) non-random missing data (that is, the failure to report facilities that are actually present) and (2) substitution effects between nearby facilities (both between the closest facilities of different types and also between one facility and the next nearest facility of the same type).

It could be hypothesized that service availability has no incremental impact on prevalence once the total environment has been taken into account. In this scenario, a more modern or developed community will be more likely to have family planning services—diverse, numerous, nearby, and of high quality—in the same way that it is more likely to have, for example, a secondary school, a post office, and a theater or cinema. Health and family planning services will be an aspect of modernity itself and not an independent lever that can be manipulated separately from the larger socioeconomic environment.

If it were found that, indeed, the SA variables do not help predict local prevalence once the socioeconomic environment has been taken into account, it would *not* follow that services are unimportant. Rather, the inference would be that the impact of services depends upon, and is part of, the whole process of development. If this were the case, it would be most effective to place new family planning services and improve existing services in those communities that are already the most developed in other ways.

Alternatively, it might be found that family planning services have an effect on contraceptive prevalence even after adjustment for other influences. This would suggest that providing services has an impact independent of the stage of local development. Facilities could be placed and improved in communities where use and development were low, and an impact could still be anticipated.

As is usually the case when two such diametrically opposed alternatives are presented, in most countries the truth probably lies somewhere in between.

Referring back to the causal model, assume now that the X and W blocks of variables have been taken into account and the task is to determine whether the addition of the Z variables helps to predict Y. Essentially, this amounts to first fitting Y with the full set of socioeconomic indicators in the X and W blocks and regarding the residual deviance from that model as a new baseline deviance. Then the various components of service availability are added to this new baseline model, and, as each component is added, it is examined for significant reductions in the deviance—i.e., improvements in fit.

Table 8 presents the measures of fit for the models described above. A new baseline model, referred to as Model X + W, is estimated using the two blocks of socioeconomic variables. The bulk of the information is clearly in the X block, which includes the within-cluster distributions of marital status, literacy, source of household water, electricity, and television. The W block, which comes from the SA Module and consists of type of locality, presence of a secondary school, presence of a post office, and presence of a theater or cinema, gives a marginal reduction in deviance of only 14.49 points at a cost of 8 degrees of freedom. This is not a statistically significant reduction. (The 95th percentile of the chi-square distribution with 8 degrees of freedom is 15.5.) Nevertheless, the W variables will be incorporated into the baseline model because of their substantive interest.

The various components of block Z are added one at a time, separately rather than cumulatively. Only one of these components is significant at the .05 level: the measures of distance and time to the nearest facilities. This block reduces the new baseline deviance by 94.02 points at a cost of 44 degrees of freedom. This is less than half the effect these variables had in the previous section of this paper, but is still far above the threshold of significance (the 95th percentile of chi-square with 44 degrees of freedom is 60.5). Table 8 also shows that the distance indicators and the time indicators are highly significant separately and that each set adds significantly to the other.

Table 8 Quality of fit to the prevalence of current use of modern methods of contraception in rural Guatemala, using the X and W blocks of socioeconomic indicators and components of the Z block for program and service indicators

Model	Predictors	Deviance (d.f.)	Improvement	
			Over:	Deviance (d.f.)
Original baseline	none	410.18 (115)		
Model X	socioeconomic, from individual questionnaire	146.72 (103)	original baseline	263.46* (12)
Model W	socioeconomic, from SA module	294.09 (107)	original	116.09* (8)
Model X+W	socioeconomic, from both questionnaires	132.23 (95)	original baseline	277.95* (20)
			Model X	14.49 (8)
			Model W	161.86* (12)
Model X+W+ diversity	socioeconomic plus diversity of facilities within 30 km	128.40 (90)	Model X+W	3.83 (5)
Model X+W+ presence	socioeconomic plus presence of specific facilities within 30 km	118.78 (85)	Model X+W	13.44 (10)
Model X+W+ access	socioeconomic plus distance and travel time to nearest facilities	38.21 (85)	Model X+W	94.02* (44)
			Model X+W+ presence	80.57* (34)
			Model X+W+ distance	42.60* (19)
			Model X+W+ access	46.40* (20)
Model X+W+ quality	socioeconomic plus quality of nearest facilities	122.24 (90)	Model X+W	9.99 (5)
Model X+W+ program	socioeconomic plus presence of local distribution program	132.20 (94)	Model X+W	0.03 (1)

* Improvement is significant at .05 level

In contrast, the indicators for the existence of a program, and service diversity, presence, and quality are all insignificant when the socioeconomic environment has been taken into account. The difference due to the five quality measures is 9.99, which is below the critical value of 11.1 for chi-square with 5 degrees of freedom. None of these five indicators is significant by itself. It is possible that more thorough searching would produce a significant set of indicators, but it would appear that the very powerful block of distance and time measures would swamp any conceivable set of quality measures.

The three models that include X + W and either or both distance and travel time have low residual deviances compared to their degrees of freedom. All three of these models fit the data quite well, particularly the model that includes both distance and travel time. It is, of course, possible that other measures could further reduce the deviance by a statistically significant amount, but a chi square statistic of 38.21 with 51 degrees of freedom is so low (at approximately the 10th percentile of the distribution) that it could easily be attributed entirely to sampling error.

Service Availability as a Mechanism for Development Effects

As was shown earlier, when there is a local distribution program in rural Guatemalan clusters, the contraceptive prevalence rate is 14.6 percent; when no such program exists, the prevalence rate is only 3.5 percent. Clearer evidence of program impact than this 4:1 ratio could not be hoped for. Yet after controlling for the socioeconomic environment, the effect of a local program disappears entirely. How can this finding be interpreted? Does it mean that programs have no impact and are a waste of money? Is it a meaningless statistical artifact?

A distribution program and family planning facilities are not placed in various communities at random, following some kind of experimental design. It is apparent that in Guatemala they are a part of the development process, rather than independent of that process. In terms of the causal model, they are, to a very large degree, the mechanisms by which the socioeconomic environment leads to variations in contraceptive use. As Mauldin, Berelson, Lapham, and others have demonstrated in comparative analyses, the impact of a program usually reflects the cultural and developmental setting (Mauldin, Berelson, and Sykes, 1978; Mauldin and Lapham, 1987).

Table 9 treats clusters as units and shows the proportion of clusters in the sample that have a distribution program according to the type of locality and the presence of a secondary school, post office, or theater or cinema. There is a strong relationship between each of these variables—except for a post office—and the presence of a program. As was seen earlier, contraceptive prevalence is higher in a nuclear aldea, dispersed caserio, and pueblo, and communities with a secondary school or theater/cinema. These also tend to be the communities that have a program—with the notable exception of the 36 dispersed caserios, which tend not to have programs but to have relatively high prevalence rates, a finding that would be interesting to pursue further.

Table 9 Percentage of clusters with a contraceptive distribution program, by type of locality, presence of secondary school, presence of post office, and presence of theater or cinema

	Number of Clusters	Percentage with a Distribution Program
Type of Locality	(Chi square 27.6 with 5 d.f.)	
Finca	15	13
Aldea, nuclear	18	72
Aldea, dispersed	36	19
Caserio, nuclear	6	50
Caserio, dispersed	36	14
Pueblo	5	60
Secondary School	(Chi square 6.2 with 1 d.f.)	
Present	22	50
Absent	94	23
Post Office	(Chi square 0.0 with 1 d.f.)	
Present	30	30
Absent	86	28
Theater or Cinema	(Chi square 7.2 with 1 d.f.)	
Present	21	52
Absent	95	23

There is also a strong relationship with the indicators drawn from the survey of individual women. For example, pooling together the 33 communities with a program, 52 percent of the women have electricity in their homes and 33 percent have a television set. In the remaining 88 clusters without a program, only 20 percent of the women have electricity in their homes and 8 percent have a television set. This is not to imply that a program will lead to higher television ownership or vice versa, but to show that programs tend to be located in the same communities where development and the standard of living are more advanced.

If the presence/absence of a distribution program is regressed on the full set of variables of types X and W, using logit regression and treating clusters as units, the residual deviance is only 89.81 with 95 degrees of freedom (reduced from a baseline deviance of 138.54 with 115 degrees of freedom). This is less than the expected value and so low that the remaining unexplained variation in the presence/absence of a program can be attributed to sampling error.

Similarly, it can be shown that the density of most facilities is much higher in the more developed clusters. Although there is a risk of confusing the causal priority of these variables, it is perhaps easiest to present the association between density of services and electrification, a good indicator of development, by giving the pooled percentage who have electricity in clusters at specified distances from the nearest facilities. Table 10 gives these percentages. It suggests that areas close to facilities will tend to be more modern, with especially strong gradients for pharmacies and private physicians and weak or ambiguous gradients for hospitals and clinics.

Table 10 Percentage of women with electricity in household, by distance from nearest facility of each type

Distance	Hospital	Health Clinic	Center	Private Pharmacy	Physician
0-4 km	6[a]	43[a]	36	41	56
5-9 km	11[a]	43	24	31	34
10-19 km	59	39[a]	10	25	27
20-29 km	31[a]	33[a]	15[a]	17[a]	19[a]
30+ km	23	26	10[a]	3[a]	12[a]
Not named	29	27	52[a]	20	26

[a] Figures based on fewer than 10 clusters.

Because the service availability variables are categorical, it is difficult to carry out a formal decomposition in which coefficients are attached to direct and indirect pathways, as could be done with systems of linear equations. However, after adjusting for the general socioeconomic environment, these are the general findings:

1. The presence of a program and the presence of specific facilities somewhere within 30 km are aspects of general development. In Guatemala, local distribution programs and all types of family planning facilities tend to be distributed around the rural areas in the same way as electrification and other community facilities, such as secondary schools and cinemas. They tend to be found in areas where the population is better educated and better off economically.

