Statistical Interactions in Studies of Physician Utilization: Promise and Pitfalls
Author(s): David L. Ronis and Kimberley A. Harrison
Source: *Medical Care*, Vol. 26, No. 4 (Apr., 1988), pp. 361-372
Published by: Lippincott Williams & Wilkins
Stable URL: http://www.jstor.org/stable/3765319
Accessed: 05/02/2009 13:44

MEDICAL CARE
April 1988, Vol. 26, No. 4

Statistical Interactions in Studies of Physician Utilization

Promise and Pitfalls

DAVID L. RONIS, PHD,* AND KIMBERLEY A. HARRISON, MHSA†

It has been suggested that use of interactive statistical models would greatly increase the proportion of variance accounted for by studies of physician utilization. The purpose of this paper is to evaluate and describe the benefits and pitfalls of using interactive statistical models of physician utilization. The paper presents Monte Carlo simulation data and real world data to determine how much more of the variance in physician utilization can be accounted for by interactive regression models. Results indicate that adding interaction terms is unlikely to produce large increases in variance accounted for. The usefulness of interactive models is particularly low when there is substantial measurement error in the predictor variables. Other advantages and disadvantages of interactive models are discussed, including 1) improved understanding, 2) inflation of alpha, 3) sensitivity to transformations and scale of measurement, and 4) confounding of interaction effects with nonlinear effects. Key words: health care use; physician utilization; access; multivariate statistics; linear regression. (Med Care 1988; 26:361–372)

Mechanic[1] reviewed studies of the decision to consult a physician and concluded that studies using large samples and powerful statistical techniques found trivial psychosocial and organizational effects. Mechanic suggested that this was due to the conceptual approaches, measures, and techniques of data aggregation and analysis used. Rundall[2] extended Mechanic's discussion by arguing that the data analyses have not appropriately represented the theoretical model. Specifically, the most frequently used theoretical model is interactive while data analyses have used additive models.

Rundall's comment[2] focused on Ronald Andersen's behavioral model[3] of health services utilization. According to this model, physician utilization is a function of three variables: 1) *need* i.e., illness; 2) enabling factors, i.e., the *ability* to obtain services; and 3) a *predisposition* to use services. Physician utilization is predicted only if all three factors are present. This implies a three-way interaction among the predictors. Since the effects of the three predictor variables are not additive, a nonadditive statistical model is required to fully represent and test the theory. This can be accomplished by including product terms as predictors in the

* From the University of Michigan Institute for Social Research, Ann Arbor, Michigan.

† From Blue Cross/Blue Shield of Michigan, Detroit, Michigan.

Preparation of this paper was supported by the University of Michigan Department of Psychology and Institute for Social Research, by Grant CA31520 from the National Cancer Institute and by NIMH Grant MH 16892.

Address correspondence to: David L. Ronis, Michigan Health Care Education and Research Foundation, 600 Lafayette East, Mail Code: E163, Detroit, MI 48226.

multiple regression. Presence of an interaction is indicated when the product terms account for a significant amount of variance above and beyond that accounted for by the original predictors.‡ Interaction terms may be added to regression models whenever researchers want to find out whether the effect of one independent variable on a dependent variable depends on the level of another independent variable.[4,5]

Rundall[2] suggested that researchers' failure to use interactive statistical models "could account for the low predictive power found with empirical tests of the behavioral model of physician utilization" (p. 104). The current paper presents Monte Carlo and survey results to determine how much more variance can be accounted for by including interaction terms in the statistical models. In addition, the paper discusses other advantages and disadvantages of interactive statistical models. Other ways to improve prediction and understanding of physician utilization are also mentioned.

Although the paper focuses on models of physician utilization, the results and discussion are applicable to interactive regression models of other dependent variables. Such models have been used, for example, by researchers studying the joint effects of stress and social support on health.[6,7]

Monte Carlo Study

Monte Carlo studies use random number generators and computation to study the properties of statistical methods. Monte Carlo research requires three steps: 1) specifying a model (or models) for generating the data, 2) sampling data generated by the model(s), and 3) analyzing the sampled data. Thus, Monte Carlo studies indicate what results would be obtained by various statistical analyses if the data were produced by various *known* processes. They are particularly useful for comparing the performance of different statistical methods and for studying the effects of violations of assumptions on statistical analyses. These questions cannot be answered by analyzing real data because we do not know precisely the causal mechanisms that generated any set of real data. Sometimes it is possible to derive an answer to such questions. We were not, however, able to derive answers to the main questions addressed in this paper, and suspect that no such derivation is possible in this case.

The general approach was to generate realistic data sets in which utilization (the dependent variable) was a highly interactive function of need, ability, and predisposition (the predictor variables). Then the data were analyzed with an additive statistical model and with an interactive statistical model. The variance accounted for in the two analyses were compared to determine how much the predictive accuracy could be improved by adding interaction terms.

