WORLD FERTILITY SURVEY TECHNICAL BULLETINS

AUGUST 1978

NO. 3/TECH. 597

Standardization

T. W. PULLUM

INTERNATIONAL STATISTICAL INSTITUTE
Permanent Office · Director: E. Lunenberg
Prinses Beatrixlaan 428
Voorburg, The Hague,
Netherlands

WORLD FERTILITY SURVEY
Project Director:
Sir Maurice Kendall, Sc. D., F.B.A.
35-37 Grosvenor Gardens
London SW1W OBS, U.K.

The World Fertility Survey is an international research programme whose purpose is to assess the current state of human fertility throughout the world. This is being done principally through promoting and supporting nationally representative, internationally comparable, and scientifically designed and conducted sample surveys of fertility behaviour in as many countries as possible.

The WFS is being undertaken, with the collaboration of the United Nations, by the International Statistical Institute in cooperation with the International Union for the Scientific Study of Population. Financial support is provided principally by the United Nations Fund for Population Activities and the United States Agency for International Development.

This paper is one of a series of Technical Bulletins recommended by the WFS Technical Advisory Committee to supplement the document *Strategies for the Analysis of WFS Data* and which deal with specific methodological problems of analysis beyond the Country Report No. 1. Their circulation is restricted to people involved in the analysis of WFS data, to the WFS depositary libraries and to certain other libraries. Further information and a list of these libraries may be obtained by writing to the Information Office, International Statistical Institute, 428 Prinses Beatrixlaan, Voorburg, The Hague, Netherlands.

STANDARDIZATION

WFS/TECH. 597

By: T.W. PULLUM
Director
Center for Studies in Demography
and Ecology, DK-40
University of Washington
Seattle, WA 98195, U.S.A.

C O N T E N T S

ACKNOWLEDGEMENTS

Several individuals, both inside and outside the World Fertility Survey, have assisted in the preparation of this document.

Special acknowledgement should be made of the contributions of two persons: D.I. Pool made a strong case for the value of standardization, and without his initial impetus this bulletin might never have been prepared. R. Little raised the issue of interaction and made many other suggestions. Very helpful comments by R. Carleton, G. Kalton, and N. Ryder on an earlier draft are also acknowledged.

1. INTRODUCTION

This technical bulletin is oriented around two problems. The first, which is relatively specific, is a clarification of the impact of educational level upon cumulative fertility in Malaysia, drawn from Malaysia's Country Report No.1. The second, more general problem concerns the systematic interpretation of the other many-way tabulations recommended for Country Report No.1 in all countries participating in the World Fertility Survey. In a sense, the first problem is simply an illustration, a specific instance of the more general problem. But the impact of education on fertility may also be regarded as a fundamental and prototypical question in its own right.

We shall describe a procedure for interpreting tables in Country Report No.1, or similar tables, which does not require access to the complete data file and does not require access to a computer. It may be applied during the preparation of that report, or later in more specialized reports. The procedure is a form of direct standardization, which is widely known to demographers.

If a researcher has access to the complete data file and adequate computational facilities, it will be preferable to use a procedure such as path analysis (see WFS Technical Bulletin No.2, *Path Analysis and Model Building*). If one lacks access to the complete file but has the computational facilities to apply log linear models, then these too will be preferable (see WFS Technical Bulletin No.4, *Generalized Linear Models for Cross-Classified Data from the WFS*). Standardization is regarded as a supplementary or a clarifying tool for analysis by cross-tabulation when analytic resources are limited.

Many readers who have experience with standardization will find it described here in an unfamiliar way. We have made an effort to determine the technique's limitations, to draw parallels with other procedures, and to clarify the nature of the demographic assumptions. As a result, users of the technique as described herein may have considerable confidence in their interpretations. Despite the relatively deep treatment of what is often considered an elementary procedure, the development is non-mathematical and, it is hoped, will be understandable to most authors of reports on country surveys. The computational procedures are illustrated in detail, and can be easily done without a computer.

WFS can, on request, provide a FORTRAN computer program to perform these procedures. The program is intended for use by people without access to the original data and uses cell frequencies or means as input.

2. COMPOSITIONAL VARIABLES

Much of the analysis of large sample surveys is based upon aggregate measures. These are means, rates, or proportions, calculated for subgroups which share certain characteristics. For example, in a fertility survey one might calculate the mean parity of different education groups, and interpret differences in these means, from one group to another, to be a consequence of educational differences. Even if the author did not actively draw such an interpretation, readers of such means would be inclined to do so.

Aggregate measures reflect more, however, than the variable(s) defining the aggregation. To continue the above example, there may be differences between education groups in other variables associated with fertility, such as age at marriage, marital duration, ethnicity and so on. One can imagine a situation in which there would be no differences in mean parity if attention were restricted to subgroups having the same values on these other variables. Demographers have applied the adjective "compositional" to any variable which is not itself of interest in a particular context but which, through relation to a predictor of interest, can influence the values of the dependent variable.

When an aggregate measure is found to vary across subgroups, then that is a real and often important finding. It is the interpretation of this finding which requires elaboration. We need some way to compute the *net* effect of a variable, as well as the *total* effect. We shall attempt in this document to give meaning to these terms.

3. STANDARDIZATION: ITS USUAL FORM

Demographers have long used standardization to "remove" compositional variables. The most familiar application is to international comparisons of the crude death rate (CDR), the number of deaths in a year per thousand of population. The CDR is equivalent to a weighted average of the age-specific death rates, in which the weights are simply the proportions of the population in the several age groups. The age distribution is the compositional variable. Typically, in developing countries, which have experienced high rates of natural increase for at least a generation, these weights will be relatively small for the older ages, in which most deaths occur, and large for the ages of low mortality. In a developed country, by contrast, the population tends to be older and a relatively high proportion is in the ages of high mortality. Thus the CDR can be lower for the developing country, even if all its age-specific rates are individually higher than those of the developed country. The CDR, as a measure, is closely related to the age distribution, and some effort is required to remove this relationship.

In this elementary form, standardization involves 1) the choice of a standard, or reference population; and 2) the substitution of this population's composition for that of each population of interest. When several populations are then compared, differences will be net of the compositional variable, because all populations will have been assigned an identical distribution on that variable.

The principal shortcoming of the procedure is that the choice of the standard population is somewhat arbitrary. Most commonly, it is some country for which accurate data have long been available, such as England and Wales, or else one of the populations of interest. The standard is virtually never, in this context, taken to be the aggregate of all populations combined. There are at least two reasons for this: The first is that one might wish to add another country to the set of populations of interest, and any such addition would require recalculation of the standard distribution and then of all the standardized rates. Secondly, when comparing national entities, it is not at all clear whether they should be included in proportion to their population sizes, or all should be given equal weights.

It is essential to note that in this context, interpretations are always limited to differences, between countries, in the standardized quantities. The analyst pays no attention to differences between the pairs of standardized and unstandardized quantities; these reflect nothing more than the differences between the population of interest and the standard in their distributions on the compositional variable. These differences can be assessed in a more direct manner, if they are of interest. When the standard population is chosen arbitrarily, they are not of interest.

For examples and discussions of this form of standardization see Jaffe (1951), Barclay (1958), Kitagawa (1964), and Shryock and Siegel (1971).

4. STANDARDIZATION OF SUB-POPULATIONS

In the analysis of a single data set, such as a census or a cross-sectional survey, it is possible to apply a variant of standardization as described above. The population may be divided into several aggregates, such as ethnic groups, with the objective of comparing the values of a dependent variable in these sub-populations. The dependent variable, as before, is an aggregate measure such as mean parity.

Comparisons across sub-populations are hampered by statistical associations that may exist between the variable which defines the sub-populations and some other variable. For example, in comparing the mean parities of several educational categories, the conclusions will be more complex if education and marital duration are associated. Marital duration has a clear, largely biological relationship to parity and if, say, the higher education groups have disproportionately high numbers of women with short durations of marriage, then the high education groups will have low fertility for that reason alone. Marital duration is a compositional variable.

This is a situation in which standardization may be considered, although the choice of the standard population is not immediately clear. There are three reasonable possibilities. First, the distribution of the control variable, which would be applied within each category of the predictor variable, could be that of another country, as in the previous discussion. Second, the overall distribution of the control variable, within the country, could be applied. The standardized quantities would then be interpreted as the values that the aggregate measure would assume if each category of the predictors had the same distribution on the control variable as the overall population. Third, the standard could be one of the sub-populations, such as the highest education group.

The first of these possibilities might be considered for international comparisons of several different countries, although in any case there is the complication that some variables (such as ethnic group or region) are unique to each country. As with the third possibility, there is a serious problem of arbitrariness. It is not clear how one could select a reference or standard country for international work; nor is it clear how one would identify the sub-population within a country which could serve as a standard for internal comparisons.

The second choice, the application of the overall distribution on the compositional variable to each category of the predictor, will be recommended in this document. This choice is not wholly satisfactory, because this overall distribution is a weighted average of the distributions within each level of the predictor. These weights might be argued to be transitory and changing. For example, considering education as a predictor, the country might be experiencing a rapid rise in the level of education, and hence a rapid rise in the importance of the more highly educated groups in the overall composition. Nevertheless, the overall distribution is the least arbitrary distribution in that it applies when individuals are drawn randomly from the population, without our knowledge of their value on either the dependent variable or the predictor.

Standardization on the overall distribution of the compositional variable has a considerable history of use in demography and public health research. The use of this standard was promoted for purposes of multivariate interpretation by Rosenberg (1962). It appears not to have received critical evaluations from statisticians, except for Kalton (1968) and Bishop, Fienberg, and Holland (1975).