2. The density of services, which is assumed to be the critical factor underlying the distance measure, has a major impact on current use that goes beyond the other characteristics of the women and their communities and remains important in itself. The effect of density appears strongest for pharmacies and physicians, both of which are private and which involve resupply of condoms and pills. Although this paper has not yet looked separately at different methods and the specific use of different sources, it appears that hospitals are used more for sterilization—the principal contraceptive method in Guatemala—and gradations of time and distance are less important for that purpose.

3. There is no evidence that variations in the quality of services are so great, given the other major features of the local socioeconomic and family planning environment, that they have any additional impact on contraceptive use. Only five characteristics have even zero-order (controlling for other variables in block Z but not for the X or W types of variables) relationships with current use—namely the number of doctors in the hospital, the number of nurses in the clinics and health centers, the provision of female sterilization in the nearest hospital, and the provision of pills by the nearest private physician. However, these characteristics are closely enough related to the other variables in the model that they have no identifiable residual impact. It is of course possible that statistical significance would be achieved with a larger data base.

Variation within Clusters

The statistical model allows for the effect (upon Y) of interactions between cluster-level variables (Z) and within-cluster deviations on socioeconomic variables (X'). Note that these deviations are calculated from cluster means, rather than from grand means, as in the interaction terms found by Avery (1988). As a rule, deviations from cluster means will be smaller than deviations from grand means. Also, as a rule, interaction effects will be much smaller than main effects, and are rarely significant when main effects are not significant. For these reasons, instead of proceeding to the within-cluster phase of the statistical model, this section will illustrate a different strategy for relating variation within clusters to variation between clusters, taking individuals as the units of analysis.

The survey of individual women contained a question about the latest source of family planning services. In Table 11 this source is cross-tabulated with method used for the 252 current users in the 116 rural clusters. This table corresponds to Table 5.14 of the main report (INCAP and IRD, 1989), except that table refers to all of Guatemala and this table comes from the Standard Recode File, in which sources are grouped into four main categories plus "other".

The types of sources listed in the Standard Recode File, as well as the types listed in the questionnaire itself, do not neatly correspond to the five types in the SA Module. As mentioned previously, the five types of facilities in the SA Module are hospital, clinic, health center, pharmacy, and private physician ("medico particular"). In the individual questionnaire, the woman was given these twelve choices: private hospital, public hospital, Roosevelt Hospital, private clinic, health center, health post, IGSS, Aprofam, pharmacy, health promotor, community distributor, and other place. In the Standard Recode File these were compressed into the following five categories: government clinic/pharmacy, government home/community delivery, private clinic/delivery, private pharmacy, and other. (The headings of Table 11 are slight modifications of these labels.) In view of the lack of consistency in the minds of the architects of these categories, one cannot help but wonder how accurately the interviewer and respondent were able to classify specific facilities. The following strategy for analyzing method and source will be limited by these ambiguous categorizations.

Table 11 Distribution, by current method and source, of the 252 users of modern contraceptive methods in the 116 rural Guatemalan clusters where the SA Module was administered

	Government		Private			
Method	Hospital Center	Distribution Program	Clinic	Pharmacy	Other	Total
Pill	27	10	8	6	1	52
IUD	1	0	12	0	0	13
Injection	0	0	3	0	0	3
Foam, etc.	2	0	2	2	0	6
Condom	2	0	0	13	0	15
Sterilization						
Female	54	0	92	0	3	149
Male	1	0	12	0	1	14
TOTAL	87	10	129	21	5	252

Note: Data on source are drawn from DHS survey of individual women.

Some 87 percent of the current users are located in just seven cells of Table 11, listed here in declining order of frequency:

Number of Women	Method	Source
92	Female sterilization	Private clinic/physician
54	Female sterilization	Govt hospital/center
27	Pill	Govt hospital/center
13	Condom	Private pharmacy
12	IUD	Private clinic
12	Male sterilization	Private clinic
10	Pill	Govt distribution Program

Each of two methods is provided almost entirely by a single source: IUDs are limited to private clinics, and condoms are limited to private pharmacies. Female sterilization, the leading method, is divided between two main sources, either a private clinic/physician or a government hospital/health center. The pill is the method least concentrated among one or two sources, but it is found mostly at government hospitals/health centers. It appears that substitution between different kinds of facilities will be greatest if a woman wishes to use the pill and least if she wishes to use an IUD or condoms.

How can this information be related to that in the SA Module? First comes an investigation of how the method currently used corresponds with information in the Module about the local availability of that method. The more methods that are available locally, the more choices the woman will have, and whatever method the woman is using must, de facto, be available somewhere.

Table 12 examines method choice relative to the local environment of method options for the same group of 252 users. The first column gives the number of current users for each contraceptive method. The remaining columns list the numbers of women in each of the current user groups who are reported by the SA Module to have access to each of the six methods within a 30 km radius. For example, 225 out of the 252 women are reported in the SA Module to have at least one source of condoms within 30 km; only 47 of them have access to male sterilization within 30 km.

Table 12 Contraceptive method currently used matched against availability of the method, among the 252 users of modern methods in the 116 rural Guatemalan clusters where the SA Module was administered

		Number of women for whom method is available at one or more sources within 30 km radius[b]					
						Sterilization	
Current Method[a]	Number of Users	Pill	IUD	Injection	Condom	Female	Male
Pill	52	50[c]	18	19	50	5	4
IUD	13	13	6[c]	6	13	2	2
Injection	3	3	3	3[c]	3	3	1
Foam, etc.	6	5	0	2	5	1	1
Condom	15	13	5	3	13[c]	15	12
Sterilization							
Female	149	125	61	58	128	43[c]	34
Male	14	13	8	4	13	3	3[c]
TOTAL	252	122	101	95	225	62	47

[a] Reported in the individual survey
[b] Reported in the SA survey
[c] Indicates diagonal cell where the number who have access to the method should be compared with the number who are using the method.

Out of the 225 women who have access to condoms, only 13 are using condoms. Clearly condoms are not a preferred method. Of more significance, however, is that the number of current users of condoms is 15. That is, two of the women who *are* using condoms do not, according to the SA Module, have a source within 30 km. Similarly, of the 52 current users of the pill, only 50 have convenient access; the other 2 do not. Of the 13 users of the IUD, less than half (6) have access to the method within 30 km. Likewise, only 43 of the 149 women who are sterilized, and only 3 of the 14 women whose husbands are sterilized, have access within 30 km.

There are two plausible interpretations of this observation that a total of 128, or about half, of the women apparently do not have access to the method they are currently using. One interpretation is that these women (or their husbands, in the case of male sterilization and condoms) are fully prepared to travel beyond the 30 km limit, particularly if the need for services is infrequent. The greatest relative gap exists for sterilization, followed by the IUD, methods which, respectively, require only a single or an annual visit to a provider. Methods requiring the most frequent re-supply, pills and condoms, show the smallest gap of just two women each. Injection shows no gap, but only three women are involved.

Indeed, Table 12 shows almost an inverse relationship between access to a method and the proportion of users who choose that method. The methods can be ranked according to the ratio of the number of users to the number of women with access to the method. Although this is a very crude measure, and is complicated by the fact that degree of access is both a response to demand and a stimulus to demand, it does indicate that much more is involved in the use of a method than access to it:

Method	Ratio of Users to Access
Female sterilization	149/62 = 2.40
Pill	52/122 = .43
Male sterilization	14/47 = .30
IUD	13/101 = .13
Condom	15/225 = .07
Injection	3/95 = .03

Female sterilization, in particular, is apparently sought out, and women are willing to travel some distance for it. Other methods, particularly the condom, injection, and the IUD, are conspicuously rejected in favor of female sterilization and the pill. The clear implication is that if access is to be improved, it should be improved mainly for female sterilization and the pill.

An alternative interpretation of Table 12 is that the SA Module is not turning up service points which are in fact within the 30 km limit or at least are close enough to be accessible to the 128 users whose method is not available within 30 km. It is suspected that some service points have indeed been missed, because the SA Module records only the nearest facility of each type and because of the evidence described earlier that "not named" does not always mean "absent." It is impossible, however, to determine conclusively whether such omissions are important enough to alter the interpretation just given.

The next step would be to investigate how the source currently used corresponds with information in the SA Module about the local availability of that type of source—that is, to examine choice of facility within the local environment of alternative facilities. Table 13 lists the presence of facilities and the current choice in a format analogous to that of Table 12. A table of this sort would be revealing if the same classification system had been used in both questionnaires. Since this obviously was not the case, no interpretation will be attempted here.

Table 13 Source of family planning method currently used matched against access to that source, among the 252 users of modern methods in the 116 rural Guatemalan clusters where the SA Module was administered

		Number of women for whom source is available within 30 kilometer radius[b]				
Current Source[a]	Number of Users	Hospital	Clinic	Hospital Center	Pharmacy	Physician
Hospital or Health Center	87	24	27	80	74	40
Community Distribution	10	2	5	10	10	7
Clinic	129	32	40	107	100	57
Pharmacy	21	3	4	18	16	11
Other	5	3	3	5	5	3
Total	252	64	79	220	205	127

[a] Reported in the individual survey
[b] Reported in the SA survey

Conclusions and Recommendations

This paper has suggested an analytical model that views service availability as an intervening variable mediating the impact of development upon contraceptive use. The model, however, also allows for the additional, direct effect of service availability on contraceptive prevalence, even after the level of local development has been taken into account. An examination of data from rural Guatemala did indeed find strong, additional effects for the nearness to services (mainly from non-hospital sources) within a 30 km radius. Proximity added considerably to the simple presence of such services within 30 km. In contrast, the few indicators of quality were not found to be significant.