Methods

The study examined the proportion of variance accounted for by additive versus interactive statistical models as a function of three other factors: 1) the pattern of *correlation* among true scores on the predictor variables, 2) the *function* relating scores on utilization with true scores on the predictor variables, and 3) the amount of *error* of measurement of the predictor variables. (The theory of measurement error[8] postulates that every observed score on any measure includes two components: a "true score" and a random measurement error. The true score, by definition, includes no

‡ In testing interactive models, it is essential that the models be hierarchical: that is, a product term should never be included in a regression model unless all of its constituents are also included in the model.[4] For a two-variable product to be included in the model, the two separate variables must also be included. For a three-variable product to be included, the three separate variables and their three two-way products must be included. As a result, stepwise regression analysis cannot be used with interactive models, unless the procedure is constrained to follow these hierarchical rules.

measurement error and is the average value that would be obtained from an infinite number of repeated measurements.) The three factors (correlation, function, and error) were varied to ensure representation of a wide variety of possible "real worlds" and to assess the impact of these variations on the usefulness of interaction terms.

Correlation. Three different patterns of correlation among the true scores on the predictors were considered plausible enough for inclusion: 1) all three predictors (need, ability, and predisposition) positively correlated with each other, 2) all three variables uncorrelated, and 3) ability and predisposition positively correlated with each other, but negatively correlated with need. A high magnitude of correlation (0.6) was used.

Functions. Five different functions were used to generate true scores on the dependent variable (utilization) from true scores on the predictors. The equations for these functions are listed in Table 1. In Function 1, utilization equals the product of the predictors. If any predictor equals zero, utilization will equal zero. This is consistent with a literal interpretation of Andersen and Newman.[3] Functions 2 and 3 are variations on Function 1 designed so that need is more important than the other two predictors—as is typically the case.[1] Function 2 gives need extra weight by adding need to the three-way product term. Function 3 gives need extra weight by adding two terms (each including need) to the three-way product term. Functions 4 and 5 include two-way interactions but no three-way interaction. In Function 4, utilization equals the product of need and ability. In Function 5, utilization equals the sum of all three two-way products of predictors.

Three features are shared by all five functions. First, all are highly interactive. Second, they do not include any nonlinear components (curvature). Third, they include no disordinal (or crossover) interactions. Disordinal interactions are those in which the direction (sign) of the effect of one predictor variable depends on the level of another predictor variable. Need (illness) and ability (enabling factors) would have a disordinal interaction if, for example, need increased utilization among people with high ability to obtain care but decreased utilization among people with low ability to obtain care. (In some of the literature on regression analysis, disordinal interactions are referred to as qualitative interactions, while ordinal interactions are referred to as quantitative interactions.) In research on utilization of health care, disordinal interactions do not seem very likely.

Conclusions of the study cannot be safely generalized to functions having different characteristics, e.g., nonlinear effects or disordinal interactions. On the other hand, the conclusions of the Monte Carlo study are not restricted to studies of physician utilization. They are equally applicable to any substantive domain that can be described by equations of the forms listed in Table 1. We believe that interactions in most areas of health services research are ordinal in form

TABLE 1. The Five Functions Generating Scores on Utilization from True Scores on the Predictors[a]

Function	Equation
Function 1	U = NAP + E
Function 2	U = NAP + 0.25N + E
Function 3	U = NAP + 0.5NA + 0.5NP + E
Function 4	U = NA + E
Function 5	U = NA + NP + AP + E

[a] In these equations, U represents the score on utilization, N represents the true score on need, A represents the true score on ability to obtain care, P represents the true score on predisposition, and E represents random measurement error in utilization. True scores on the predictors (N, A, and P) were uniformly distributed in the range from zero to one. The random measurement error (E) was normally distributed with a variance selected so that 10% of the variance in the dependent measure (utilization) would be due to measurement error. The addition of measurement errors produced some negative values. Since negative scores for utilization do not make sense, each negative score was replaced by a randomly selected score between zero and the true score.

and are similar to functions included in the study. In many cases, the theoretical and analytical models are less interactive than the equations in Table 1.[9] In those cases, the proportionate increase in variance accounted for would be less. However, the findings concerning the effects of measurement error (see below) would apply to those domains as well.

Error. Multiple regression analyses unrealistically assume that the predictors are measured without error. In order to examine the impact of this factor, four different levels of error were added to the predictor variables. The measurement errors were drawn from normal distributions with means of zero and variances selected so that the reliabilities of the predictors (defined as the proportion of true variance to total variance in the measure) were 1.00, 0.90, 0.70, and 0.50. Studies of the reliability and validity of survey measures suggest that the lower two levels are more common.[10,11] The addition of errors created some negative scores. Since negative scores are not normally observed in this area of research, each negative score was replaced by a randomly selected score between zero and the true score. The reliability of the dependent measure (utilization) was fixed at 0.90. Although this value is higher than typical in health services research, its exact value is much less critical than that of the reliability of the predictor variables, since measurement error in the dependent variable does not violate an assumption of the statistical model.