In motivating and illustrating the procedure, we shall employ notions of explanation and elaboration derived from Lazarsfeld (see Kendall and Lazarsfeld, 1950), as did Rosenberg (1962), but we shall lean even more heavily on the notions of total and net effects in the literature on path analysis (see Duncan, 1975, and Technical Bulletin No.2). The statistical terminology, in particular the reference to adjusted effects, is similar to that of multiple classification analysis, or MCA (see Andrews, Morgan, Sonquist, and Klem, 1973), which, like multiple regression, is a particular case of the general linear model. The major reason for putting standardization into this framework is that in practice it will be a less sophisticated alternative to these other techniques, and it is desirable to use a consistent terminology.

Also, in the interest of consistency, we shall replace the term *compositional variable* with the more widely used term *control variable*. The latter is more readily applied at all levels of measurement, and not just when the variables are categorical. It is commonly used in the social sciences outside of demography, despite its origins in the "controlled experiments" of other fields, and in practice simply refers to any variable which obscures a relationship of interest.

5. DEMOGRAPHIC CONSIDERATIONS

All tables recommended for Country Report No.1 include current age or years since first marriage as a demographic control and it is these variables, sometimes in combination with others, whose effect we propose to "remove" with standardization, in order to clarify relationships between variables of central interest. It is important, however, to understand the roles played by age and marital duration, both in so far as they are similar and in so far as they are different. We shall briefly summarize the nature of their impact on cumulative fertility, because that is the dependent variable of the main example, but rather similar comments would apply for most other dependent variables in Country Report No.1.

As a preface to these comments it must be made explicit that the measured values of age and marital duration, even if these can be obtained without any measurement error, are primarily indicators of social and cultural phenomena which influence an individual woman's fertility in a causal manner. They are themselves merely amounts of elapsed time: years since birth for the former and years since first marriage for the latter.

At the simplest level, of course, age and marital duration indicate the woman's accumulated biological potential for childbearing. That is, if we know either of these numbers then an effective upper limit can be placed on any estimate of her cumulative fertility or parity. Moreover, any individual woman's parity bears a monotonic relationship to her age and marital duration. These facts alone will lead to a strong statistical association.

Current age identifies when the woman was born. Women who were born at about the same time, who comprise a birth cohort, will have accumulated many shared cultural and socializing experiences. On the one hand, they will have been exposed to a background of some norms and behaviour patterns which will have changed very little. These will be in common to women of all birth cohorts, except that older women will have proceeded farther along their life cycle. On the other hand, there may have been changes in some areas, such as politics, economics and public health. There may have been general trends towards greater participation in the modern sector of the labour force, rural to urban migration, greater knowledge of family planning practices, and so on. These changes will have had a greater impact on younger women, because of their timing in relation to the woman's stage in her life. Thus, if efficient contraception is of only recent introduction, it cannot have affected the early fertility of older women. Age reflects both current and cumulative exposure to a mixture of both stable and changing social influences on fertility.

Marital duration, the other main WFS control variable, is obtained by subtracting Date of First Marriage from Date of Interview. Women who were married at about the same time comprise a marriage cohort.

If all women marry at about the same age, and there has been no trend in this age, then marital duration will sort the women in exactly the same way as age and will serve to indicate the same unmeasured characteristics as does age.

When age at marriage does vary considerably and/or there is a time trend, then there is a loss in the precision of marital duration as compared with age in indicating birth cohort phenomena. Suppose, for example, that there has been a range of ten years in age at first marriage - say, that most women married when between 15 and 25 but with a broad dispersion in that range. Then women who are matched in marital duration may well vary by a decade in age. The introduction of modern contraception and changes in attitudes toward abortion, for example, may have come at different points in their married lives.

Many groups will show a stable pattern of variation in fertility according to age at marriage, even within marriage cohorts. In Malaysia, for example, women who marry in their late teens are consistently more fertile than those who marry at a later age, even within marriage cohorts.

On the other hand, marital duration is nearly always an improvement over age as an indicator of the woman's potential fertility. Since most women reach menarche within a narrow range of age, we can use current age to obtain good estimates of how many years a woman has been fecund and can set an upper limit on her probable cumulative fertility. Marital duration permits a refinement: it gives an estimate of accumulated exposure to conception and childbearing, particularly if there is little premarital sexual activity and little marital separation or disruption.

There are life cycle phenomena as well that are better dated with reference to time of first marriage, than time of birth. For example, in some developing countries it is considered improper for a woman to continue childbearing after she has become a grandmother. Thus all her childbearing will be compressed into approximately the first twenty years of marriage. In developed countries, very little fertility occurs after just the first ten years of marriage. Generally, the extent to which life cycle effects may be dated better by the woman's age or by her marital duration depends on such things as the extent of fertility control and the strength of the association between marital status and childbearing.

Since age and fertility are related at the level of the individual, it follows that for a group of women, as in a WFS sample, the age distribution will be related to the fertility distribution. In particular, the age distribution of a group will be related to the group's mean parity. At this level, however, there is no analytic content to the relationship. The age distribution of a group of women aged 15 to 49, say, reflects nothing more than the annual numbers of births from 15 to 49 years ago, and the pattern of mortality during the past 49 years. These phenomena lead to the existence, at the date of the survey, of certain numbers of women of certain ages; and although each woman's age does relate to her fertility, the aggregate fertility of the group will be dominated essentially by historical accidents. The age distribution is irrelevant to an interpretation (although not to a pure description) of differential fertility.

The impact of marital duration at the aggregate level is somewhat different. The distribution of marital duration, which is a characteristic of the group, depends in a compound manner upon the age distribution and upon the historical distribution, within birth cohorts, of age at marriage. The first of these two distributions is of no analytic interest, as mentioned above. The second, however, is a typical policy-related variable. Age, in other words, is mainly a "nuisance" variable in analysis at the aggregate level; but age at marriage, like education, ethnic group, etc., is not. Marital duration is therefore a compound of one variable of no real interest and one of analytic value.

A brief comment is required on the use of parity as a measure of fertility. A woman's current parity has been reached in a series of steps or transitions. Each of these unit increases has been a consequence of a mixture of some circumstances which existed early in the woman's background and others which were contemporary with the transitions. Some of these circumstances were contextual, and others were specific to the woman. A thorough analysis of fertility would take into account the length of time associated with each transition (i.e., birth) which occurred, and the various statuses associated with the woman during each transition. Our use of current parity in the present bulletin does not imply any preference for this crude measure. It is due simply to the relatively accurate measurement possible with this variable and to its dominant role in the First Country Report. Subsequent technical bulletins will describe more sensitive measures of fertility, along with procedures which are more appropriate to them.

6. DEVELOPING A CAUSAL MODEL

Three types of variables are present in this discussion: the dependent variable; predictors (one or more); and controls (any number). The classification of a variable into one of these types is not intrinsic to the variable itself but is an expression of a model. The dependent variable is easily selected in general, although there are many variables which may play this role in one model and not in another. For example, a variable measuring the use of contraception may be dependent in one analysis but in another context may be considered to be a predictor of fertility.

After a dependent variable, such as parity, has been selected, the researcher identifies a set of variables which are related to it, in the sense that a change in the value of any one of these variables is believed to result in a change in the value of the dependent variable. These other variables may be differentiated into predictors or controls. Some variables may be classified in either manner; for example, in this document we shall usually treat level of education as a predictor, but sometimes as a control. Certain other variables are consistently treated by demographers as controls.

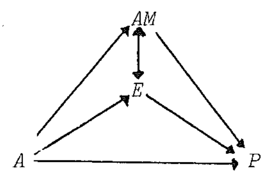
There are three possible ways of incorporating a control variable, such as age or marital duration. First, one may subdivide the population (or sample) into subgroups which are homogeneous with respect to the control, and analyze each subgroup separately. The second possibility is to redefine the dependent variable, for example, as a rate. Third, the control may be included explicitly as a covariate, in which case it is manipulated quantitatively in just the same manner as a predictor variable. We shall follow the third course in this document.

In our context, all women who are tied on each predictor and control are grouped together, in the cells of a cross-tabulation whose rows, columns, and panels consist of categories of the predictors and controls. The dependent variable will then be measured by a mean or proportion in each cell of the table.

Corresponding to each cross-tabulation is a model which, in more or less detail, specifies the nature of the relationship among the component variables. It is this model, in fact, which has led to the cross-tabulation. Whether or not it has been formally stated, one can say that a model of this sort is implicit in any type of analysis.

We shall use the term *causal model,* but *causal* should not be taken to presume the absence of reciprocal effects or of additional, unspecified variables. A "causal model" is simply an abstraction to facilitate statistical analysis. Regardless of the units of analysis, it may be conveniently represented by the same kind of diagram used for a path analysis, in which the unit is generally the individual woman, without aggregation (see Technical Bulletin No.2). We shall illustrate this representation for a basic combination of four variables which correspond to Table 2.2.6 of the *Guidelines for Country Report No. 1* (WFS Basic Documentation No.8).

Let P represent the mean parity of a group of women. The groups will be defined by all combinations of the remaining categories. A major topic of interest is the impact of level of education, E, on parity. As argued above, some measure of exposure to childbearing is also required as a control. For this purpose we shall bring in both components of marital duration. Let A and AM be defined to be current age and age at marriage, respectively. A diagram involving all four of these variables and allowing for all logically possible effects would appear as follows:



In this diagram, a single-headed arrow indicates an unambiguous direction of causation. Age is clearly prior to all the other variables, because as a number it simply identifies year of birth. Under the assumption that virtually all fertility is marital, age at marriage will also be prior to parity. (If this assumption is not appropriate, the easiest course would be simply to omit age at marriage altogether). Education is placed before parity under the assumption that young women will not have children while in school. A double-headed arrow is placed between AM and E; for most young women education is completed before marriage, but in some cases education may have been terminated simply so that marriage could occur, or continuing to go to school may have been simply an alternative to marriage. The ordering of the variables is derived largely from their temporal sequence for an individual woman, but in some other models there may be a more theoretical basis for the ordering.