This conceptualization of family planning services as an intervening variable—in terms of both their presence and their utilization—is fully consistent with prevailing theoretical frameworks for the determinants of fertility and fertility regulation.[1] However, most analyses of the SA Module thus far have been limited to zero-order relationships or have made statistical adjustments without providing an analytical framework. In their recent comparative analysis using the module, Tsui and Ochoa (1989) found a strong impact of service proximity on contraception, but they did not adjust for the local context or indicate how this might be done. Marcotte (1989), in his analysis of the effect of access to a health center on DPT vaccination in Colombia, did separate analyses for urban and rural areas, for literate and illiterate mothers, and for mothers within 4 km of a health center or farther away, but did not have a model for integrating these determinants into a multivariate analysis. Avery (1988) took

[1] See, for example, Bulatao and Lee (1983, vol. 1, pp. 1-26). Jain (1989) presents an alternative in which services are exogenous rather than endogenous.

into account various individual characteristics but did not aggregate characteristics of the cluster. In their analysis of contraceptive use and method choice in Costa Rica, Hermalin, Riley, and Rosero-Bixby (1989) did use community-level, socioeconomic variables. The main distinction between their approach and ours is that they did not partition variation into that between clusters and that within clusters, and they worked solely with individual women as the units of analysis.

This paper's second intended contribution has been a statistical model that takes into account the true number of data points by examining both variation between clusters and variation within clusters. For a dramatic illustration of the need for this, consider once more the evaluation of the importance of hospitals in rural Guatemala. The usual strategy would be to attach the cluster-level information to the 3,043 women in the rural clusters and look for variation in their use of contraceptives. However, as has already been emphasized, there are only 116 clusters, and the information about hospitals is identical for all the women within any one of these clusters. Thus, the approach has been instead to see whether knowledge of the location of hospitals helps to determine the location of the 252 rural contraceptive users.

The limitation on data about hospitals, however, is even more serious than just indicated. There are only 50 clusters in which a hospital is mentioned, and only 20 different hospitals are named. Moreover, in the 23 clusters located within 30 km of a hospital, only 10 separate hospitals are named! Cuilapa and Nacional hospitals each fall within 30 km of four clusters; Roosevelt and Quiche hospitals, three clusters; and Progreso, Solola, and Jutiapa hospitals, two clusters. Escuintla and Santa Cruz are the only hospitals that are within 30 km of just one cluster.[2] The data quality measures (e.g., number of doctors and number of methods provided) should be the same for all of these hospitals, although the consistency of the data was not actually verified. Thus there are only 10 distinct observations on the quality of hospital services and only 23 distinct observations on variations in access within the 30 km limit. It would be fallacious and highly misleading to proceed as if there were 3,043 independent observations and to use statistical tests that assume such independence.

The approach taken here has almost certainly been excessively lenient in its interpretation of the independence of clusters. It has been assumed that they are statistically independent data points, but the figures just given for hospitals demonstrate that this is probably not a valid assumption. In a random sampling of 116 clusters of 25 to 30 households in rural Guatemala, it seems unlikely that four of them would be within 30 km of one hospital, four within 30 km of another hospital, etc. (To assess the likelihood of this happening by chance, it would be necessary to look at the density of hospitals and the general population in rural Guatemala.) A more thorough analysis would take account of the clustering of the clusters, so to speak, in the sample design.

[2] An unnamed hospital ("sin nombre") is also located within 30 km of one cluster. If it is the same as one of the other hospitals listed, there would be nine, rather than ten distinct hospitals.

The most serious concerns about the SA Module involve its very design as an instrument for evaluating service availability and its impact upon contraceptive use and other outcomes. First and foremost, the effective number of data points for cluster-level data is the number of clusters, not the number of women. This was illustrated earlier with a 95 percent confidence interval for the percentage of women who live in an area with a local distribution program. Avery (1988), Tsui and Ochoa (1989), and others describe access in terms of the number or percentage of women with various kinds of access as if this were assessed separately and independently for each woman. In fact, the standard errors of all estimates of access depend totally upon the number of clusters, *not* the number of women. To further illustrate this point, Table 14 presents estimates of the percentage of women who are within 30 km of each type of facility and 95 percent confidence intervals for these percentages. These intervals have widths of 9.3 percent to 18.0 percent and would be considered unacceptably large for most purposes.

Table 14 Point and interval estimates (with 95 percent confidence) of the percentages of women in rural Guatemala who have access to basic family planning services from a local program or from a source within 30 km

Source	Coefficient	Standard Error	Percent	Confidence Interval (%)	Width of Confidence Interval
Local program	-0.6892	0.1964	33.4	(25.5-42.5)	17.0
Hospital	-1.267	0.2236	22.0	(15.4-30.4)	15.0
Clinic	-0.9299	0.2055	28.3	(20.9-37.1)	16.3
Health Center	2.703	0.3788	93.7	(87.7-96.9)	9.3
Pharmacy	1.579	0.2464	82.9	(75.0-88.7)	13.8
Doctor	0.0493	0.1857	51.2	(42.2-60.2)	18.0

Note: The coefficient and standard error come from logit regression of binary variable (presence vs. absence), with no covariates, with clusters as units and weights proportional to the number of women in the cluster. These are converted from the logit scale to percentages.

Second, the SA Module collects more information about the nearest facility of each type than can possibly be analyzed—there are too many variables for the number of observations—but does not accurately convey the density of services and the degree of substitution of one facility for another. Distance to the nearest facility is an indicator of access, but it is also an indicator of density. Some of the resources spent on collecting detailed information about the nearest facility of each type (and more such details are included in the plans for DHS-II than for DHS-I) would be better invested in finding out whether there is a second or third facility in the vicinity, even if less information were collected about such places.

Third, because of the rapid saturation of degrees of freedom and the difficulty of achieving statistical significance for individual indicators of quality, it is suggested that an overall index of the quality of a facility be developed, perhaps following the criteria suggested by Bruce (1990). A summation scale or other synthesis of "general quality" could be created that would best express the hypothesized importance of quality. There are not enough cases to use data-based scaling procedures such as factor analysis with any confidence.

Fourth, although the cluster-level data recorded by the SA Module and the individual-level data of the basic DHS questionnaire can be easily linked by data processing, there are no genuine links between the women in the survey and the specific facilities mentioned. It is not known whether the woman is aware of a facility or has used it. If she has not visited a facility, perhaps it is because she has never heard of it—not because its hours are limited. If she has visited it but not returned, perhaps it is for reasons completely unrelated to the service quality.

It should be possible to collect the cluster-level data *before* the individual-level data and to provide the interviewers in each cluster with a list of the nearest facilities. Specific questions could then be asked about the knowledge and use of these facilities. The main survey includes questions about local providers, of course, but they cannot at present be reliably linked with the specific facilities described in the SA Module. At the very least, the categorization of facilities should be the same in both surveys.

Fifth, it appears that sometimes facilities were present but were not mentioned. The category "not named" in Table 6 consistently has a higher prevalence rate—sometimes much higher—than the category ">30 km." Whether due to a flaw in the design or the execution, such omissions can completely undermine an analysis with so few units. It is essential that the absence of a facility be distinguished from the failure to locate a facility.

Sixth, the basic strategy of piggy-backing the SA Module onto the DHS cluster design should be re-evaluated, particularly if the goal is to evaluate the impact of the quality of services. This goal would be better met by drawing a sample of providers and then evaluating their ability to attract clients and retain them (for methods other than sterilization). A provider survey would be the most direct way to learn how far and how long people are willing to travel, how they match a method with a facility, which factors lead to a first visit, and which factors lead to continued visits. While the SA Module can shed some light on these questions, it is far from the ideal mechanism for answering them.

References

Avery, Roger C. 1988. Family Planning Services and Contraceptive Behavior in the Dominican Republic. Population Studies and Training Center, Brown University, Providence, R.I.

Billy, John O.G. et al. 1989. Report from Battelle Memorial Institute to National Institute of Child Health and Human Development on contract "The Influence of Contextual Effects on Fertility and Fertility Regulation in the United States."

Blalock, H. 1984. Contextual-Effects Models: Theoretical and Methodological Issues. *Annual Review of Sociology*, 10.

Blalock, H.M. 1985. Cross-Level Analyses. In *The Collection and Analysis of Community Data*, ed. J.B. Casterline, 187-206. Voorburg, Netherlands: International Statistical Institute.

Boyd, L.H. Jr., and G. Iversen. 1979. *Contextual Analysis: Concepts and Statistical Techniques.* Belmont, California: Wadsworth.

Bruce, Judith. 1990. Fundamental Elements of the Quality of Care: A Simple Framework. *Studies in Family Planning* 21(2):61-91.

Bryk A.S., and S.W. Raudenbush. 1987. Application of Hierarchical Linear Models to Assessing Change. *Psychological Bulletin* 101:147-158.

Bulatao, R.A. and R.D. Lee (eds). 1983. *Determinants of Fertility in Developing Countries.* New York:Academic Press.

Farkas, G. 1974. Specification, Residuals and Contextual Effects. *Sociological Methods and Research* 2:333-361.

Hermalin, Albert I., Ann P. Riley, and Luis Rosero-Bixby. 1989. A Multi-Level Analysis of Contraceptive Use and Method Choice in Costa Rica. Population Studies Center, University of Michigan.

Hirschman, Charles, and Philip Guest. 1990. Multi-Level Models of Fertility Determination in Four Southeast Asian Countries: 1970 and 1980. *Demography* 27(3):369-396.

Instituto de Nutrición de Centro América y Panamá (INCAP) [Guatemala] and Institute for Resource Development/Westinghouse (IRD). 1989. *Encuesta Nacional de Salud Materno Infantil 1987.* Columbia, Maryland: IRD.

Jain, Anrudh K. 1989. Fertility Reduction and the Quality of Family Planning Services. *Studies in Family Planning* 20:1-16.

Kish, Leslie. 1965. *Survey Sampling.* New York: John Wiley and Sons, Inc.

Marcotte, John. 1989. The Impact of Community Health Facilities on Prenatal Care and Vaccination in Colombia. Population Studies and Training Center, Brown University, Providence, R.I.

Mason, W.M., G.Y. Wong, and B. Entwisle. 1983. Contextual Analysis Through the Multilevel Linear Model. In *Sociological Methodology 1983-84*, ed. S. Leinhardt, 72-103. San Francisco: Jossey Bass.