Statistical model. The generated data were analyzed using two statistical models: the *full* interactive linear regression model and the *additive* linear regression model. The full model is represented as:

$$(1)\quad U = \beta_0 + \beta_1 N + \beta_2 A + \beta_3 P + \beta_4 NA + \beta_5 NP + \beta_6 AP + \beta_7 NAP + \epsilon,$$

where U, N, A, and P are the *measured* levels of utilization, need, ability, and predisposition, respectively; β_0 through β_7 are the regression coefficients estimated from the data; and ϵ is the error in predicting utilization from the predictor variables. The additive model is represented as:

$$(2)\quad U = \beta_0 + \beta_1 N + \beta_2 A + \beta_3 P + \epsilon.$$

The four factors in the study were crossed with each other producing a 3 (correlation) $\times$ 5 (function) $\times$ 4 (error) $\times$ 2 (statistical model) factorial design. Fifteen samples of 300 cases were generated and analyzed for each cell in the design.

Major Results and Interpretation

This section presents the major findings from the Monte Carlo study. A complete description of the results and further methodological details are presented in Ronis.[12] The effects of statistical model and measurement error can be seen in Figure 1, which shows values of R^2 averaged over functions, patterns of correlation, and samples.

The full model accounted for more of the variance ($R^2 = 0.694$) than the additive model ($R^2 = 0.564$), a difference or *advantage* of 0.13. This advantage was substantially more than the increase of 0.006 that would be expected *by chance* when four additional predictor variables are added to three predictors that already account for 56% of the variance with a sample size of 300.[13,14] Ratios between the R^2 values for the full model and those for the additive model were 1.39, 1.22, 1.27, 1.23, and 1.09, respectively, for Functions 1–5. Thus, the full model typically accounted for about one and a quarter times as much variance as the additive model. Of course, if the true function relating the criterion variable (utilization) to its predictors is less interactive than the functions included in the simulation, the full model would produce a smaller increase in R^2.

Measurement error had two important effects. The addition of measurement error

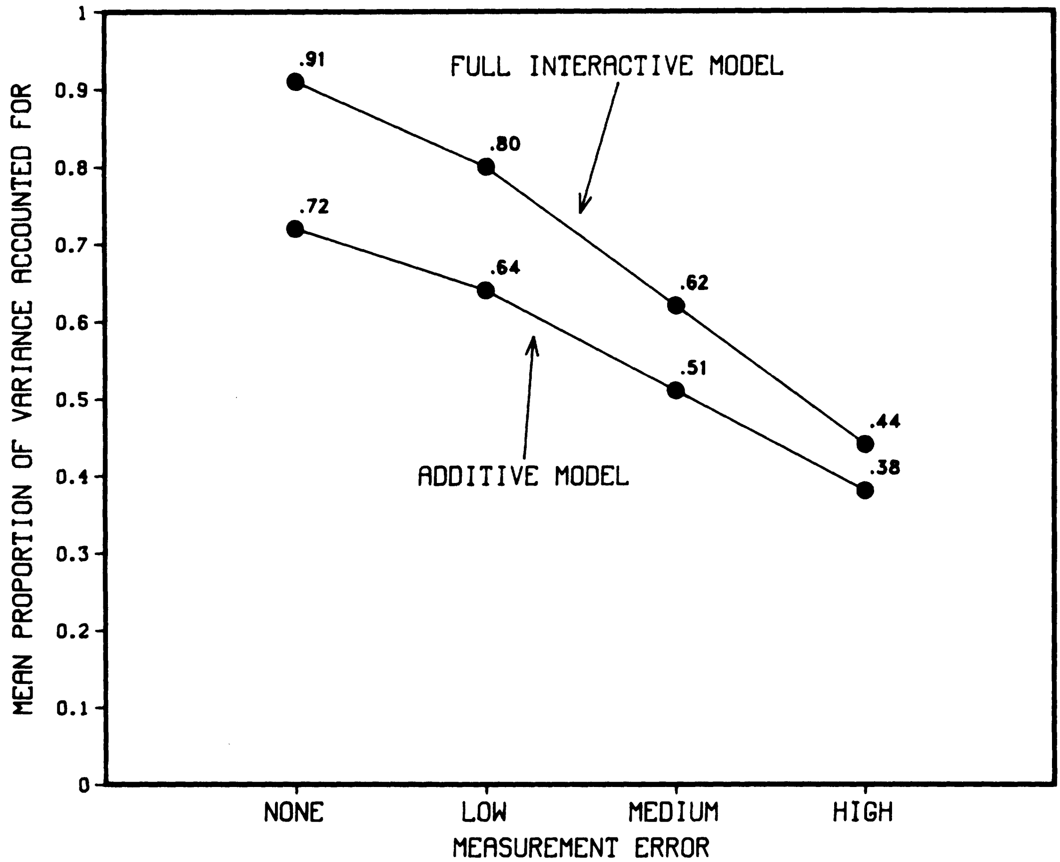

FIG. 1. Mean proportion of variance accounted for (R^2) as a function of measurement error for two statistical models (full and additive).