Other variables, such as Region, could logically be included in this diagram. In fact, even when these other variables are omitted from the explicit representation, they will still play a role in any estimation. For example, the path from A to P actually represents a summation of a whole set of complex causal chains passing through any number of unmeasured variables. It would be simplistic to think that this "direct" path, or any other in the diagram, shows an immediate causal influence. Rather, each line is a shorthand representation from which the actual mechanics of the relationships are excluded, mainly because we lack adequate measurements.

7. ADAPTING THE MODEL TO AVAILABLE DATA

With every methodology one must make numerous compromises between the underlying objectives and the available data and tools. The substantive objective of the main example in this technical bulletin is a clarification of the impact of education upon cumulative fertility in Malaysia. In pursuing this objective a large number of compromises have been required, and it is important that they and the reasons behind them be made explicit, particularly if this document is to be a useful guide to similar analyses.

Every analyst will recognize the basic measurement problems, such as the use of reported, unverified responses. Hardly any variable can be taken at face value. For example, educational level is merely a proxy measurement of a set of characteristics which bear upon fertility. No one would argue that the hours spent sitting in the classroom, or the ingestion of what is formally taught there, have much to do with fertility. Rather, it is the assimilation of certain modern ideas, and the development of an identification with a socio-economic group, that will - or may - be reflected in later fertility. As mentioned earlier, parity itself is a deficient measure of fertility, and a variable so seemingly straightforward as age has its value as an indicator of more subtle characteristics.

In a WFS survey, as in any survey, we are provided with a limited amount of information. More to the point, since we are assuming that the file of individual women is not available, and the data come to us as cross-tabulations, we can only consider a few variables simultaneously. If there are more than three or four variables, or if there is more than minimal categorization of these variables, then cell frequencies will be too small.

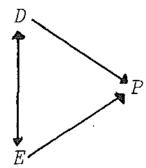
In Malaysia, we confront a particular obstacle to the use of the model described in the previous section. Malaysia is composed of three distinct ethnic groups, and to be complete, any analysis must incorporate ethnicity. If ethnicity were added to the four-variable model, we would have five variables and a large number of empty or nearly empty cells in the table. Therefore, in exchange for adding this variable a further compromise is required. We shall replace age and age at marriage by their difference, marital duration. Mean parity is given within all combinations of marital duration, education, and ethnicity in Table 2.2.7 of Country Report No.1.

A statistical control for marital duration is rather different, it must be noted, from a simultaneous control on age and age at marriage. On the average, a woman with a shorter duration of marriage will be younger. This is because most marriages occur in a narrow range of ages. It is also the case that a woman with a shorter marital duration will tend to have married later. In Malaysia, for example, a woman who has a marital duration five years less than another woman will (on the average) be about four years younger and will have married when about one year older than the other woman. However, there is a good deal of variability about this average. It is thus not possible to interpret variations according to marital duration as being a consequence of variations in a specific one of its component variables.

If a good deal of fertility occurs outside of marriage or outside of any recognized unions, then marital duration should not be used as a control variable, and the modification of the basic four-variable model described above would lie in the omission of age at marriage.

For the Malaysian data a check was made on the possibility that age and age at marriage jointly contain information (in the present context) going beyond that contained in marital duration. The four-variable model on page 12 was analyzed, using standardization, and was found to indicate almost precisely the same effects of education on parity as the following condensed model. Such a check should always be made, because in some countries the condensed model with marital duration would probably be an inadequate representation of a more complete model which used both age and age at marriage. In the case of Malaysia, we are justified in making a consistent use of this simplification both before and after the inclusion of ethnicity.

The role of ethnicity will be defined later; the initial model to which we shall apply the Malaysian data is summarized by the following three-variable diagram:



Here P represents mean parity within all combinations of education (E) and marital duration (D). The double-headed arrow between D and E indicates that effects operate in both directions: a) women with longer marital durations tend to be older, to have been educated at a time when educational opportunities were more limited, and therefore to have lower levels of education; and b) women with higher levels of education tend to have postponed marriage and for that reason have shorter marital durations. However, D and E are both clearly prior to P, so the other arrows are single-headed.

8. THE IMPORTANCE OF STATISTICAL INTERACTION

In some situations standardization is inappropriate, and it is always necessary to check for this possibility. The critical circumstance is that the effects of the predictor and control variables upon the dependent variable must be <u>additive</u>. The absence of this condition in the data, i.e., non-additivity or interaction, is an equally important obstacle to other techniques, such as path analysis and multiple classification analysis. Social scientists have not been sufficiently wary of non-additivity; it is noteworthy that, as Kalton (1968) has pointed out, the central illustration of standardization in Rosenberg's article (1962) was inappropriate for the technique. (Also see Atchley, 1969).

When the statistical model is some variant of the general linear model, standard tests can be applied, yielding clear guidelines as to the advisability of omitting interaction terms. Generally, however, we are considering contexts in which the technical capacity for these tests - as for these other methods themselves - is lacking. Moreover, in our applications the within-cell variability is not known. Therefore, we are limited to informal, approximate checks. The recommended checks should have the incidental feature of helping the analyst to understand the data better, whether or not they indicate that standardization is justified.

The following artificial data give an example of interaction. Let us suppose that education and duration have both been dichotomized into categories of equal size.

TITLE: Mean parity of women having specified levels of education and marital duration (base frequencies in parentheses).

		Education	
		Low	High
Marital duration	Low	2.0 (50)	2.0 (200)
	High	6.0 (200)	4.0 (50)

The base frequencies (given in parentheses) show clearly that women of high education tend to have low marital duration, and women of low education tend to have high marital duration. This sort of association between the predictor and the control is not the sort of interaction we worry about; in fact, it is the very circumstance that suggests the need for multivariate analysis. Rather, the interaction which must be checked for is of a higher order. It pertains to the joint impact of the predictor(s) and control(s) upon the dependent variable.

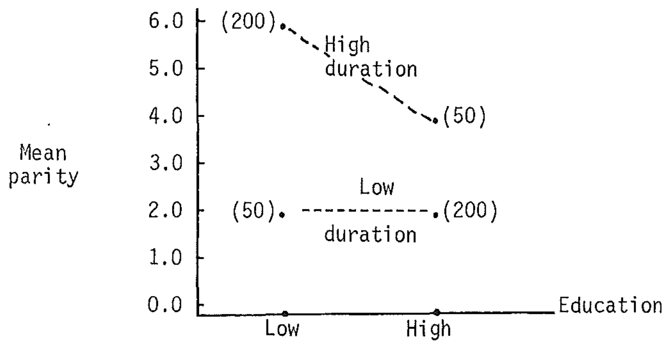
If we "hold duration constant" by restricting attention to the low duration group (top row of the table) we see that the women of high education have the same mean parity as those of low education (2.0 - 2.0 = 0.0). On the other hand, women of high duration (the bottom row of the table) have two children less, if they have high education (4.0 - 6.0 = -2.0). The effect of education is dramatically different within different levels of the control variable. This is the essence of interaction.

The implications of this sort of pattern are serious. Suppose that we ignored marital duration and computed the mean parities within the two education groups. In the low group, the mean would be (200·6 + 50·2) /250 = 5.2, and for the high education group, (50·4 + 200·2)/250 = 2.4. For reasons described earlier, the analyst might be motivated to standardize on marital duration, applying the overall distribution on this variable to each education group. The standardized mean parities would then be (250·6 + 250·2)/500 = 4.0 for the low education group and (250·4 + 250·2)/500 = 3.0 for the high education group.

The standardized mean parities by themselves imply that, controlling for marital duration, education has a negative impact on parity. But the table shows quite clearly that this impact is quite different within different levels of marital duration. Another name for this phenomenon is specification; the effect of the predictor varies according to the specified level of the control variable.

When this circumstance exists, the analyst is not justified in summarizing the data in a way which obscures or hides the interaction. On the contrary, some of the most useful and interesting findings of a survey are of this type, and they should be brought out.

One of the simplest ways to check for interaction (regardless of whether one hopes to find it or hopes not to find it) is graphical. The preceding table could be graphed as follows:



The two dotted lines refer to the two rows of the table. The frequencies associated with each point of the graph (i.e., with each cell of the table) are shown in parentheses. Less weight should be given to points based on lower frequencies. Non-additivity, or interaction, will be indicated graphically if the dotted lines are not parallel. Generally, when there are more than two categories in the predictor variable, interaction is implied when the dotted line segments do not have a constant vertical displacement.

In practice, the conclusions from a graphical check will be mixed. The line segments will rarely be exactly parallel; but if the departures are small, or if the case bases are so small that the points should not be taken too seriously (we suggest that cells with fewer than 20 cases be ignored), then the interaction may be considered negligible. It happens, more than occasionally, that the assumption of additivity is then acceptable except for a few cells of the table. The analyst may then proceed with the standardization, but in the discussion should explicitly mention and describe the interaction effects in those cells. We shall illustrate this practice in our examples. Approximate non-graphical checks are suggested in Appendix II.

It should be noted that non-additivity is probably to be expected with certain combinations of variables. For example, one might expect fertility differentials between sub-groups to increase steadily as marital duration increases. Certainly, if one had great detail in the measurement of duration one would expect a smaller differential soon after marriage than, say, thirty years later. The logarithm of parity suggests itself as a means of bypassing this problem; indeed, in some other Technical Bulletins we shall advocate the log transformation.

As will be seen later herein, however, with Malaysian data, interactions do not seem to result from our not taking logs. Moreover, there is a complication when logs are taken of cell entries which are arithmetic mean parities. (Arithmetic and geometric means differ from one another, to a degree which depends on the extent of aggregation.) Therefore, in the present context we shall not apply any transformations of means.

If the cell entry is a proportion P, and particularly if it varies much in the range 0 to 1, then one may wish to replace it by the logit, $\log P \,/\, (1\text{-}P)$, in order to reduce interaction.