Mauldin, W.P., B. Berelson, and Z. Sykes. 1978. Conditions of Fertility Decline in Developing Countries 1965-1975. *Studies in Family Planning* 9(5):89-147.

Mauldin, W.P. and R.J. Lapham. 1987. The Measurement of Family Planning Inputs. In *Organizing for Effective Family Planning Programs*, ed. Robert J. Lapham and George B. Simmons, 545-582. Washington, D.C.: National Academy Press.

McCullagh, P., and J.A. Nelder. 1983. *Generalized Linear Models.* London: Cambridge University Press.

Pullum, T.W. 1989a. Models for Multi-Level Analysis Using Variation Between and Within Levels. Population Research Center, University of Texas at Austin. Included as chapter 11 in Billy, et al. 1989.

Pullum, T.W. 1989b. The Timing of First Births in the Unites States: A Multi-Level Analysis Using The 1982 National Survey of Family Growth. Population Research Center, University of Texas at Austin, included as chapter 12 in Billy, et al. 1989.

Pullum, T.W., J. Billy, and K. Brewster. 1987. An Assessment of Procedures to Detect Contextual Influences on Fertility and Contraception. Paper presented at the 1987 Annual Meeting of the Population Association of America.

Tsui, Amy Ong, and Luis Hernando Ochoa. 1989. Service Proximity as a Determinant of Contraceptive Behavior: Evidence from Cross-National Studies of Survey Data. Paper presented at the IUSSP Seminar on the Role of Family Planning Programs as a Fertility Determinant, Tunis, June 26-30.

Appendix A:
SA Questionnaire for Rural Guatemala

Ministerio de Salud Pública y Asistencia Social/INCAP

ENCUESTA NACIONAL DE SALUD MATERNO INFANTIL

CUESTIONARIO DE LA COMUNIDAD

GUATEMALA - 1987

MUNICIPIO ______________________________

AREA (URBANA=1; RURAL=2; U-R=3) [ ]

NUMERO DE SEGMENTO [ ][ ][ ]

NOMBRE DEL ENTREVISTADOR: ______________________________________ FECHA [ ][ ][ ] (D) (M) (A)

Nº	PREGUNTAS Y FILTROS	CODIGOS	PASE A PGTA.
2	TIPO DE LOCALIDAD NOMBRE: ______________________________	FINCA .... 1 ALDEA .... 2 CASERIO .... 3 PUEBLO 4	-> 3 -> 3
2A	TIPO DE ALDEA O CASERIO	NUCLEAR .... 1 DISPERSO 2	->104
3	NUMERO DE HABITANTES DE LA LOCALIDAD	MENOS DE 500 .... 1 500 A <2.000 .... 2 2.000 A <5.000 .... 3 5.000 A <10.000 .... 4 10.000 A <20.000 .... 5 20.000 A <50.000 .... 6 50.000 A <100.000 .... 7 100.000 O MAS 8	 -> 7 (6, 7, 8)
4	PRINCIPAL VIA DE ACCESO A LA LOCALIDAD	CARRETERA PAVIMENTADA .... 1 CARRETERA DE TERRACERIA EN BUEN ESTADO .... 2 CARRETERA DE TERRACERIA EN MAL ESTADO .... 3 CARRETERA DESTAPADA O CARRETEABLE .... 4 CAMINO VECINAL O CAMINO DE HERRADURA .... 5 VIA FLUVIAL 6	
5	DISTANCIA EN KILOMETROS A LA LOCALIDAD MAS CERCANA DE 20.000 HABITANTES O MAS	MENOS DE 10 .... 1 10 A <20 .... 2 20 A <30 .... 3 30 A <50 .... 4 50 A <100 .... 5 100 Y MAS 6	
6	TIPOS DE TRANSPORTE PUBLICO COMUNMENTE USADOS PARA LLEGAR A LA LOCALIDAD MAS CERCANA DE 20.000 HABITANTES O MAS. (Señale todas las alternativas que se apliquen)	BUS O CAMIONETA .... 1 TAXI .... 1 BOTE O LANCHA .... 1 NINGUNO 1	

Nº	PREGUNTAS Y FILTROS	CODIGOS	PASE A PGTA.
7	¿Está cubierto este segmento por un programa de distribución comunitaria de anticonceptivos?	SI .... 1 NO 2	 -> 8
7A	¿Que métodos anticonceptivos tienen disponibles y cuál es su costo?	METODOS / COSTO ______________ [ ][ ].[ ][ ] ______________ [ ][ ].[ ][ ] ______________ [ ][ ].[ ][ ]	
8	¿Está cubierto este segmento por un promotor voluntario de planificación familiar?	SI .... 1 NO 2	 -> 10
8A	¿Cada cuánto tiempo visita el promotor el segmento o el vecindario?	NUMERO DE VECES POR AÑO [ ][ ]	
8B	¿Qué métodos anticonceptivos distribuye el promotor y cuál es su costo?	METODOS / COSTO ______________ [ ][ ].[ ][ ] ______________ [ ][ ].[ ][ ] ______________ [ ][ ].[ ][ ]	
10	¿Está cubierto el segmento por una comadrona o partera?	SI............................ 1 NO............................ 2	

12 DISPONIBILIDAD DE SERVICIOS PUBLICOS MAS CERCANOS AL SEGMENTO	EXISTEN EN LA LOCALIDAD SI NO (1)	DISTANCIA EN KILOMETROS (Menos de 1 anote 00; 97 o más anote 97) (2)	TRANSPORTE MAS COMUN Motorizado = 1 Animal = 2 A pié = 3 Bicicleta = 4 A otro = 5 (3)	TIEMPO QUE SE DEMORA PARA LLEGAR HASTA ALLI (en minutos) (4)	COSTO DEL TRANSPORTE (5)
A. EDUCACION					
1. Escuela primaria	1 2	[ ][ ]	[ ]	[ ][ ][ ]	[ ].[ ][ ]
2. Escuela secundaria	1 2	[ ][ ]	[ ]	[ ][ ][ ]	[ ].[ ][ ]
3. Superior / Técnica	1 2	[ ][ ]	[ ]	[ ][ ][ ]	[ ].[ ][ ]
A. SERVICIOS GENERALES					
1. Oficina de correos	1 2	[ ][ ]	[ ]	[ ][ ][ ]	[ ].[ ][ ]
2. Mercado semanal	1 2	[ ][ ]	[ ]	[ ][ ][ ]	[ ].[ ][ ]
3. Teatro o cine	1 2	[ ][ ]	[ ]	[ ][ ][ ]	[ ].[ ][ ]
4. Alcantarillado	1 2				
5. Recolección de basuras	1 2				

13 DISPONIBILIDAD DE SERVICIOS DE SALUD MAS CERCANOS AL SEGMENTO	EXISTEN EN LA LOCALIDAD (1)	DISTANCIA EN KILOMETROS DESDE EL SEGMENTO Menos de 1=00 97 y más =97 No sabe =98 (2)	TIPO DE TRANSPORTE MAS COMUN (3)	TIEMPO QUE SE DEMORA EN LLEGAR HASTA ALLI Menos de 1=000 No sabe =998 (4)	NUMERO DE DOCTORES (MEDICOS) No tiene = 00 97 y más = 97 No sabe = 98 (5)	NUMERO DE ENFERMERAS No tiene = 00 97 y más = 97 No sabe = 98 (6)	DIAS EN QUE ESTA ABIERTO AL PUBLICO (7)	TOTAL DE HORAS DE SERVICIO AL PUBLICO (8)	SERVICIOS DISPONIBLES (9)	AÑO EN QUE COMENZO A OPERAR EL SERVICIO (10)
A HOSPITAL (Nombre) __________ __________	SI ...... 1 NO ...... 2 ¿Dónde? __________ __________ __________ (Localidad)	__|__ Si 30 Kms. o más	Motorizado 1 Animal 2 A pié 3 Bicicleta 4 Otro 5	__|__|__ (en minutos)	__|__	__|__	Lunes 1 Martes 1 Miercoles 1 Jueves 1 Viernes 1 Sabado 1 Domingo 1 Otro 1	__|__ __|__ __|__ __|__ __|__ __|__ __|__ __________	Materno/ Infantil 1 Emergencia 1 Planif. Familiar 1 Rehidratacion Oral 1 General 1	__|__ __|__ __|__ __|__ __|__
B CLINICA (Nombre) __________ __________	SI ...... 1 NO ...... 2 ¿Dónde? __________ __________ __________ (Localidad)	__|__ Si 30 Kms. o más	Motorizado 1 Animal 2 A pié 3 Bicicleta 4 Otro 5	__|__|__ (en minutos)	__|__	__|__	Lunes 1 Martes 1 Miercoles 1 Jueves 1 Viernes 1 Sabado 1 Domingo 1 Otro 1	__|__ __|__ __|__ __|__ __|__ __|__ __|__ __________	Materno/ Infantil 1 Emergencia 1 Planif. Familiar 1 Rehidratacion Oral 1 General 1	__|__ __|__ __|__ __|__ __|__
C CENTRO DE SALUD (Nombre) __________ __________	SI ...... 1 NO ...... 2 ¿Dónde? __________ __________ __________ (Localidad)	__|__ Si 30 Kms. o más	Motorizado 1 Animal 2 A pié 3 Bicicleta 4 Otro 5	__|__|__ (en minutos)	__|__	__|__	Lunes 1 Martes 1 Miercoles 1 Jueves 1 Viernes 1 Sabado 1 Domingo 1 Otro 1	__|__ __|__ __|__ __|__ __|__ __|__ __|__ __________	Materno/ Infantil 1 Emergencia 1 Planif. Familiar 1 Rehidratacion Oral 1 General 1	__|__ __|__ __|__ __|__ __|__
D FARMACIA (Nombre) __________ __________	SI ...... 1 NO ...... 2 ¿Dónde? __________ (Localidad)	__|__ Si 30 Kms. o más	Motorizado 1 Animal 2 A pié 3 Bicicleta 4 Otro 5	__|__|__ (en minutos)					¿Hay sales de rehidratación oral? SI .... 1 -> NO 2	__|__
E MEDICO PARTICULAR (Nombre) __________ __________	SI ...... 1 NO ...... 2 ¿Dónde? __________ (Localidad)	__|__ Si 30 Kms. o más	Motorizado 1 Animal 2 A pié 3 Bicicleta 4 Otro 5	__|__|__ (en minutos)					¿Tiene sales de rehidratación oral disponibles? SI .... 1 -> NO 2	__|__