steadily decreased the proportion of variance accounted for, from 0.817 for error-free predictors to 0.413 for predictors with reliabilities of 0.5. In addition, measurement error reduced the fit of the full model more than it reduced the fit of the additive model. Thus, error decreased the advantage of the full model—from 0.188 for error-free predictors to 0.063 for predictors with reliabilities of 0.5.§

Interactive statistical models are particularly susceptible to the effects of measurement error because their predictions depend on the *joint* level of the predictors. A large measurement error on *one* predictor can lead to a very inaccurate prediction. For example, an inaccurately recorded level of zero on any predictor may lead to a prediction of low utilization, no matter how high the levels of the other predictors. An additive model would be less affected by this error.

§ Some supplementary computer runs were conducted using a symmetric unimodal distribution—beta (4,4) and skewed distributions—beta (1,2) and beta (2,1) instead of uniform distributions for the true scores of the predictors. These runs revealed the same main effects and interaction of measurement error and statistical model. Compared with the results for the uniform distribution, interaction terms were less useful for the symmetric unimodal distribution, more useful for the positively skewed distribution, and less useful for the negatively skewed distribution.

Discussion

On average, for the five highly interactive functions studied, the interactive statistical model accounted for about one and a quarter times as much variance as the additive model. So, if a study using the additive model accounted for 20% of the variance (as was typical of past studies[1]), the full model might account for about 25% of the variance. While this is a useful improvement, it still leaves 75% of the variance unexplained. This suggests the use of additive models is not a major cause of the low predictive accuracy of large scale studies. Researchers should consider using interactive models, but they cannot expect too much help from this improvement alone.

In the simulation, predictive accuracy was decreased by measurement error in the predictors. Measurement error decreased the fit of the interactive model more than it decreased the fit of the additive model. An implication of this finding is that investigators using interactive models should be particularly concerned about the reliability and validity of their measures. The development and refinement of multiple-item scales should be very useful in studies designed to investigate interactive models.

The Monte Carlo study allowed us to estimate the potential usefulness of interaction terms (when the data are known to be highly interactive) and provided insight as to when and why interaction terms are more or less useful. Monte Carlo studies cannot, however, indicate how useful such interaction terms will be in practice when the underlying function and level of error is unknown. For that information we must turn to real-world data—in this case, to survey data on physician utilization. Comparison of the Monte Carlo and real-world data can also provide information that could not be derived from either type of data alone.

Real-World Studies

Past Survey Research

We know of only two published studies of physician utilization that have used interactive statistical models. The first was reported by Sharp, Ross, and Cockerham in 1983[15]; the second was reported by Arling in 1985.[9] (A third study concerning use of mental health services is also somewhat relevant.[16] It is not discussed here, however, because it is not possible to determine the contribution of the interaction terms from the published report.) The study by Sharp, Ross, and Cockerham[15] examined the interaction between symptoms experienced (a need or illness factor) and perceptions of the significance of symptoms (a predisposing factor). Specifically, Sharp et al.[15] asked 618 adult residents of Illinois to indicate which of 11 symptoms they had experienced during the past year and which of the 11 symptoms were serious enough to prompt a trip to the doctor. Three indices were derived from these questions: 1) *Symptoms*—the number of symptoms experienced; 2) *Attitude Toward Symptoms*—the number of symptoms judged serious enough for medical attention; and 3) *Salient symptoms*—the number of symptoms that were both experienced and judged serious. The salient symptoms variable is an interaction term or, more precisely, the sum of eleven theoretically meaningful interaction terms: one symptom $\times$ attitude interaction term for each symptom.

The additive linear regression model predicting physician utilization from symptoms and attitude accounted for 14.0% of the variance. The interactive model including all three predictors accounted for 15.5% of the variance in physician utilization, or one and one ninth as much variance as the additive model.[‖] This difference is statistically significant and many times as great as the increase of 0.002 that would be expected *by chance* when a third predictor is added. Although the finding contributes to our understanding of physician utilization, the increase in variance accounted for is not very substantial.