9. APPLICATION: EDUCATION AND FERTILITY IN MALAYSIA

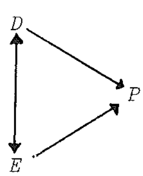
At last we are ready to apply standardization to clarify the impact of education upon cumulative fertility in Malaysia, using tables in the Country Report of the Malaysian Fertility and Family Survey, which was conducted in 1975. The variables may be referred to as P, D, and E, where

P = mean parity
D = marital duration (years since first marriage)
E = educational level,

with categories of D and E defined as follows:

CATEGORY		PER CENT OF SAMPLE
$D(1)$:	0-9 years	37.2
$D(2)$:	10-19 years	29.0
$D(3)$:	20 or more years	33.8
and $E(1)$:	No education	35.5
$E(2)$:	Religious education	0.8
$E(3)$:	Other non-formal education	2.9
$E(4)$:	Less than 7 years' formal education	48.7
$E(5)$:	7-12 years' formal education	11.0
$E(6)$:	More than 12 years' formal education	1.1

The educational categories are numbered as in the Country Report, but we note that for analytical purposes categories 2 and 3 should be treated separately, since they disrupt the ordinality of the other four categories. In practice one would probably combine categories 5 and 6, because the latter is so small, but we shall maintain the original classification. Table 1 presents the observed values of P within all combinations of D and E. (This table was actually obtained as a condensation of Table 2.) The bottom row of the table shows that mean parity decreases as education increases. However, the frequencies (in parentheses) in the columns show an association between D and E; higher educated women tend to have been married a shorter time than lower educated women. (Thus, the proportion having duration 0-9 years is .14, .46, .73 and .66 in education categories 1, 4, 5, and 6, respectively). It is our objective to "remove" the effect of the association in order to learn the effect of education *per se*. The model may again be diagrammed as follows:



The first step is to check for interaction between the control (D) and the predictor (E) in their effect on the dependent variable (P). Figure 1 shows graphically how mean parity varies with education; points which are connected have the same value on the control. Note that the education categories have been re-ordered in the figure, to aid interpretation of the graph. The same conclusion, with the same considerations, would be reached if any other ordering of the educational categories had been used.

The higher line segments in Figure 1 are generally steeper than the lower line segments. However, our evaluation of this graph is that standardization is appropriate for these data. This judgement is based on the following points:

a) The change in parity from no education for some formal education (1 to 4) is not the same for all levels of the control. However, there is considerable vertical separation, implying that the interaction term is small compared to the main effect D.

b) The line segments connecting the two highest categories of education (5 to 6) are not parallel, but two of the points involve low frequencies and low statistical reliability.

c) The line segments joining the two non-formal categories (2 and 3) are not quite parallel, but they also involve two points with low sample frequencies and they are considerably displaced.

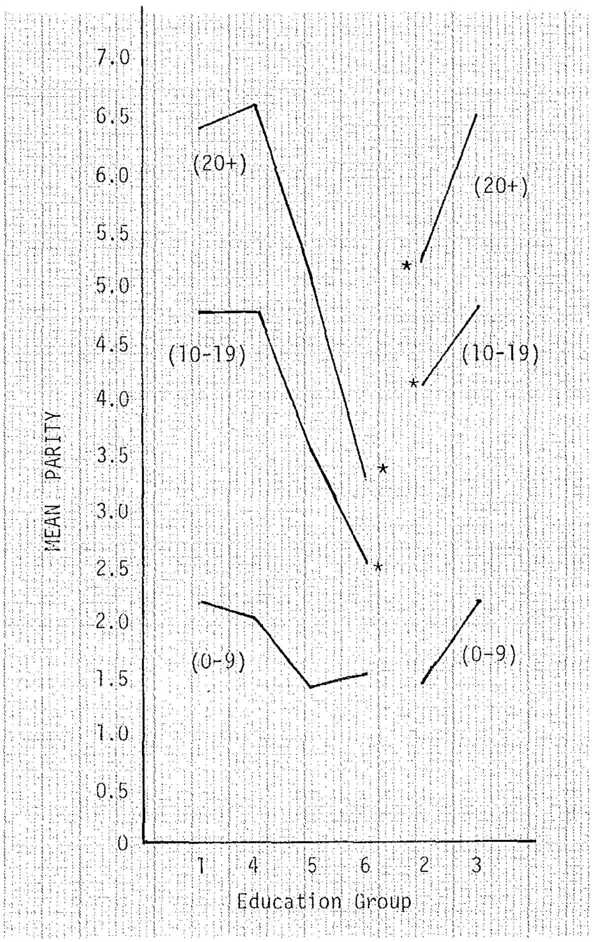
When the approximate test described in the first part of Appendix II is applied, the only interaction which appears to be significant (at the .01 level) stems from the entry in row 1 and column 4 of Table 4, which is about 10 per cent lower than one would expect if there were no interaction at all. This cell has the largest base frequency in the entire table.

TABLE 1 Mean number of children ever born, according to level of education, within broad groups of years since first marriage, for all ever-married women. (Malaysian Fertility and Family Survey, 1975, categories defined in text. Base frequencies are given in parentheses).

Years since first marriage		Level of Education $E(1)$	$E(2)$	$E(3)$	$E(4)$	$E(5)$	$E(6)$	All
	$D(1)$	2.15 (305)	1.43 (22)	2.15 (56)	1.99 (1413)	1.38 (508)	1.47 (46)	1.87 (2350)
	$D(2)$	4.75 (607)	4.10 (7)	4.83 (57)	4.73 (1012)	3.54 (130)	2.53 (18)	4.63 (1831)
	$D(3)$	6.39 (1329)	5.19 (19)	6.47 (67)	6.59 (652)	5.06 (59)	3.17 (6)	6.40 (2132)
	All	5.37 (2241)	3.31 (48)	4.61 (180)	3.87 (3077)	2.10 (697)	1.89 (70)	4.20 (6313)

FIGURE 1

A GRAPHICAL PRESENTATION OF TABLE I



NOTES:

1. *Years Since First Marriage given in parentheses.*
2. *Frequencies less than 20 are indicated by an asterisk (*).*

Of course, considerable interpretation of the data is possible from Figure 1 alone (or from Table 1 directly, for that matter, although the graphical presentation of the numbers is easier for most persons to comprehend). The general pattern is that, within levels of the control variable, a) women with religious education have only slightly higher fertility than those with 7-12 years formal education (categories 2 and 5); b) there is little difference in fertility between women of no education, other non-formal education, and some education (categories 1, 3, and 4); c) otherwise, increasing education results in lower fertility (categories 4, 5, and 6), although less so for women of less than 10 years' marital duration; and d) comparing extremes, women of the highest education have 30-50 per cent fewer children than those with no, or some, or other non-formal education. All of these findings differ from those based simply on the bottom row of Table 1.

The overall mean parity is 4.20. The final column of Table 1 gives the mean parity within each category of duration, and it is convenient to express the effect of each category as the deviation of the category-specific mean from the overall mean. These deviations, as a set, will be termed the *total effects of duration on mean parity*.

The category means are as follows:

	$D(1)$	$D(2)$	$D(3)$
LINE 1 Mean parity (P):	1.87	4.63	6.40

The deviations from the overall mean, line 1 minus 4.20, are

LINE 2 Effect of $D(i)$ on P:	-2.33	+0.44	+2.20

In other words, a woman married 0-9 years ago will have (on the average) 2.33 fewer children than the sample mean, and so on.

From the bottom row of Table 1 we get the mean parity of each educational category and then the deviations from the overall mean of 4.20, the total effects of education on mean parity.

		$E(1)$	$E(2)$	$E(3)$	$E(4)$	$E(5)$	$E(6)$
LINE 3	Mean Parity	5.37	3.31	4.61	3.87	2.10	1.89
LINE 4	Effect of $E(j)$ on P: (line 3 minus 4.20)	+1.17	-0.89	+0.41	-0.33	-2.10	-2.31

The role of marital duration is so obvious in relation to fertility that it would be unwise to place much emphasis on its role as a statistical predictor. We turn at once to the role of education. According to the total effects, women with seven or more years of formal education (categories 5 and 6) average two or more children fewer than the mean, and three or more children fewer than those with no education at all (category 1). But we have already noted that these marginal effects are misleading. The question is whether women of similar duration, but differing education, have differing levels of fertility.

We control, or "hold constant", the variable of marital duration by the technique of standardization. That is, the overall distribution of marital duration (given by frequencies in parentheses in the final column of Table 1), is applied within each education group. The standardized mean in the first educational category, for example, will be $(2350 \cdot 2.15 + 1831 \cdot 4.75 + 2132 \cdot 6.39)/6313 = 4.34$. The standardized mean parities for all educational groups are given in Line 5.

		$E(1)$	$E(2)$	$E(3)$	$E(4)$	$E(5)$	$E(6)$
LINE 5	Mean P standardized for D:	4.34	3.48	4.39	4.34	3.25	2.35

The overall mean usually changes, although very slightly, following standardization. There is no constraint that it should remain fixed, as there is in some statistical models. The new mean, which is calculated as a weighted average of the numbers in Line 5 with weights proportional to the number of women in each educational category, is 4.19. The deviations of the preceding numbers from this mean may be interpreted as <u>the net effect of a given level of education on mean parity, when duration of marriage has been controlled</u>:

	$E(1)$	$E(2)$	$E(3)$	$E(4)$	$E(5)$	$E(6)$
LINE 6 Net effect of $E(j)$ on P:	+0.15	-0.72	+0.20	+0.15	-0.94	-1.84

It must be noted that the role of marital duration has not been completely removed because in Table 1 it was given in only three very broad categories. Within each of these categories, better educated women will continue to have relatively shorter durations, both because they will be younger and because they will have married later. At best we will have eliminated most of the effect of this variable.

Even more important, it must be made explicit that, as with all techniques restricted to cross-sectional data, we have only "removed" a statistical association. These net effects do not necessarily represent "causal" effects. If, say, it had been possible to assume an unambiguous causal sequence to the predictor and the control, then a stronger statement could be made. In practice, of course, unambiguous causal sequences are infrequently found.