14 DISPONIBILIDAD DE SERVICIOS DE PLANIFICACION FAMILIAR MAS CERCANOS AL SEGMENTO	EXISTEN EN LA LOCALIDAD (1)	DISTANCIA EN KILOMETROS DESDE EL SEGMENTO < de 1 = 00 97 y más = 97 No sabe = 98 (2)	TIPO DE TRANSPORTE MAS COMUN (3)	TIEMPO QUE SE DEMORA PARA LLEGAR HASTA ALLI < de 1 = 000 No sabe = 998 (4)	METODOS DISPONIBLES (5)	COSTO DE LOS METODOS Contribucion = 77 o 777 Gratis = 88 o 888 No sabe = 98 o 998 (6)	AÑO EN QUE COMENZO A OFRECERSE EL METODO POR PRIMERA VEZ (7)	NUMERO DE DOCTORES (MEDICOS) PARA PLANIFICACION FAMILIAR No tiene = 00 97 y más = 97 No sabe = 98 (8)	NUMERO DE ENFERMERAS PARA PLANIFICACION FAMILIAR No tiene = 00 97 y más = 97 No sabe = 98 (9)	DIAS DE SERVICIO PARA PLANIFICACION FAMILIAR (10)	TOTAL DE HORAS DE SERVICIO PARA PLANIFICACION FAMILIAR (11)
A HOSPITAL (Nombre) ________ ________	SI ...... 1 NO ...... 2 ¿Dónde? ________ ________ (localidad)	|_|_| Si 30 Kms. o más	Motorizado 1 Animal 2 A pié 3 Bicicleta 4 Otro 5	|_|_|_| (en minutos)	Píldora 1 DIU 1 Inyección 1 Condón 1 Esteril. Femenina 1 Esteril. Masculina 1 Otro 1			|_|_|	|_|_|	Lunes 1 Martes 1 Miercoles 1 Jueves 1 Viernes 1 Sabado 1 Domingo 1 Otro 1	________
B CLINICA APROFAM 1 OTRA 2 Nombre ________ ________	SI ...... 1 NO ...... 2 ¿Dónde? ________ ________ (localidad)	|_|_| Si 30 Kms. o más	Motorizado 1 Animal 2 A pié 3 Bicicleta 4 Otro 5	|_|_|_| (en minutos)	Píldora 1 DIU 1 Inyección 1 Condón 1 Esteril. Femenina 1 Esteril. Masculina 1 Otro 1			|_|_|	|_|_|	Lunes 1 Martes 1 Miercoles 1 Jueves 1 Viernes 1 Sabado 1 Domingo 1 Otro 1	________
C CENTRO DE SALUD (Nombre) ________ ________ ¿Es abastecido por APROFAM? SI ...... 1 NO 2	SI ...... 1 NO ...... 2 ¿Dónde? ________ ________ (localidad)	|_|_| Si 30 Kms. o más	Motorizado 1 Animal 2 A pié 3 Bicicleta 4 Otro 5	|_|_|_| (en minutos)	Píldora 1 DIU 1 Inyección 1 Condón 1 Esteril. Femenina 1 Esteril. Masculina 1 Otro 1			|_|_|	|_|_|	Lunes 1 Martes 1 Miercoles 1 Jueves 1 Viernes 1 Sabado 1 Domingo 1 Otro 1	________
D FARMACIA (Nombre) ________ ________	SI ...... 1 NO ...... 2 ¿Dónde? ________ ________ (localidad)	|_|_| Si 30 Kms. o más	Motorizado 1 Animal 2 A pié 3 Bicicleta 4 Otro 5	|_|_|_| (en minutos)	Píldora 1 DIU 1 Inyección 1 Condón 1 Esteril. Femenina 1 Esteril. Masculina 1 Otro 1						
E MEDICO PARTICULAR (Nombre) ________ ________ ¿Es abastecido por APROFAM? SI ...... 1 NO 2	SI ...... 1 NO ...... 2 ¿Dónde? ________ ________ (localidad)	|_|_| Si 30 Kms. o más	Motorizado 1 Animal 2 A pié 3 Bicicleta 4 Otro 5	|_|_|_| (en minutos)	Píldora 1 DIU 1 Inyección 1 Condón 1 Esteril. Femenina 1 Esteril. Masculina 1 Otro 1						

Appendix B:
Statistical and Computing Details

The Basic Statistical Model

Introduction

This section, extracted from Pullum (1989a), will describe the logic of the statistical model in terms of an interval-level dependent variable. The next section will make modifications for a binary dependent variable, which is the type used in the body of the paper. The body of the paper refers to variables of type W, as well as Y, X, and Z. Here it is not necessary to distinguish between types W and Z, so only the latter will be discussed.

Individuals (within contexts) and contexts will be referred to with the subscripts i and j, respectively. The words "micro" and "macro" will sometimes be used to describe the two levels of analysis. Three basic classes of variables will be distinguished, two at the individual level and one at the group level.

Let Y_{ij} be the dependent variable for person i in group j, for example, desired family size or the type of contraceptive currently used. Assume it is measured as an interval-level variable. Its mean in group j is $Y_{.j}$.

Let X_{ij} be a micro level predictor, for example, years of education, with the sole restriction that X_{ij} does in fact vary across individuals within each group j. Its mean in group j is $X_{.j}$.

The other class of variable assumed to be related to the dependent variable is Z_j, which may be described as an integral macro variable. That is, it is an intrinsically macro-level characteristic (for example, the presence or absence of a subsidized family planning clinic) which is not built up from individual measurements.

The unit of analysis for the statistical work is defined by the dependent variable, which is assumed to be measured at the level of the individual. The goal is to generate predicted values which can be compared with the observed values; this must be done for individuals.

Substantive theory guides the selection of variables and how they are ordered within classes. For example, X_{ij} should be regarded as a generic element from a class of variables $X1_{ij}$, $X2_{ij}$, etc. These various predictors can be ordered into a causal model on the basis of theoretical reasoning and empirical evidence. Most fertility models make no specific mention of context, and model building operates entirely within this class of variables—a severely restricted class, it may be argued. Similarly, the ordering of variables within type Z_j is a matter for more general theory.

Earlier Contextual Models

Existing contextual models are primarily of two types, which for convenience will simply be labelled conventional linear modelling (CLM) and hierarchical linear modelling (HLM).

Conventional linear modelling consists of treating individuals as the units of analysis, attaching or linking the characteristics of aggregates to the individuals, and estimating the parameters of a single equation. The importance of the context is indicated by the coefficient for the aggregate or group-level variables, with particular attention paid to interactions, represented as products of the group-level and individual-level variables. The importance of the context is also expressed in terms of the increase in the explained sum of squares when the group-level variables are added to the model.

Specifically, in the general model for the kinds of variables described here, Y_{ij} is regressed upon X_{ij}, and upon the group-level variables $X_{\cdot j}$ and Z_j, and upon the interaction terms $X_{ij}X_{\cdot j}$ and $X_{ij}Z_j$. The model can include interactions among variables defined solely at the individual level or interactions among variables defined solely at the aggregate level, but these are incidental to the terms specified. Expressed as an estimated equation, the general model has the form

$$Y_{ij} = b_0 + b_1X_{ij} + b_2X_{\cdot j} + b_3Z_j + b_4X_{ij}X_{\cdot j} + b_5X_{ij}Z_j + e_{ij},$$

where e_{ij} is a residual or error term. This model has been discussed in detail by Boyd and Iversen (1979) and by Blalock (1984, 1985). Boyd and Iversen limited themselves to predictor variables of types X_{ij} and $X_{\cdot j}$, severely limiting the scope of a contextual analysis, but the inference is that if they had included a variable of type Z_j, then they would also have included an interaction term such as $X_{ij}Z_j$.

Hierarchical linear modelling has been used in contextual analysis by Mason, Entwisle, Hermalin, and Wong (see, for example, Mason et al., 1983) and by Bryk and Raudenbush (1987). Again, specific applications vary somewhat, but the essence of this formulation is two sets of equations, one at the micro level and the other at the macro level. For the generic variables, the micro-level model is a set of estimated equations of the form

$$Y_{ij} = a_{0j} + a_{1j}X_{ij} + d_{ij} \quad ,$$

with one such equation for each group j, and the macro-level model is a set of two estimated equations of the form

$$a_{0j} = c_{00} + c_{01}X_{\cdot j} + c_{02}Z_j + f_{0j}$$

and $$a_{1j} = c_{10} + c_{11}X_{\cdot j} + c_{12}Z_j + f_{1j}.$$

($X_{.j}$ is not always included in the macro model.) When the macro equations are substituted into the micro model, the following equation is obtained:

$$Y_{ij} = c_{00} + c_{10}X_{ij} + c_{01}X_{.j} + c_{02}Z_j + c_{11}X_{ij}X_{.j} + c_{12}X_{ij}Z_j + [(d_{ij} + f_{0j}) + X_{ij}f_{1j}].$$

Here d_{ij}, f_{0j}, and f_{1j} are error terms. This combined equation differs from the conventional linear model in two important ways. First, it is generated by two levels or sets of equations with their own error terms. The parameters which are estimated in the first equation are regarded as random variables in their own right, with a specified distribution (for example, a normal distribution) across aggregates. Second, the two-level specification requires a modification of the estimation procedure using special computer programs and empirical Bayes assumptions about the prior distribution of the parameters in the micro equation.