‖ These results were derived from the correlation matrix presented by Sharp et al.[13]

Arling[9] examined physician visits as a function of 13 need, ability (enabling), and predisposing variables and eight of their two-way interactions in a sample of 2,051 Virginians over 60 years of age. Three of these interactions were statistically significant, increasing R^2 from 17.9% to 18.7%. Although this increase was about twice as large as the increase of 0.004 that would be expected by chance even if eight predictors were added, the interactive model accounted for only about one and one twentieth as much variance as the additive model. These two studies indicate that inclusion of interaction terms suggested by theoretical analysis can produce a modest increase in variance accounted for. They do not indicate how much more of the variance can be accounted for by models including the full set of interactions suggested by the behavioral model of physician utilization.

New Survey Research

To further test the usefulness of interaction terms in predicting physician utilization, measures of the relevant variables were included in a recent survey conducted by the authors. The data were analyzed using additive and interactive statistical models.

Methods. Respondents in the survey were a probability sample of 619 Detroit area women. Data were collected in personal interviews. Physician utilization was assessed by questions about surgeries and hospitalizations, office visits for treatment of medical problems, and office visits for checkups. Need or illness was assessed by a general rating, questions about how many times the respondent stayed in bed or cut down on normal activities because of health problems, and the occurrence of other health problems. Ability to obtain health care was assessed by questions about family income, health insurance coverage, travel time and waiting time to obtain health care, busyness—employment and children, and a general rating of ease or difficulty of obtaining medical care. Indices of utilization, need, and ability were created by summing the scores on the relevant items.¶ Items with strong positive skews were log transformed before summing. Several potential predisposing factors were also examined: education, age, race (dichotomized: black vs. other), and marital status (dichotomized: married vs. other).

Results. Complete data for regression analysis were obtained from 492 respondents. Four separate regressions were computed to predict utilization from need, ability, and each of the four predisposing variables. Four additional regressions were conducted including the predictors mentioned above and all two- and three-way products of the three predictors in each analysis. Adding the interaction terms (products) produced a significant increase in R^2 for the model using marital status as a predisposing factor but not for the other analyses. In the significant case, the additive model accounted for 23.6% of the variance, while the interactive model accounted for 25.3% of the variance. Though the difference between these two R^2 figures was several times as great as the difference of 0.006 expected by chance with the addition of four predictors, the ratio between these figures was only 1.07.

Discussion

It is notable that the improvement in prediction of utilization in the three real data sets was substantially less than in the Monte Carlo study. Two factors may explain this difference. First, the true function relating utilization to need, ability, and predisposition may be less interactive than the func-

¶ Items composing the utilization index were weighted according to the relative degree of demand each placed on health care services. Office visits for checkups were given a weight of one. Office visits for treatment of medical problems were given a weight of two. Surgeries and hospitalizations were given a weight of four. Items composing the need and ability indices were summed without weighting. Items that were on different scales were transformed to z-scores (standardized) before summing so they would receive equal weight.

tions included in the simulation. This seems to be the case in the Arling study where several variables had significant effects on utilization without being involved in significant interactions. Second, the measures in the three studies may be less reliable or valid or both than the average level in the simulation. Both of these factors probably reduced the contribution of the interaction terms in the survey data.

General Discussion

Advantages and Disadvantages of Interaction Terms

The analyses described above indicate that interaction terms can produce a statistically significant improvement in predictive accuracy. Another potential advantage of including interaction terms is improved understanding of the causes of physician utilization. The studies by Sharp et al.[15] and by Arling[9] provide excellent illustrations of this advantage. By combining interaction terms with path analysis, Sharp et al. were able to provide evidence that symptoms were not the direct causes of utilization. Rather, symptoms combined interactively with specific attitudes to produce *salient symptoms.* Salient symptoms (a partially physiologic and partially psychosocial variable) mediated the effects of symptoms and attitudes on utilization behavior. Evidence for this process could not have been obtained without using interaction terms.

Inclusion of interaction terms also increased understanding in the Arling[9] study. For example, *impairments* in activities of daily living interacted with medical conditions and with social support. These interactions indicated that different variables influenced physician utilization among persons with and without such impairments. Medical conditions significantly increased utilization among persons with zero to two impairments but not among persons with three or more impairments. On the other hand, social support significantly increased utilization among persons with impairments, but not among persons without impairments. These findings make sense and increase our understanding of physician utilization. They could not have been obtained without examining interactions among the predictors. See LaRocco et al.[7] for an example of the theoretical utility of interactive models in another substantive domain.

There are, however, disadvantages of using interaction terms that should be considered. These disadvantages are not restricted to studies of physician utilization but apply to use of interaction terms in regression analyses generally.