Line 6 implies that (given the control for duration):

1) women with no education or non-formal education, or less than seven years of education (categories 1, 3, and 4, respectively) have nearly identical fertility, .15 to .20 children above the mean;

2) the depressing effect (on fertility) of formal education takes place gradually, rather than occurring abruptly with 7 or more years of education, which would have been concluded from the total effects in line 4 (see categories 4, 5, and 6);

3) religious education (category 2) depresses fertility nearly as much as 7-12 years of formal education (category 5), by approximately .8 children below the mean.

As we should expect, there is a very close correspondence between these observations and those given for the graphical representation (Figure 1).

The standardized means (Line 5) are obviously closer to their mean, 4.19, than the unstandardized means (Line 3) are to their overall mean, 4.20.

The simplest way of measuring this change is in terms of the range. The unstandardized or marginal means within categories of education ranged from 1.89 to 5.37, a spread of 3.48. The standardized means, however, range from 2.35 to 4.39, a spread of 2.04. The latter range is only about three-fifths of the former.

A more comprehensive indicator of the reduced variability in mean parity following standardization comes from an analogy with the analysis of variance. The initial "between category sum of squares" is defined to be the sum obtained by squaring the effects (or deviations) in Line 4, respectively multiplying by the number of women in each educational category, and adding. These computations give the sum 6,920. When the same procedure is applied to the standardized effects in Line 6, a smaller quantity is obtained, viz., 998. This number is only 14 per cent of 6,920, and we may say, therefore, that the marginal variability in mean parity, according to education, has been reduced by 86 per cent by standardizing on broad categories of marital duration.

The change in the between-category sum of squares, before and after standardization, is a convenient guide to the importance of the control variable. However (see Appendix I), it is quite possible for this quantity to increase after standardization, when the control variable acts as a *suppressor* in the terminology of Lazarsfeld. And in no case is the calculation of a percentage change in this sum of squares to be considered a substitute for detailed examination of the effects.

Complementary to the net effects is the notion of *indirect effects*. These represent the portion of the total effect which is due to the association with marital duration. These are obtained simply by subtracting the net effects (Line 6) from the total effects (Line 4). These are given below in Line 7 (with Lines 4 and 6 repeated):

		$E(1)$	$E(2)$	$E(3)$	$E(4)$	$E(5)$	$E(6)$
LINE 4	Total effect of E on P:	+1.17	-0.89	+0.41	-0.33	-2.10	-2.31
LINE 6	Net effect of E on P, controlling for D:	+0.15	-0.72	+0.20	+0.15	-0.94	-1.84
LINE 7	Indirect effect of E on P, through D:	+1.02	-0.17	+0.21	-0.48	-1.16	-0.47

Note that a) the weighted sum of each of these lines is zero, when the weights are the proportions of the women in each educational category; b) Line 4 is the sum of Lines 6 and 7; and c) when the overall mean, 4.20, is added to each term in Line 4, the marginal mean parities in the respective educational categories will be returned.

For example, women with no education will have an average of .15 additional children as a net effect (from Line 6) but 1.02 additional children (from Line 7) through the indirect mechanism of greater marital duration. For these women, most of the effect of their lack of education is indirect. They are quite close to the mean overall parity once their marital duration has been taken into account.

At the other extreme, the women with more than 12 years of education (a small group, of course) have 1.84 fewer children as a net result of their education, and another .47 fewer as an indirect result. The large size of the net effect implies that these women have much reduced fertility within marriage.

The net and indirect effects are consistent in sign, except for educational category 4, less than seven years of formal schooling. The women in this category have a small positive net effect, the same as those women with no education at all, but it is overbalanced by an indirect effect of -.48 children.

To summarize the structure of the preceding discussion, we have inspected these relationships:

1. The total effect of duration on parity;
2. The total effect of education on parity;
    a. the net effect of education on parity, controlling for marital duration;
    b. the indirect effect of education on parity, due to marital duration.

A comparison was then made between 1 and 2, and between 2a and 2b. The comparison of the two total effects (1 and 2) indicated the relative importance of duration and education as predictors of parity; the comparison of direct and indirect effects of education (2a and 2b) indicated the extent to which the total effect of education is due to the association with marital duration.

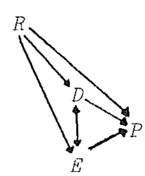
10. INCORPORATING ETHNICITY

Another variable may be added to the preceding discussion: ethnic or racial group. Actually, the analysis would otherwise be incomplete. The four categories for Malaysia are given below.

CATEGORY	PER CENT OF SAMPLE
$R(1)$: Malay	56.7
$R(2)$: Chinese	33.4
$R(3)$: Indian	9.3
$R(4)$: Other	0.6

The fourth category is simply a residual and is quite small. We shall not drop it, however, because our procedure involves standardizing on the marginal distribution of the whole sample. It will be included in the computations but will be largely ignored in the discussion.

Each woman's ethnic category is determined at the time of her birth, so this new variable stands logically prior to the three already in the analysis. The causal diagram is clearly as follows:



Ethnicity (R) may affect any of the other three variables, but can itself be affected by none of them.

Some of the observed relationship between education and mean parity may result simply from the prior impact of ethnicity. One can imagine, for example, an extreme situation in which a) all the women in Ethnic Group *X* - and only these women - have high education, and b) all the women in Ethnic Group *Y* - and only these women - have low education. Then, because ethnicity unambiguously precedes education, any observed differential between the parities of the low and high education groups would be misleading. The ultimate source of the differential (at least within this simple model) would be ethnicity, and it would be transmitted through the intervening variable of education.

Therefore, as we continue to refine our assessment of the impact of education (E) on mean parity (P) we need to control simultaneously for marital duration (D) and a clearly prior variable, ethnicity (R). The types of effects to be considered may be classified as follows:

1. The total effect of duration on parity;
2. The total effect of education on parity;
   a. The net effect of education on parity, controlling for marital duration and ethnicity;
   b. The indirect effect of education on parity due to the association with marital duration and the common antecedent, ethnicity;
3. The total effect of ethnicity on parity;
   a. The net effect of ethnicity on parity, given a control for education and duration;
   b. The indirect effect of ethnicity on parity, through education and duration.

Table 2, which is a reproduction of Table 2.2.5D in the Malaysian Report, gives the mean value of P for each combination of D, E, and R. The Table has 3x6x4 = 72 interior cells, many of them with small frequencies.

The first two sets of total effects in the above outline will be unchanged from the three-variable analysis, and are as given in Line 2 and Line 4 above. It is proposed to obtain new net effects (2a) by standardizing on a joint control variable, ethnicity x duration, with 4x3 = 12 categories.

TABLE 2

(Table 2.2.5D taken from the *Malaysia Fertility and Family Survey - First Country Report*)

ETHNIC GROUP	TOTAL	NO EDUCATION	RELIGIOUS EDUCATION	NON-FORMAL EDUCATION	LESS THAN 7 YEARS	7 - 12 YEARS	MORE THAN 12 YEARS
		LEVEL OF EDUCATION					
YEARS SINCE FIRST MARRIAGE TOTAL							
TOTAL							
NUMBER	6,313	2,241	48	180	3,077	697	70
MEAN NUMBER ..	4.2	5.4	3.3	4.6	3.9	2.1	1.9
MALAY							
NUMBER	3,580	1,422	44	57	1,725	314	18
MEAN NUMBER ..	4.2	5.2	3.3	4.2	3.8	1.8	2.1
CHINESE							
NUMBER	2,106	644	1	109	1,044	279	29
MEAN NUMBER ..	4.2	5.7	6.0	4.7	3.8	2.1	1.8
INDIAN							
NUMBER	589	166	3	13	303	91	13
MEAN NUMBER ..	4.6	6.1	3.7	5.9	4.4	3.0	2.0
OTHER							
NUMBER	38	9	-	1	5	13	10
MEAN NUMBER ..	2.3	3.1	0.0	1.0	3.4	2.2	1.5
YEARS SINCE FIRST MARRIAGE UNDER 10							
TOTAL							
NUMBER	2,350	305	22	56	1,413	508	46
MEAN NUMBER ..	1.9	2.2	1.5	2.1	2.0	1.4	1.5
MALAY							
NUMBER	1,212	150	21	17	747	262	15
MEAN NUMBER ..	1.7	1.7	1.5	2.2	1.8	1.3	1.8
CHINESE							
NUMBER	915	126	-	35	542	192	20
MEAN NUMBER ..	2.1	2.6	0.0	2.2	2.2	1.4	1.4
INDIAN							
NUMBER	206	26	1	3	122	46	8
MEAN NUMBER ..	2.1	2.5	0.0	1.7	2.2	1.8	1.3
OTHER							
NUMBER	17	3	-	1	2	8	3
MEAN NUMBER ..	1.2	2.7	0.0	1.0	0.0	1.3	0.7

TABLE 2 (cont'd)

(Table 2.2.5D taken from the *Malaysia Fertility and Family Survey – First Country Report)*