Strong arguments can be made for either the CLM or the HLM approach, and the objective here is to build upon both of them.

Boyd and Iversen (1979) stated their approach in terms of micro and macro equations (omitting a variable of type Z_j), but the equation they actually estimated, using ordinary least squares, was the single-equation CLM as given above. Their equations for the coefficients did not include any error components, so the error structure after substitution remained simple. For that reason their work is in the CLM tradition rather than an example of HLM.

The hierarchical model is an improvement, to be sure, but its specification of the macro model, in which regression coefficients are redefined to be randomly distributed dependent variables, is difficult to interpret. There is no clear parallel between the micro and macro models, because the dependent variables in the macro model express the *relationships* among micro variables rather than the *levels* of variables. There is particular difficulty interpreting the macro level equation for the intercept term (a_{0j}), because in any estimation procedure the intercept will be a function of the means ($Y_{.j}$ and $X_{.j}$) and the slope (a_{1j}; specifically, $a_{0j} = Y_{.j}-a_{1j}X_{.j}$) and is often regarded as a nuisance parameter.

The HLM is difficult to estimate and very difficult to extend to additional levels of aggregation. The promoters of HLM have not indicated the sensitivity of the conclusions to alternative specifications of the underlying prior distribution of the micro-level parameters, or the impact upon the final parameter estimates of using HLM rather than the much simpler CLM.

More importantly, the statement of the micro and macro models, whether by Boyd and Iversen (1979) or by Mason et al. (1983), does not lead to any kind of partitioning of explained variation into components for the micro and macro levels which would differ from the usual partitioning into components for different variables, regardless of level.

Hierarchical linear modelling was initially developed by Dempster, Rubin, Laird, and other Bayesian statisticians for reasons far removed from a desire to model contextual effects. They desired to improve or "tighten" within-group estimates of relationships between micro-level variables when the within-group samples were small. In order to do this, they assumed that the parameters being estimated for each group are themselves distributed systematically according to a distribution, perhaps normal or log-normal or beta, with so-called hyper-parameters. If the hyper-parameters were further hypothesized to be conditional upon some other variable, then a linear term for such a variable could be included in the "macro" equation.

The empirical Bayes framework allows information about the collection of all groups to be used to adjust within-group relationships, typically adjusting "wild" within-group estimates from a small group to bring them closer to the overall mean estimate. In other words, the intended purpose of these procedures is to improve the estimates of the parameters of the micro model, especially for the groups with small samples. For this purpose, the interpretation of the macro model and its parameters is not problematic.

In contextual analysis, the main interest is in the parameters of the macro model, rather than the far greater number of parameters in the micro model, and there is less reason to employ empirical Bayes estimation procedures. The estimates from the CLM (the b's) will differ very little from the corresponding macro-level estimates from the HLM (the c's) because both sets of estimates are based on the pooling of all groups. When the within-group samples are large (for example, if they arise from separate World Fertility Survey samples), even the within-group estimates will be affected only slightly by the use of HLM rather than separate least-squares estimation within each group.

A New Approach

We conceptualize a model which may be characterized as hierarchical analysis of covariance with random effects. (The term "random effects" is used because in most applications, the groups for which data are available within each level, e.g. counties or clusters, are a random sample of a population of groups at that level.)

Our first and perhaps most fundamental observation is that the distinction between micro and macro variables compels a distinction between within-group and between-group sources of variation. Because they are, by definition, the same for all persons in group j, a group mean or a macro variable cannot possibly be a cause of variation ***within*** group j *except* through the mechanism of an interaction with a micro-level predictor. Otherwise, the effect can only be on the mean level of Y_{ij} within group j, that is, on $Y_{.j}$. This is an important assertion, which, although elementary, appears to distinguish our approach from most others. For example, Blalock (1984) discusses a minimal form of the CLM, the model $Y_{ij} = a + b_1X_{ij} + b_2X_{.j} + e_{ij}$. Since $X_{.j}$ is the same for all individuals in group j, it cannot possibly be a source of within-group variation in Y. It may indeed affect the group mean $Y_{.j}$, and thereby the score Y_{ij}, but that assertion should be expressed in a separate equation.

This observation implies that a model should be stated on two levels, corresponding to the two terms Y_{ij} and $Y_{.j}$. We prefer to work with the two statistically independent terms on the right hand side of the decomposition,

$$Y_{ij}-Y_{..} = (Y_{ij}-Y_{.j}) + (Y_{.j}-Y_{..}) = Y'_{ij} + Y''_{.j}.$$

A prime (') will be used to indicate a deviation of an individual value from a group mean and a double prime (") to indicate a deviation of a group mean from an overall mean. First, we have the *micro model*, in which the deviation $Y_{ij}-Y_{.j} = Y'_{ij}$ is expressed as a function of deviations of the form $X_{ij}-X_{.j} = X'_{ij}$. This function may bring in interactions with Z_j, which may be indicated as products but are not limited to that form. This equation is *not* estimated in each context or group, as in the approach of some other researchers (e.g., Mason et al., 1983), but once for all contexts combined. In addition, the individual-level predictors may be stated as functions of one another and of the aggregate variables.

The second equation will be referred to as the *macro model* and states that $Y''_{.j}$ is a function of the mean of the individual-level predictor and the values of the other variables. It may be reasonable to specify some interaction terms as well, corresponding to the ones in the micro model. In additional equations, the macro level predictors (expressed as deviations from their grand means) may be stated as functions of one another.

A statistical test would work from the decomposition of the total sum of squares

$$\begin{aligned}\Sigma(Y_{ij}-Y_{..})^2 &= \Sigma(Y_{ij}-Y_{.j})^2 + \Sigma N_j(Y_{.j}-Y_{..})^2 \\ &= \Sigma Y'^2_{ij} + \Sigma N_j Y''^2_{.j},\end{aligned}$$

where N_j is the number of cases in group j.

The basic idea of combining a micro model and a macro model is certainly not new with us. As described above, it is the basis of the book by Boyd and Iversen (1979), although their work has the serious flaw that they actually estimated only a single combined equation, and it is also central to the approach of Mason, Wong, and Entwisle (1983). Boyd and Iversen sometimes express predictor variables as deviations, using the term "centering," but they do not proceed to express the dependent variable as a deviation and make the decomposition given above.

Our approach differs from those in the literature in the following important ways. First, for Boyd and Iversen and Mason et al., the interactions between $X_{.j}$ and the macro predictors emerge from a substitution process. All of the micro and macro equations are initially assumed to be linear, and the interaction terms arise when the macro equations for the coefficients are substituted back into the micro equation. By contrast, we allow for the inclusion of interaction terms in the basic equations. There is no loss of conceptual clarity if an interaction is assumed directly in the initial equation rather than assumed indirectly by the form of the second equation and "derived" by a substitution. In virtually every other use of a product or interaction term in regression, the interaction is assumed initially rather than "derived", even though it always *could* be generated with an extra step which would complicate the error structure. We also allow the interaction to be totally general, not just multiplicative.

Second, their regression equation for the Y intercept is equivalent to an equation for $Y_{.j}$, because the Y intercept is equal to $Y_{.j}$ minus a weighted sum of the means of the predictors (the weights being the regression coefficients). These authors never make it explicit that this secondary regression is fundamentally a regression of $Y_{.j}$ on the means of the other predictors. If the macro model of these authors is taken literally, it is totally expressed in terms of the coefficients from the micro model and appears to ignore the more basic statistical quantities, the means. Surely, a mean is much more interpretable than a Y intercept, and we are explicit in our use of the means.

Most important, we express variables as deviations and develop causal models at each level of aggregation, with the possibility of cross-overs between levels.

We shall now provide some greater specificity, using the generic variables Y_{ij}, X_{ij}, and Z_j.

Step 1. Partitioning of Variation. Assume that there are individuals labeled i and one level of aggregation with groups labeled j; a micro level dependent variable Y_{ij}; a micro level predictor variable X_{ij}; and a macro predictor Z_j.

The deviation between Y_{ij} and the grand mean $Y_{..}$ is partitioned into within-group and between-group deviations:

$$Y_{ij}-Y_{..} = (Y_{ij}-Y_{.j}) + (Y_{.j}-Y_{..}) = Y'_{ij} + Y''_{.j} \,. \tag{1}$$

Step 2. The Micro Model. Next, equations are developed for the components given above, beginning with the within-group term $Y_{ij}-Y_{.j} = Y'_{ij}$. This equation is referred to as the micro equation because its dependent variable is at the micro level, even though it contains macro predictors in the form of interactions. The model refers to all groups, so the coefficients do not require subscripts for group j.

The micro level equation is based on two assumptions: first, that all variables should be expressed as deviations from their means at the next level of aggregation, as a kind of purging of main effects; and, second, that the effect of X_{ij} upon Y_{ij} (more precisely, the effect of X'_{ij} on Y'_{ij}) depends upon the level of the macro variable Z_j, which for consistency is expressed as a deviation from its mean at the next higher level of aggregation. This deviation is labeled Z"; there is no term labeled Z'. An interaction of the form $X'_{ij}X''_{\cdot j}$ is also included because $X''_{\cdot j}$ can be regarded as a macro variable in its own right. Because of the use of deviations, the Y-intercept in this equation will be zero and can be omitted.

A main effect for Z is deferred to the macro model. It can be shown that if such an effect were included here, its estimate would be zero. Therefore

$$Y'_{ij} = b_{yx}X'_{ij} + b_{yxx}X'_{ij}X''_{\cdot j} + b_{yxz}X'_{ij}Z''_{j} + e_{yij}, \tag{2}$$

where e_{yij} is the residual or unexplained variation in Y'_{ij}. It is important that the interaction term be included. This is the only location in which it is plausible to insert an interaction between the micro and macro predictors.

The micro model also can include equations for the individual-level predictors. In this case there is only one such predictor, $X_{ij}-X_{\cdot j} = X'_{ij}$. From the preceding discussion, it should be evident that there is no plausible source of within-group variation in X with the variables at hand, so there can be no other equations.