Increasing the Number of Predictors and Tests. Interaction terms greatly increase the number of significance tests that can be performed. For example, a researcher using just six dichotomous or continuous predictor variables can test 15 two-way interactions, 20 three-way interactions, and 22 higher-order interactions. With ten such predictors, there are 45 two-way interactions, 120 three-way interactions, and 848 higher-order interactions. Even if testing is limited to two-way interactions, increasing the number of tests performed increases the chance that one or more effects will be found to be statistically significant *by chance alone.* In other words, alpha (the type I error rate) will be inflated and researchers will be reporting more significant effects that cannot be replicated. Another problem with including a large number of interaction terms in the regression model is the loss of degrees of freedom from the residual or error term. The loss of degrees of freedom reduces the sensitivity of the significance tests and the precision of the parameter estimates.[4,17] These problems can be reduced by testing only a small number of theoretically interesting interactions (see Sharp et al.[15]). When many interactions are tested, the alpha level for each test should be reduced.

Multicollinearity: High Correlations Among Predictors. Product terms are often highly correlated with one or more of their components. This collinearity reduces

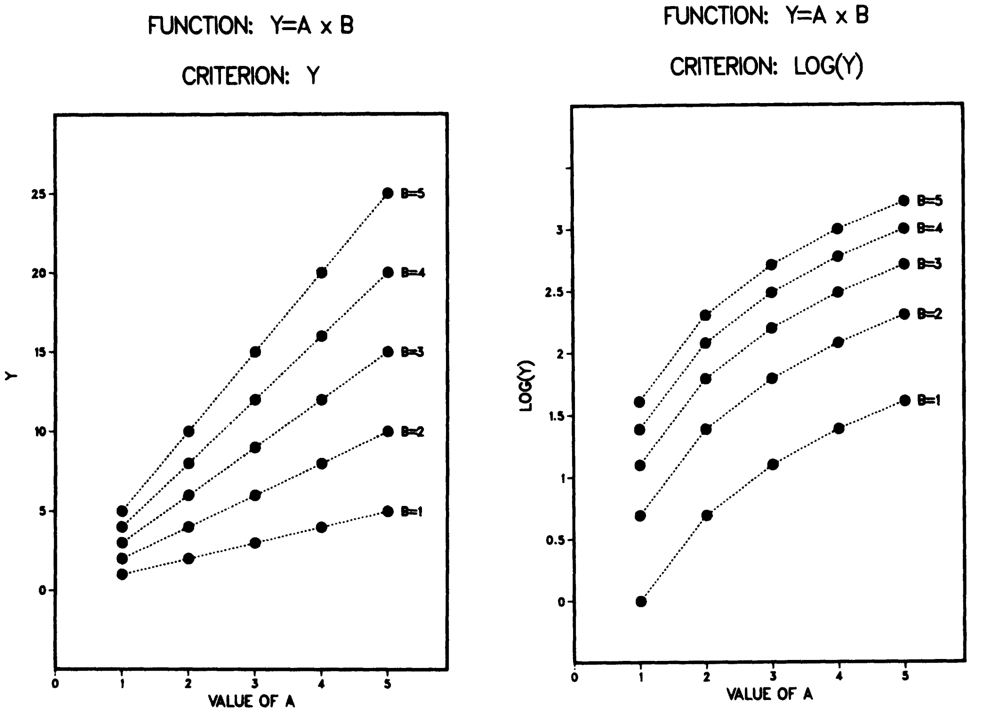

FIG. 2. Transformations can eliminate interactions: Y (left panel) and log(Y) (right panel) as functions of A and B. In both panels, the data were generated by Y = A × B.

the sensitivity of the significance tests and the precision of the parameter estimates. (It may also cause rounding errors or other computation problems especially for poorly written programs.[18]) Severe multicollinearity is less likely if only a few interaction terms are included in the model. Some techniques have been developed to reduce the effects of multicollinearity, including ridge regression and principal components regression. Although these techniques are substantially more complex than ordinary least squares regression, they are useful in some circumstances.[18,19]

Sensitivity to Transformations and Scale of Measurement. A third problem with using interaction terms is that the presence and form of ordinal (qualitative/noncrossover) interactions generally depend on the scale of measurement. A log transformation of the dependent measure, for example, will often reduce or eliminate an interaction. Main effects (i.e., nonproduct terms) are generally less affected by transformations.[20,21]

The reason that log transformations can eliminate interactions can be traced back to algebra. If the criterion variable Y (utilization or anything else) was equal to $A \times B$, and the predictors were A and B, the effect of A on Y would depend on the level of B. That is, A and B would interact and including a product term ($A \times B$) in the regression model would improve predictive accuracy. If a log transformation were performed on Y, the criterion variable [log (Y)] would equal log (A) + log (B). Log (Y) would be an additive (but nonlinear) function of A and B. The effect of A on log (Y) would not depend on the level of B. That is, A and B would not interact. If A and B were independent of each other, including a product term would not improve predictive accuracy.