ETHNIC GROUP	TOTAL	LEVEL OF EDUCATION: NO EDUCATION	RELIGIOUS EDUCATION	NON-FORMAL EDUCATION	LESS THAN 7 YEARS	7 - 12 YEARS	MORE THAN 12 YEARS
YEARS SINCE FIRST MARRIAGE 10 - 19							
TOTAL							
NUMBER	1,831	607	7	57	1,012	130	18
MEAN NUMBER ..	4.7	4.8	4.1	4.8	4.8	3.5	2.5
MALAY							
NUMBER	1,020	357	7	20	591	42	3
MEAN NUMBER ..	4.6	4.5	4.1	4.9	4.6	4.1	3.7
CHINESE							
NUMBER	617	202	-	32	320	59	4
MEAN NUMBER ..	4.8	5.1	0.0	4.6	4.9	3.3	2.5
INDIAN							
NUMBER	181	46	-	5	99	27	4
MEAN NUMBER ..	4.8	5.3	0.0	6.0	5.0	3.3	2.8
OTHER							
NUMBER	13	2	-	-	2	2	7
MEAN NUMBER ..	2.4	2.5	0.0	0.0	4.5	2.0	1.9
YEARS SINCE FIRST MARRIAGE OVER 19							
TOTAL							
NUMBER	2,132	1,329	19	67	652	59	6
MEAN NUMBER ..	6.4	6.4	5.2	6.5	6.6	5.1	3.2
MALAY							
NUMBER	1,348	915	16	20	387	10	-
MEAN NUMBER ..	6.1	6.0	5.1	5.3	6.5	5.5	0.0
CHINESE							
NUMBER	574	316	1	42	182	28	5
MEAN NUMBER ..	6.8	7.2	6.0	6.8	6.6	4.6	2.8
INDIAN							
NUMBER	202	94	2	5	82	18	1
MEAN NUMBER ..	7.1	7.5	5.5	8.4	7.0	5.6	5.0
OTHER							
NUMBER	8	4	-	-	1	3	-
MEAN NUMBER ..	4.6	3.8	0.0	0.0	8.0	4.7	0.0

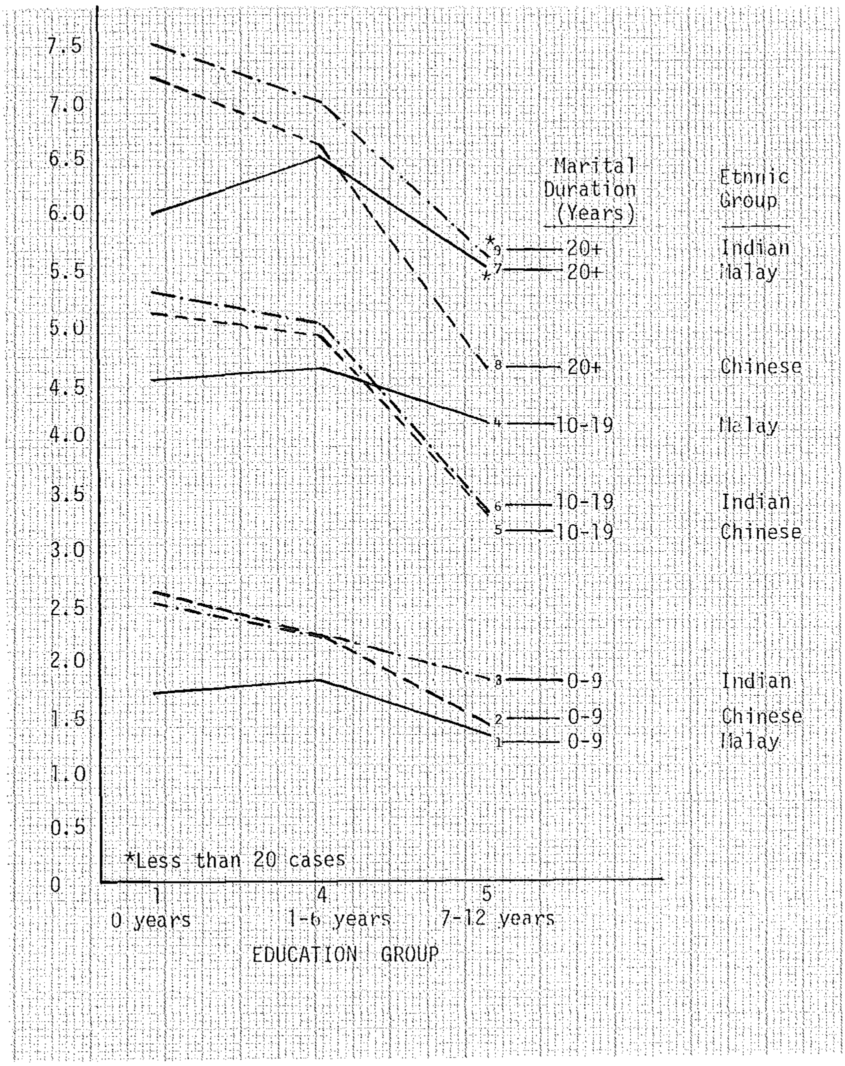
That is, we shall determine the proportion of women in each of these 12 categories, and then use these proportions as weights to calculate mean parity within each education group.

Before actually applying the method, a graphical check is again in order. For this purpose, we shall omit categories 2, 3, and 6 of the Education variable and the final, "Other" category of the Ethnicity variable. (However, no cells are omitted from any actual calculations to follow). The graph will therefore contain only 3x3x3 = 27 points, rather than 72, but will retain 6,015, or 95.3 per cent of the women in the sample. The check will thus ignore the statistically unstable cells. The conclusions of the next paragraph are supported by the approximate test suggested in Appendix II.

The graph is given as Figure 2 below; again, the positions on the horizontal axis refer to values of the predictor, in this case with a natural ordering, and points connected by line segments share the same value on the control variable - in this case, a joint variable. The vertically separated line segments are reasonably parallel, with an obvious exception. For the Malay women, an increase from 0 to 1-6 years of education leads to a slight increase in mean parity. Otherwise, it is uniformly the case that an increase in education leads to a reduction of parity, and the reduction from the middle to the high category is sharper than the one from the low to the middle category. We shall proceed with the standardization, focussing the discussion on the differential parity of the middle and high categories (4 and 5) of Education. Subsequently it will be shown we can more confidently deal with the degree of non-additivity indicated in Figure 2.

FIGURE 2

A GRAPHICAL PRESENTATION OF TABLE 2



When P is standardized on R and D, the mean parity and effects for each educational category are as follows:

		$E(1)$	$E(2)$	$E(3)$	$E(4)$	$E(5)$	$E(6)$
LINE 8	Mean P standardized for R and D	4.36	2.76	4.23	4.33	3.38	1.93
LINE 9	Net effects (Line 8 minus 4.19):	+.16	-1.43	+.04	+.14	-.82	-2.26
LINE 10	Indirect effects (Line 4 minus Line 9):	+1.01	+.54	+.37	-.47	-1.28	-.05

The overall mean under the standardization is 4.19, and when it is subtracted from the entries in Line 8 we get Line 9, corresponding to the net effects (2a) in the outline. When these are subtracted from the total effects we get the part of the total which is due to the relationship with duration and ethnicity, (2b) in the outline.

If Lines 9 and 10 are compared with Lines 6 and 7, respectively, in which the control for marital duration was given, it will be seen that the additional control for ethnicity has virtually no impact on the effects for Education categories 1, 4, and 5, which comprise over 95 per cent of the population. The direct effects of religious and maximum formal education (categories 2 and 6) as depressants of fertility seem even stronger than before, but we emphasize that the analyst should focus on categories 1, 4, and 5, because they contain most of the sample, and especially on the latter two, where there appears to be additivity.

The main conclusion would be that as a woman moves from education category 4 to 5, i.e., from 1-6 to 7-12 years of education, the net impact on her fertility will be to have one child (.14 + .82) less. Looking back at the bottom row of Table 1, we see that the initial impression would have been that she would have 1.8 fewer children, giving a substantial exaggeration of the education effect. Line 9 also indicates, as did Line 6, that there is <u>no</u> reduction in fertility as a woman moves from category 1 to 4, i.e., from none to 1-6 years of education, whereas Table 1 would have implied a drop of 1.5 children.

We have noted that the effects of type 2b in the classification given in Line 10 consist of two kinds combined: the association with D and effects due to the common antecedent, R. One could develop methods for subdiving these (as well as the effects of type 3b), but we shall not do so in the present document.

Consider now the effect of ethnicity, part 3 of the outline, using the data in Table 2.

The mean parity (P) within each ethnic group is given by Line 11:

		$R(1)$	$R(2)$	$R(3)$	$R(4)$
LINE 11	Mean parity (P)	4.17	4.16	4.65	2.37

The deviations from the overall mean are then as follows:

		$R(1)$	$R(2)$	$R(3)$	$R(4)$
LINE 12	Effect of $R(k)$ on P: (Line 11 minus 4.20)	-0.03	-0.04	+0.45	-1.83

The third ethnic group, the Indians, has approximately half a child more than the Malays or Chinese. The fourth group has markedly lower parity, but since this group is so small, the numbers are unreliable, and it would be unwise to pay much attention to it.

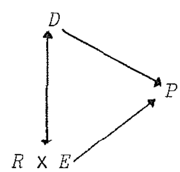
Table 2 also shows that the Malays have lower fertility than the Chinese or Indians within each category of marital duration, but when all categories of marital duration are combined, they are nearly identical with the Chinese. It would clearly be useful to have some method for expressing the impact of ethnicity *per se* on mean parity.

Our objective is to determine whether the extra half a child born to Indian women is simply a consequence of the other measured variables (E and D) or whether it is due to a direct effect - that is, to racial characteristics which we have not actually measured.

As usual, we make a graphical check, for possible interaction, again based on the 27 cells which contain 95.3 per cent of the sample. This appears as Figure 3. Although this figure consists of exactly the same points as Figure 2, connected by line segments in a different way, we see a greater degree of interaction, i.e., a greater departure from parallelism by vertically separated line segments.

The interaction shown in Figure 3 below is so great, in fact, that we shall <u>not</u> proceed with the standardization in part 3 of the outline. Some researchers might do so, but we shall instead turn to another formulation which will respond as well to the degree of interaction found in Figure 3.