Step 3. The Macro Model. Next the deviation $Y_{\cdot j}-Y_{\cdot\cdot} = Y''_{\cdot j}$ is expressed as a function of the group-level deviations:

$$Y''_{\cdot j} = c_{yx}X''_{\cdot j} + c_{yz}Z''_{j} + e_{yj}\,. \tag{3}$$

Here, e_{yj} is a residual term; it is required to sum to zero, so a constant term can be omitted from the equation. All of the values of variables here are specific for group j. It is possible to include an interaction between Z and the group mean of X, but this is not necessary and would not be interpretable as a contextual effect.

A similar equation is formed for each of the micro level predictors; in this simple case there is only one:

$$X''_{\cdot j} = c_{xz}Z''_{j} + e_{xj}\,. \tag{3'}$$

Additional equations can be developed in which groups j are the units. Direct and indirect effects and the other products of causal models are applicable exactly as if the units were individuals, except for the important distinction that in the macro-level analysis, group j must be weighted in proportion to the size of the group, N_j.

Step 4. The Reconstituted Equation. Addition of (2) and (3) produces an equation for Y_{ij}-Y..:

$$Y_{ij}\text{-}Y.. = b_{yx}X'_{ij} + c_{yx}X''_{\cdot j} + b_{yxx}X'_{ij}X''_{\cdot j} + c_{yz}Z''_{j} + b_{yxz}X'_{ij}Z''_{j} + e_{yij} + e_{yj} \quad (4)$$

The coefficients on the right hand side may be interpreted as the following kinds of influences on the deviation Y_{ij}-Y.. or, equivalently, on the individual outcome Y_{ij}:

- b_{yx} is the effect of individual variation in X;
- c_{yx} is the effect of group variation in X;
- b_{yxx} is the effect of interactions between individual variation in X, on the one hand, and group variation in X, on the other hand;
- c_{yz} is the effect of group variation in Z;
- b_{yxz} is the effect of interactions between individual variation in X, on the one hand, and group variation in Z, on the other hand; and
- $e_{yij} + e_{yj}$ is the residual.

Except for b_{yx} (and the residuals), all of these terms can be regarded as indicating contextual effects.

Modifications to the Basic Model for a Binary Outcome

Introduction

Next to be considered are modifications to the statistical model when the dependent variable is binary—for example, whether or not the woman is a current user of modern contraception. Many outcomes related to fertility and health are binary in nature. For a more complete discussion and an application to the timing of the first birth in the United States, see Pullum (1989b).

The most commonly employed transformation of a binary outcome is the logit of the conditional probability that the criterion outcome has occurred. If Y=1 for occurrence and Y=0 for non-occurrence, then the logit of Y is logit(Y)=log[Pr(Y=1)/Pr(Y=0)]. This transformation will be expressed as an additive or linear function of selected covariates and interactions among those covariates.

Here, as in virtually every analysis of individual-level data, the statistical analysis works from a measure of the variability in the dependent variable and then expresses both the substantive importance and the statistical significance of a predictor in terms of the extent to which the inclusion of that predictor reduces variation in the dependent variable. A general label for the original variation in the dependent variable, in the absence of any predictors, is "baseline deviance." For the most familiar case of a normally distributed, interval-level variable, the baseline deviance is the usual sum of squared deviations about the mean.

The total sum of squares can be motivated as the appropriate element (for linear models with normally distributed errors) from a larger class of measures, referred to as deviances, for the family of exponential models (McCullagh and Nelder, 1983). For different kinds of dependent variables and error distributions, the deviance will be defined differently. In general, the baseline deviance is -2 times the natural logarithm of the maximized value of the likelihood function for the baseline model in which no covariates are employed. For the most familiar models (cursorily designated as normal) as well as for the models used in this paper (designated as logit probability), the deviances have chi-square distributions.

Let Y be the dependent random variable, which is 1 if a woman is a current user of modern contraception and 0 otherwise. The general model allows for each woman to have a risk R of having the outcome Y=1. In fact, R is 1 for all women with this type of outcome. However, if Y were an outcome such as a child death or a contraceptive failure, R would sometimes be set to a value between 0 and 1 to adjust for censoring. Therefore, R shall be included in this discussion, but it should be understood that R=1 identically for all women.

For a logit probability model, Y is assumed to have a binary distribution with denominator R ($0<R\leq 1$), and the logit of the probability that Y=1 is a linear function of any covariates. In the baseline model, $\text{logit}(\hat{Y}/R)=\log[\hat{Y}/(R-\hat{Y})]=b_0$, where $\hat{Y}$ is the fitted value of Y and b_0 is a constant. Therefore, $\hat{Y}=\exp(b_0)/[R+\exp(b_0)]$ for a woman with risk R, and, in particular, if R=1, then $\hat{Y}=\exp(b_0)/[1+\exp(b_0)]$.

If a binary variable Y with risk R has a logistic distribution, then its deviance (McCullagh and Nelder, 1983:25) is

$$-2\Sigma[Y_i \ln(Y_i/\hat{Y}_i) + (R_i-Y_i) \ln(R_i-Y_i)/(R_i-\hat{Y}_i)].$$

Here Y_i is the observed value of Y, and $\hat{Y}_i$ is the fitted value for woman i. Since an observed outcome Y_i can only equal 0 or 1, the contribution to the deviance for a case with observed value $Y_i=1$ simplifies to $-2\ln(\hat{Y}_i)$, and the contribution for a case with observed value $Y_i=0$ simplifies to $-2R_i\ln(R_i-\hat{Y}_i)$. When there are no covariates, $c=\Sigma Y_i/\Sigma R_i$ is the overall estimated probability of the outcome, and $\hat{Y}_i=R_ic$ is the estimated probability of the outcome for woman i, who has risk R_i.

The Baseline Deviance when Clusters are Units

It is well known that the maximum influence of characteristics of place of residence will be given, in the case of an interval-level dependent variable Y, by a one-way analysis of variance on Y across the alternative places or contexts. The proportion of variation accounted for by this breakdown, referred to as eta-squared or E^2, is the maximum that could be due to all possible covariates at the contextual level. Our own starting point will indeed be a breakdown by place of residence, but since our dependent variable is binary, rather than interval-level, an alternative to one-way analysis of variance is required.

The logit probability approach uses the logit of the conditional probability P_{ij} (for woman i in cluster j) of being a user. The corresponding probability for a randomly selected woman in the cluster is $P_{.j}$, the average of the probabilities for the women in cluster j. The data are collapsed into a single record for each cluster, putting onto that record the total risk $R_{.j}$ in the cluster (i.e., the total number of women in the cluster, for the present kind of outcome) and the total number in the cluster with outcome $Y_{ij}=1$, namely $Y_{.j}$.

The *baseline deviance* arises when the individual-level probability P_{ij} is estimated with the overall sample proportion and this is used to calculate a fitted value for the binary outcome Y_{ij} when the woman has risk R_{ij}. The formula for this deviance is

$$-2 \Sigma \{Y_{ij} \ln(Y_{ij}/\hat{Y}_{ij}) + (R_{ij}-Y_{ij}) \ln[(R_{ij}-Y_{ij})/(R_{ij}-\hat{Y}_{ij})] \}$$

with $\hat{Y}_{ij}$ equal to $R_{ij}(Y_{..}/R_{..})$, where $Y_{..}$ and $R_{..}$ are the overall sums of Y_{ij} and R_{ij}, respectively. This is precisely the same formula given earlier for the baseline deviance, except that earlier the cluster identifier j was omitted. This baseline deviance can be partitioned into two parts, referred to as the "within-cluster" and "between-cluster" deviances.

The *within-cluster deviance* is produced by generating the fitted values from the **cluster** proportions, i.e., by using the above formula with $\hat{Y}_{ij}$ estimated by $R_{ij}(Y_{.j}/R_{.j})$, where $Y_{.j}$ and $R_{.j}$ refer to the sums of Y_{ij} and R_{ij}, respectively, within cluster j. The complementary *between-cluster deviance* is produced by taking the counties as units in the general formula, with Y_{ij}, R_{ij}, and $\hat{Y}_{ij}$ replaced respectively with totals $Y_{.j}$, $R_{.j}$, and $\hat{Y}_{.j}$, and estimating $\hat{Y}_{.j}$ with $R_{.j}$ $(Y_{..}/R_{..})$. Simple algebra will verify that these two components add to the total baseline deviance.

The fact that these components are interpretable as deviances in their own right implies that the overall model can be expressed with two sub-models, one in which the within-cluster deviance is a new "baseline" deviance to be reduced by within-cluster sources of variation, and a parallel sub-model in which the between-cluster deviance is another "baseline" deviance to be reduced by between-cluster sources of variation.

Now consider the partitioning of effects or coefficients for these two sub-models. To simplify the notation, assume that there is one variable of type X and another of type Z; within cluster j, the value of the Z variable is $Z_{.j}$, and the mean value of the X variable is $X_{\cdot j}$ (if there are more variables then more subscripts can be added). The between-cluster model will be

$$\mathrm{lgt}(Y_{\cdot j}/R_{\cdot j}) = c_0 + c_1 X_{\cdot j} + c_2 Z_j + f_j,$$

where f_j is a residual for cluster j. It should be understood that $Y_{\cdot j}$ is the quantity to be fitted. Within-cluster variation in X cannot possibly add to the explanation of between-cluster variation in the outcome, so only means and cluster-level variables are included as predictors in the between-cluster model. The coefficient c_0 is the constant term, c_1 measures the effect of a unit change in the mean of X, and c_2 measures the effect of a unit change in the predictor of type Z.

The generic within-cluster model is

$$\mathrm{lgt}(Y_{ij}/R_{ij}) = \mathrm{lgt}(Y_{\cdot j}/R_{\cdot j}) + b_0 + b_1(X_{ij}-X_{\cdot j}) + b_2 Z_j(X_{ij}-X_{\cdot j}) + g_{ij},$$

where Y_{ij}/R_{ij} is the ratio of use (0 or 1) to risk for person i in cluster j; g_{ij} is a residual for person i in cluster j. The term $\mathrm{lgt}(Y_{\cdot j}/R_{\cdot j})$ is an "offset" which is treated as part of the constant on the right hand side of the equation.