Figure 2 illustrates the algebraic results described above. In both panels of the figure, the criterion variable is plotted as a function of A and B where A and B take on the integer values from one to five. In the left panel, the criterion Y equals $A \times B$. The lines are nonparallel indicating an interaction. In the right panel, the criterion log (Y) equals log (A) + log (B). The "lines" are parallel, indicating there is no interaction between A and B. (An optical illusion may cause them to appear to be nonparallel, but measurement will prove otherwise.) If both the raw and transformed measures are considered to be meaningful, interpretation of an interaction that is present for only one scoring system may be quite difficult.

Confounding of Interactions with Nonlinear Effects. One little known pitfall of using interaction terms is that interaction effects can be confounded with nonlinear effects. Whenever two predictor variables are correlated with each other, their product is correlated with the squares of both predictors. An extreme example will make this clear. Imagine two predictors C and D that are perfectly correlated and have the same mean and variance: in other words $C = D$. In this case $C \times D = C^2 = D^2$. If C^2 makes a significant contribution to predicting some dependent variable Z (above and beyond the contributions of C and D), so would $C \times D$. This would occur, for example, if $Z = C^2 + D$.

If C and D are imperfectly correlated, the confounding of C^2 with D^2 and $C \times D$ would be weaker, but still present. So when predictors are correlated, their interaction term would be correlated with their squares. A significant interaction may appear when the relationship is really additive but nonlinear. Similarly, a significant nonlinear effect may be found when the true relationship is linear but interactive.

Perhaps the best way to reduce this confusion is to include both the product term and the squares in the regression. The "real" interactive or nonlinear effects should remain significant when the others are controlled. However, this analysis will be biased by measurement error in the predictors. Unequal reliability of the predictors may be particularly troublesome. Busemeyer and Jones[22] have argued that the scale of measurement problem combined with the problem of random measurement error are so severe that regression analysis is inadequate to test for noncrossover interactions.

Other Ways to Improve Prediction and Understanding. Although interaction terms can improve prediction of physician utilization and can facilitate understanding, it is clear that they cannot do the whole job. Most of the variance remains unexplained. So it is necessary to consider other approaches to improved prediction and understanding. Improved measures seem particularly likely to be useful.

Mechanic[1] has identified several ways to improve measures of two ability/enabling factors: insurance coverage and the physician-to-population ratio. Other possible improvements include developing and using multiple-item scales and indices, reporting separate analyses for different types of utilization,[23] and including additional predictor variables. Empirical work is needed to determine which modifications and additions are useful.

Limitations of the Investigation and Directions for Future Research

In this paper we have attempted to quantify the predictive utility of including interaction terms in regression models of physician utilization, to describe and explain the benefits and pitfalls of using interaction terms, and to identify some ways to avoid or reduce the problems. We have also mentioned some alternative ways to improve prediction and understanding of physician utilization. We have not, however, attempted to quantify the magnitude of the problems, for example, to determine how frequently measurement error and correlations among predictors lead to incorrect inferences. We also have not tested any statistical techniques for dealing with measurement errors in interactive models, either by adjusting the covariance matrix or by using LISREL[24] or other similar programs. Although they are beyond the scope of the current paper, these are very interesting topics for future research.

In the Monte Carlo study and in the analysis of the Detroit-area data, we have consistently compared the additive model with the full model, and have not tested models including fewer interaction terms. This was appropriate for our purpose of determining how much more of the variance in physician utilization *can potentially* be accounted for by interactive models. From statistical theory we know that models including a subset of the interaction terms can never account for more of the variance than is accounted for by the full model. Our emphasis on the full model may, however, have led readers to believe that we disapprove of interactive models having fewer than the maximum number of interaction terms. This is the opposite of the truth. We encourage researchers to test a small number of theoretically important interactions rather than going on fishing trips with full interactive models.

Conclusions

Both the Monte Carlo study and the survey data indicate that adding interaction terms to statistical models of physician utilization or other domains characterized by noncrossover interactions can provide a statistically significant increase in predictive accuracy. Use of interactive models can also facilitate understanding. Interaction terms are unlikely, however, to produce large increases in proportion of variance accounted for. And there are several disadvantages of interactive models: inflation of alpha, reduced degrees of freedom, multicollinearity, sensitivity to scale of measurement, and potential confounding of interactive effects with nonlinear effects.

Acknowledgments

We thank R. Andersen, F. Andrews, R. Angel, G. Arling, Y. Harel, M. Kaiser, J. P. Kirscht, D. Mechanic, T. Rundall, W. Satariano, and J. E. K. Smith for comments on earlier drafts of this paper.

References

1. Mechanic D. Correlates of physician utilization: why do major multivariate studies of physician utilization find trivial psychosocial and organizational effects? J Health Soc Behav 1979;20:387.