The recourse is to a <u>joint predictor variable</u>, Ethnicity x Education ($R \times E$). We may examine the effect of this new variable on mean parity (P), controlling for duration (D), by reverting to this three-variable model:



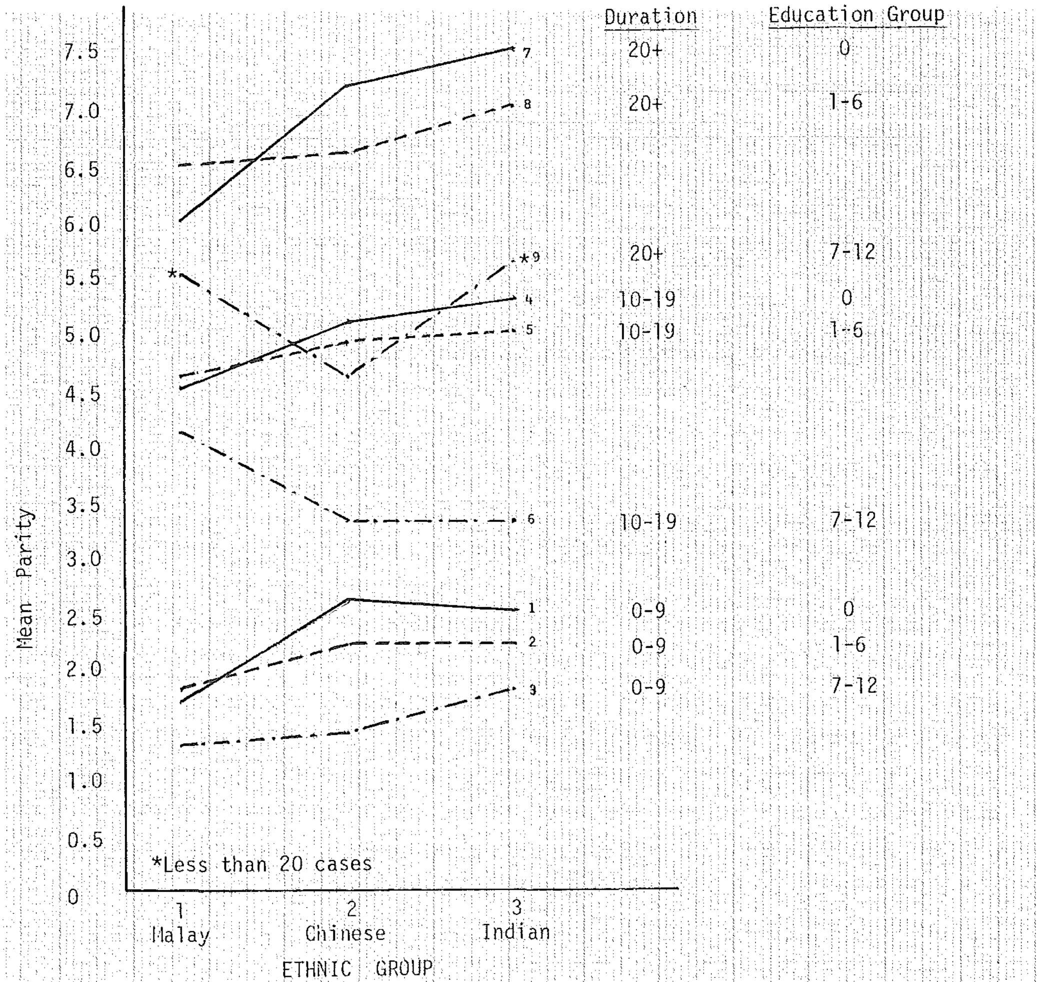
The outline for the decomposition of effects will be omitted because it is completely analogous to that given for the earlier three-variable model.

This new joint predictor has 4 x 6 = 24 categories, but, as we have already observed, 9 of them (associated with categories 1, 2, and 3 of Ethnicity and 1, 4, and 5 of Education) contain 95.3 per cent of the sample. The unstandardized mean parities (with one decimal place of accuracy) are given in tabular form in the first, or "totals" panel of Table 2.

The graphical check for interaction is given in Figure 4. Points connected by line segments have the same value on the control variable, years since first marriage. The order on the horizontal axis of the nine main categories of the joint predictor would not affect our conclusions, but is chosen to reflect both the temporal priority of ethnicity and the natural order of the main education categories.

FIGURE 3

SECOND GRAPHICAL PRESENTATION OF TABLE 2



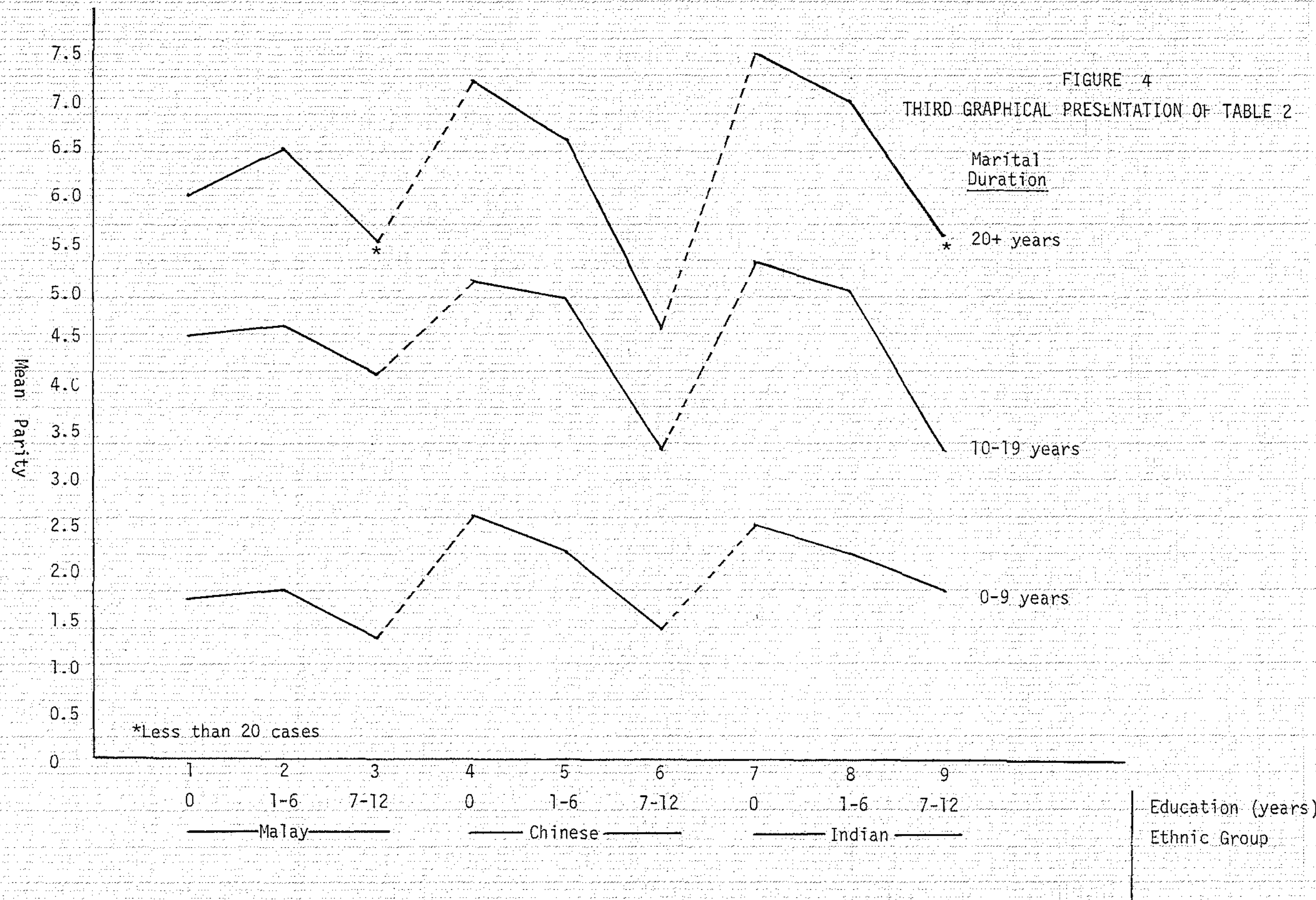


FIGURE 4
THIRD GRAPHICAL PRESENTATION OF TABLE 2

We observe less interaction here than in any of the other figures. The vertically separated lines are all very near to being parallel. The approximate test described in Appendix II supports this observation. Standardization is therefore appropriate.

Lines 13, 14, and 15 below correspond to Lines 4, 6, and 7 in Example 1. (They will be referred to as "Lines" even though they are presented as small tables). Line 13 gives the mean parities in each group minus the overall mean parity, 4.20; Line 14 gives the standardized mean parities in each group minus the overall standardized mean, 4.18; and Line 15 is Line 13 minus Line 14.

LINE 13 Total effects of $R \times E$ on P:

	LEVEL OF EDUCATION		
	$E(1)$ None	$E(4)$ 1-6 years	$E(5)$ 7-12 years
$R(1)$: Malays	+0.97	-.38	-2.39
$R(2)$: Chinese	+1.44	-.40	-2.08
$R(3)$: Indians	+1.91	+.22	-1.20

LINE 14 Net effects of $R \times E$ on P, standardized for D:

	$E(1)$ None	$E(4)$ 1-6 years	$E(5)$ 7-12 years
$R(1)$: Malays	-.21	+.02	-.65
$R(2)$: Chinese	+.70	+.29	-1.14
$R(3)$: Indians	+.83	+.46	-.66

LINE 15 Indirect effects of $R \times E$ on P, due to association with D:

	$E(1)$ None	$E(4)$ 1-6 years	$E(5)$ 7-12 years
$R(1)$: Malays	+1.18	-.41	-1.74
$R(2)$: Chinese	+0.74	-.70	-.93
$R(3)$: Indians	+1.09	-.24	-.54

Note that a) the weighted sum of each of these lines is zero, when the weights are the proportions of the women in each ethnicity x education category; b) Line 13 is the sum of Line 14 and Line 15; and c) when the overall mean, 4.20, is added to each term in Line 13, the marginal mean parities in the respective ethnicity x education categories will be returned.

The reader should briefly compare Line 14 with Line 6, which gave the net effects of education when the woman's ethnicity was ignored, in order to see the importance of ethnicity and the degree of misinterpretation that could have resulted from the omission of that variable.