In order to estimate the coefficients of the within-cluster model it is necessary to calculate the deviations of each X variable from its within-cluster mean, and to code this deviation onto the individual-level file. The Z variables will enter into this model as interactions with those deviations.

The ratios Y_{ij}/R_{ij} and $Y_{\cdot j}/R_{\cdot j}$ can range from 0 to 1. At the extremes, their logits are not defined. Defaults will either be supplied automatically by the computer package or must be specified by the user. For example, if the ratio is less than .0001 then it can be replaced by .0001, and if it is above .9999, then it can be replaced by .9999. As a rule, a change in the default will have little impact on the estimates.

If the two components are added, the combined model will be given by

$$\mathrm{lgt}(Y_{ij}/R_{ij}) = (b_0+c_0) + b_1(X_{ij}-X_{\cdot j}) + c_1 X_{\cdot j} + b_2 Z_j(X_{ij}-X_{\cdot j}) + c_2 Z_j + (f_j+g_{ij}).$$

Obtaining Results with GLIM

Although we are not aware of a better package than GLIM (the computer package "General Linear Interactive Modelling") for getting the results presented in this paper, the micro-computer version of GLIM has several limitations. These are the most relevant ones: (1) the maximum number of data points (variables x cases) is 25,000; (2) the maximum record length for data is 132 columns; (3) the maximum number of variables that can be included in a model is 31. When clusters are the units, the first limitation is not serious, but it becomes very serious if the whole file is to be used.

The following output from a GLIM run on the cluster-level file illustrates how the models are fitted with such data and also shows the equivalence of various alternatives. The listing shows that (1) it makes no difference whether a categorical predictor is coded as separate dummies or is converted to a categorical variable using the "factor" statement (it is much easier to do the latter); and (2) it makes no difference whether we attempt to fit x using binomial denominator n and no weight, or we attempt to fit p=x/n using binomial denominator 1 and weight n. It is crucial, however, that the number of women (n) be taken into account one way or the other.

Comments are inserted into the listing in upper case type. Prior to the listed statements, the following variables in the 116 clusters have been read:

n number of women in the cluster
x number of current users in the cluster
el proportion of women in the cluster who have electricity
sp1 presence code for service point 1 (hospitals)

1 not named
2 named, distance >30 km
3 named, distance <30 km

SP1 IS STATED TO BE CATEGORICAL

[i] $factor sp1 3$

ALTERNATIVE DUMMY VARIABLES ARE CONSTRUCTED

[i] $c replace sp1 with three dummies $
[i] $c sp11=1 if sp1=1, otherwise sp11=0$

[i] $c sp12=1 if sp1=2, otherwise sp12=0$

[i] $c sp13=1 if sp1=3, otherwise sp13=0$

[i] $calc sp11=%if(sp1==1,1,0)$

[i] $calc sp12=%if(sp1==2,1,0)$

[i] $calc sp13=%if(sp1==3,1,0)$

[i] $c fit count x with base n and no weight$

[i] $yvar x$

[i] $err b n$

N.B. THE LOGIT IS THE DEFAULT LINK FOR THE BINOMIAL SO THE STATEMENT "$link g$" (G FOR LOGIT) IS NOT NEEDED

```
[i]  $fit $dis e$
[o]  scaled deviance = 410.18 at cycle  4
[o]                  d.f. = 115
```

HERE WE HAVE THE BASELINE DEVIANCE AND DEGREES OF FREEDOM FOR THE NULL MODEL WITH NO COVARIATES AND CLUSTERS AS UNITS

```
[o]
[o]            estimate      s.e.       parameter
[o]      1      -2.405     0.06575         1
[o]      scale parameter taken as        1.000
```

THE OVERALL PROPORTION WILL BE EXP(-2.4050)/[1+EXP(-2.4050)]

```
[o]
[i]  $fit el $dis e$
[o]  scaled deviance = 301.94 at cycle  4
[o]                  d.f. = 114
[o]
[o]            estimate      s.e.       parameter
[o]      1      -3.156     0.1136          1
[o]      2       1.857     0.1809          EL
[o]      scale parameter taken as        1.000
```

THE FITTED LOGIT USING el WILL BE -3.156+1.857el. GET FITTED PROPORTION BY SUBSTITUTING el AND TRANSFORMING AS ABOVE.

```
[o]
[i]  $fit sp1 $dis e$
[o]  scaled deviance = 406.01 at cycle  4
[o]                 d.f. = 113
[o]
[o]          estimate     s.e.      parameter
[o]    1      -2.387     0.08586       1
[o]    2      -0.2781    0.1845      SP1(2)
[o]    3       0.1411    0.1570      SP1(3)
[o]    scale parameter taken as      1.000
```

THE FITTED LOGIT IN THE REFERENCE CATEGORY (SP1=1) IS -2.387. IN CATEGORY 2 (SP1=2) IT IS -2.387 - 0.2781. IN CATEGORY 3 IT WILL BE -2.387 + 0.1411. USING PROP=EXP(LOGIT)/[1+EXP(LOGIT)], THE CORRESPONDING PROPORTIONS ARE .0842, .0651, .0957, RESPECTIVELY.

```
[o]
[i]  $fit sp12+sp13 $dis e$
[o]  scaled deviance = 406.01 at cycle  4
[o]                 d.f. = 113
[o]
[o]          estimate     s.e.      parameter
[o]    1      -2.387     0.08586       1
[o]    2      -0.2781    0.1845      SP12
[o]    3       0.1411    0.1570      SP13
[o]    scale parameter taken as      1.000
[o]
```

IN THIS FORMAT THE EQUATION FOR THE FITTED LOGIT IS -2.387 + (-0.2781)SP12 + (0.1411)SP13, WHERE SP12 AND SP13 ARE DUMMIES DEFINED AS IN PROGRAM.

THE EQUIVALENCE VERIFIES THAT FACTOR STATEMENT CAN BE USED IN PLACE OF FORMING SETS OF DUMMY VARIABLES; NOTE HOWEVER THAT GLIM WILL NOT ACCEPT A VALUE OF 0 AS A CODE IN A CATEGORICAL VARIABLE AND WILL TAKE THE CATEGORY WITH CODE 1 AS THE REFERENCE CATEGORY.

```
[i]  $c revise model to use proportion x/n with base 1 and weight n$
[i]  $calc p=x/n$
[i]  $yvar p$
[w]  -- model changed
[i]  $calc dum=1.$
[i]  $err b dum$
[i]  $weight n$
[i]  $fit $dis e$
[o]  scaled deviance = 410.18 at cycle  4
[o]                d.f. = 115
[o]
[o]          estimate      s.e.       parameter
[o]     1     -2.405     0.06576        1
[o]     scale parameter taken as  1.000
[o]
[i]  $fit el $dis e$
[o]  scaled deviance = 301.94 at cycle  5
[o]                d.f. = 114
[o]
[o]          estimate      s.e.       parameter
[o]     1     -3.156     0.1136         1
[o]     2      1.857     0.1809         EL
[o]     scale parameter taken as  1.000
[o]
[i]  $fit sp1 $dis e$
[o]  scaled deviance = 406.01 at cycle  4
[o]                d.f. = 113
[o]
[o]          estimate      s.e.       parameter
[o]     1     -2.387     0.08588        1
[o]     2     -0.2781    0.1845         SP1(2)
[o]     3      0.1411    0.1570         SP1(3)
[o]     scale parameter taken as  1.000
[o]
[i]  $fit sp12+sp13 $dis e$
[o]  scaled deviance = 406.01 at cycle  4
[o]                d.f. = 113
```

```
[o]
[o]        estimate    s.e.      parameter
[o]   1    -2.387     0.08588    1
[o]   2    -0.2781    0.1845     SP12
[o]   3     0.1411    0.1570     SP13
[o]   scale parameter taken as  1.000
```

THE EQUIVALENCE WITH THE PRECEDING OUTPUT DEMONSTRATES THAT WE CAN TAKE P=X/N AS THE DEPENDENT VARIABLE, WITH BINOMIAL BASE 1 AND WEIGHT N, AND GET RESULTS EQUIVALENT TO FITTING X WITH BINOMIAL BASE N AND NO WEIGHT (I.E. WEIGHT 1).

We can confirm the fitted proportions for the categorical variable SP1 by simply adding up x and n for the clusters in each category of SP1 and taking the ratio. The numbers of users (sum of x) in the three categories are 148, 40, and 64, respectively; the numbers of women (sum of n) are 1759, 615, and 669, respectively. The proportions .0841, .0650, and .0957 are the same as derived from the GLIM output, except for rounding error.

The Guatemalan survey was self-weighting; all the weights were 1.0. The procedures should be modified if sampling weights are needed; otherwise, there will not be the correspondence observed in the preceding paragraph between the logit regression coefficients and the tabulations. Let w be a case's weight, normalized so that the total weight is equal to the sample size. Then on the cluster record, n should be the total weight within the cluster (not the total number of women) and x should be the sum of the weights for the current users (not the number of current users). Any runs on individual women should use a weight option with weight w.

Some computer packages that might be considered as alternatives to GLIM, for example LIMDEP, do not allow the use of weights. Some statisticians will argue that sample weights should not be used, but we prefer to use them so that, as just mentioned, the logit regressions will correspond with tabulations and other simple analyses in which weights are routinely used.

All results presented in this paper were calculated either with GLIM; with STATA, a simple micro-computer package; or with special FORTRAN programs written by the author, mainly for file construction. Because of the limitation on the number of cases that can be processed by GLIM, some data files were compressed by cumulating those women who have the same combination of covariates. The deviances in Table 4 were obtained by a special FORTRAN program to calculate baseline deviances for a multinomial outcome.