2. Rundall TG. A suggestion for improving the behavioral model of physician utilization. J Health Soc Behav 1981;22:103.

3. Andersen R, Newman JF. Societal and individual determinants of medical care utilization in the United States. Milbank Mem Fund Q 1973;51:95.

4. Cohen J, Cohen P. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates, 1975.

5. Neter J, Wasserman W. Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs. Homewood, IL: Richard D. Irwin, 1974.

6. Cleary PD, Kessler RC. The estimation and interpretation of modifier effects. J Health Soc Behav 1982;23:159.

7. LaRocco JM, House JS, French JRP Jr. Social support, occupational stress, and health. J Health Soc Behav 1980;21:202.

8. Nunnally J. Psychometric Theory. New York: McGraw-Hill, 1967.

9. Arling G. Interaction effects in a multivariate model of physician visits by older people. Med Care 1985;23:361.

10. Andrews FM. Construct validity and error components of survey measures: a structural modeling approach. Public Opinion Quarterly 1984;43:409.

11. Marquis KH, Marquis MS, Polich JM. Response bias and reliability in sensitive topic surveys. Journal of the American Statistical Association 1986;81:381.

12. Ronis DL. Statistical interactions in studies of physician utilization. In: Proceedings of the Social Statistic Section. Washington, DC: American Statistical Association, 1984.

13. Crocker DC. Some interpretations of the multiple correlation coefficient. American Statistician 1972;26:31.

14. Carpenter WT, Bartko JJ, Evans MSC. Multivariate statistics and meaning in the study of psychopathology. Schizophrenia Bulletin 1983;9:1.

15. Sharp K, Ross C, Cockerham WC. Symptoms, beliefs, and the use of physician services among the disadvantaged. J Health Soc Behav 1983;24:255.

16. Watts CA, Scheffler RM, Jewell, NP. Demand for outpatient mental health services in a heavily insured population: the case of the Blue Cross and Blue

Shield Association's Federal Employees Health Benefits Program. Health Services Research 1986;21:267.

17. Wonnacott TH, Wonnacott RJ. Regression: A Second Course in Statistics. New York: John Wiley & Sons, 1981.

18. Draper NR, Smith H. Applied Regression Analysis, 2nd ed. New York: John Wiley & Sons, 1981.

19. Morris JH, Sherman JD, Mansfield ER. Failures to detect moderating effects with ordinary least squares-moderated regression: some reasons and a remedy. Psychol Bull 1986;99:282.

20. Box GEP, Cox DR. An analysis of transformations. Journal of the Royal Statistical Society 1964;26:211.

21. Games PA. Curvilinear transformations of the dependent variable. Psychol Bull 1983;93:382.

22. Busemeyer JR, Jones LE. Analysis of multiplicative combination rules when the causal variables are measured with error. Psychol Bull 1983;93:549.

23. Wolinsky FD, Coe RM, Miller DK, et al. Health services utilization among the noninstitutionalized elderly. J Health Soc Behav 1983;24:325.

24. Jöreskog KG, Sörbom D. LISREL VI User's Guide. Mooresville, IN: Scientific Software, 1984.

CHSR Study Groups

The Committee on Health Services Research (CHSR) of the Medical Care Section of the American Public Health Association (APHA) will again sponsor study groups prior to the APHA meeting in Boston next fall. Study groups usually meet for two or more hours to discuss research issues. Several researchers in the subject area make brief presentations followed by opportunities for round table discussion. Study groups on particular subjects often continue to meet each year. The groups will meet on Saturday, November 12th and Sunday, November 13th, 1988 (definitive schedule will be announced later). Attached is a list of study groups and their leaders. If you are interested in attending one or more of these groups, please call the study group leader at the number indicated. Peter Goldschmidt, who is leading the group on access, also welcomes suggestions as to presenters and offers of help in organizing the group.

The Committee is planning the usual two sessions during the APHA program: one on current issues in methodology and the other on recent advances in health services research. The Committee will also hold its business meeting during the APHA meeting dates. For further information on the CHSR, please call the chairperson for 1987/88, Dr. R. Heather Palmer, (617) 732-1060, address: Institute for Health Research, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115.

Study Groups Sponsored by the Committee on Health Services Research of the Medical Care Section of the American Public Health Association

Group	Contact Person	Telephone
AIDS	Penelope Pine	301-597-1454
Homelessness	Rene Jahiel	212-340-6754 718-729-0808
Quality of Care	Heather Palmer	617-732-1060
Health Status & Quality of Life	Kathleen Lohr	202-334-2319
Health Care Management	Alan Cohen	609-452-8701
Technology Assessment	Stan Reiser	713-792-5140
Ambulatory Care Case Mix	Oliver Fein	212-305-6262
Access to Care: The Challenges of Cost Containment and Quality Assurance	Peter Goldschmidt	301-530-9593