The following comments are based on Line 14:

1) For the Chinese and Indians, an increase in education produces a decrease in fertility (holding duration constant). For both groups, the differential resulting from a shift from no education to 1-6 years of education is about the same: .4 of a child. The effect of a shift to 7-12 years of education is even greater for these two groups. For example, among the Chinese, women of middle education have .29 more children than the overall mean, but those with high education have 1.14 less. Indeed, this is the most dramatic differential in the table.

2) The educational effect is least for the Malay women. If anything, a shift from low to medium education produces a small increase in fertility. A shift from middle to high education gives a reduction in mean parity, but by a smaller amount than in the other ethnic groups. But one should not lose sight of the fact that the Malay women, as a group, consistently have the lowest fertility, for reasons not discernible with the present set of variables.

Next, we compare the net and indirect effects (Lines 14 and 15). When an indirect effect has the same sign as a corresponding net effect, interpretation is usually a bit simpler. In this example, the only measured co-variate within each category of ethnicity x education (R x E) is marital duration (D). The simplest situation is one in which all co-variates, measured and unmeasured, work in the same way, either to raise or to lower the mean parity. With the present model, the best one can do is to evaluate the aggregated effect of the unmeasured variables which are associated with marital duration (the indirect effect) and the aggregated effect of those which are not (the net effect).

Consider, for example, the highly educated Chinese women. Their parity is 2.08 below the overall mean. About half of this reduction (1.14) remains when we standardize on marital duration, and can therefore be said to be net of the association with duration. The remainder (.93) is the indirect effect; its negative value implies that women in this group have a relatively low marital duration. Because the net effect has the same sign, we infer that the measured and unmeasured co-variates tend to act in unison to reduce fertility in the group.

For this application, the direct and indirect effects have the same sign, or else one of them is negligible, for most categories of R x E. The most notable exception, in which suppression has occurred, would be for the medium-educated Chinese women. These women have slightly higher (+.29) parity than the overall mean when the role of duration is removed, but a negative indirect effect (-.70), possibly because they marry late, resulting in a small negative total effect (-.40).

The analyst could also disaggregate the sample into the separate ethnic groups, and repeat the analysis of Example 1 on each group. The results would be very nearly, although not exactly, as given above. The main difference would be that the overall mean above would be replaced within each ethnic group by the mean for that ethnic group, so that the reference parity would be different.

Two final comments are in order concerning our attempts to control for marital duration. First, we may be quite sure that any technique which uses ten-year intervals of duration, as we have necessarily done, will only partially control for this variable. If we passed to five-year or one-year intervals, the net effects of education, for example, would be even smaller, and might disappear altogether. This is because the better-educated women will tend to have shorter durations of marriage even within the ten-year intervals.

Second, if one seeks policy implications, the observation that most of the effect of education is indirect suggests that other ways of delaying marriage (since this is the policy-related component of shorter marital duration) may be just as important in modifying fertility. We have not separated the effect of marriage duration into its two components, but other data in the Malaysian Report (Table 2.2.3A) show that, within marriage cohorts, women who marry at age 20 and older average nearly one child less than those who marry earlier. This suggests (but does not prove) that delayed marriage is a practical mechanism through which education has its total effect upon fertility. There may be alternatives in labour force policy or laws relating to age at marriage which will produce the same results for fertility, and which may have quicker impact. (This is not to be taken as a devaluation of the other benefits of education).

11. CONCLUSION

It would be possible to provide many more examples of standardization, of course, but most of the relevant considerations have already been discussed. In most applications there will be three or four variables, with the dependent variable measured by a mean, such as mean additional number of children wanted, or by a proportion, such as the proportion wanting another child. if the dependent variable cannot be expressed by a mean and has more than three categories as for, say, type of contraceptive method, then interpretation will usually be easier if the analyst dichotomizes on the basis of one category of interest. However, the method itself does not require this; it is possible to carry along any number of categories in the dependent variable.

The procedure has two purposes. The first one is to reduce the data by eliminating one or more dimensions from a many-way tabulation, while taking into account the role of the dropped variables. Users of survey reports quite often neglect to examine, understandably enough, all the panels and all the interiors of complex tables. Instead, they tend simply to look at the margins, and thereby often reach invalid conclusions. Even the writers of the reports are liable to make the same kind of misinterpretations. Standardization can take into account the compositional or control variables in a manner which is familiar to nearly all demographers and which can be done with no special calculating equipment.

Secondly, apart from presentation, standardization is an analytic tool. It is quite analogous to multiple classification analysis (MCA) and, when appropriate, should give very nearly the same conclusions. Both methods share the limitation (not well publicized in either case) to additivity. In other words, the graphical check (or statistical check) for interaction which should precede standardization is every bit as important for MCA. For the analysis, we have suggested that one construct a model, and compare standardized and unstandardized means (or proportions) in a manner similar to path analysis, in which total effects (zero-order correlations) are decomposed into net or direct effects, indirect effects, and spurious components due to a common cause. (Path analysis, it may also be mentioned, customarily ignores interaction effects).

In applying the technique, one must avoid the subtle fallacy of regarding standardized rates as "true" or "underlying". Rather, they are the result of specific choices of control variables and standard distributions. Their interpretation must always have reference to these choices. Moreover, the analyst must always bear in mind the differences in the actual observed quantities, and guard against an over-emphasis on the differentials in the standardized quantities.

It is up to the user to decide whether to apply the procedure at all and, if it is applied, whether to choose the more elaborate, analytic form described here. The choice will depend upon the time and other resources available, as with any other procedure, but the principal advantage of standardization is that it requires less in the way of resources than other procedures known to us. When more complex methods can be used, then they most certainly should be; the essential results will be the same for any method, however, so long as the assumptions are satisfied.

REFERENCES

Andrews, F.M., J.N. Morgan, J.A. Sonquist, and L. Klem. *Multiple Classification Analysis.* Ann Arbor, Michigan, The University of Michigan, 1973.

Atchley, R. "A Qualification of Test Factor Standardization". *Social Forces* 47, 1969: 84-85.

Barclay, G.W. *Techniques of Population Analysis.* New York, John Wiley and Sons, 1958.

Bishop, Y.M.M., S.E. Fienberg, and P.W. Holland. *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, Mass., The MIT Press, 1975.

Duncan, O.D. *Introduction to Structural Equation Models.* New York, Academic Press, 1975.

Jaffe, A.J. *Handbook of Statistical Methods for Demographers.* Washington D.C., U.S. Bureau of the Census, 1951.

Kalton, G. "Standardization: A Technique to Control for Extraneous Variables". *Applied Statistics* 23, 1968: 118-136.

Kendall, P.L., and P.F. Lazarsfeld. "Problems of Survey Analysis" in R.K. Merton and P.F. Lazarsfeld, eds., *Continuities in Social Research: Studies in the Scope and Method of "The American Soldier".* Glencoe, Ill., The Free Press, 1950.

Kitagawa, E.M. "Standardized Comparisons in Population Research". *Demography* 1, 1964: 296-315.

Malaysian Fertility and Family Survey - First Country Report, Department of Statistics, Kuala Lumpur, Malaysia, 1977.

Rosenberg, M. "Test Factor Standardization as a Method of Interpretation". *Social Forces* 41, 1962: 53-61.

Shryock, H.S., and J.S. Siegel. *The Methods and Materials of Demography,* Washington D.C., U.S. Bureau of the Census, 1971.

APPENDIX I

ILLUSTRATION OF THE COMPUTATIONS

The attached worksheet indicates how standardization may be done with a hand calculator, using data from the 1974 Fiji Fertility Survey. The table relates never-use of contraception to education and parity.

The initial data which are required consist of the unstandardized percentages of never-users for each combination of education (the predictor variable) and parity (the control variable), plus the overall distribution of the control variable. These are given in columns (1), (3), (5), (7), and (9). In this example, the distribution on the control variable is given in terms of observed frequencies, but it could equally satisfactorily be given as a percentage distribution. The overall sample size is 4,928. If the last column were a percentage distribution, then, in effect, the role of the sample size would be assumed in our calculations by the number 100.

The first step is to apply the observed percentages of users to the overall distribution on the control variables. This gives the expected numbers of never-users on each combination of parity and education if the overall distribution on the control obtained within each category of education. These numbers are given in the even numbered columns. Thus, Column (2) is the product of columns (1) and (9); Column (4) is the product of (3) and (9); Column (6) is the product of (5) and (9); and Column (8) is the product of (7) and (9). For example, in Column (2), 500 is 82.7 per cent of 605, 354 is 51.3 per cent of 690, and so on.

Second, the numbers thus generated in each even-numbered column are added and divided by the total sample size, 4,928. The result is the standardized percentage of never-users in each educational category. Thus, in Column (2): 500 + 354 + ... + 338 = 1,873. Since 1,873 is 38 per cent of 4,928, we have 38.0 per cent as the standardized percentage of never-users among women with no education, and so on.

FIJI: PERCENTAGE OF ALL EVER-MARRIED WOMEN OF ALL RACES WHO HAVE NEVER USED ANY CONTRACEPTIVE METHOD, BY NUMBER OF LIVING CHILDREN AND LEVEL OF EDUCATION

(*Guidelines to Country Report No. 1*, Table 4.3.2)

Columns:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Educational Level:	Sub-populations of interest								Standard population
	None		Lower primary		Upper primary		Secondary Plus		All educational categories combined
	% Ob.	Ex. N	% Ob.	Ex. N	% Ob.	Ex. N	% Ob.	Ex. N	Ob. N.
Parity: 0	82.7	500	80.0	484	77.5	469	59.7	361	605
1	51.3	354	54.4	375	45.1	311	26.8	185	690
2	31.3	237	37.9	287	31.1	236	15.2	115	758
3	38.1	264	24.6	170	23.5	163	18.0	125	692
4	31.5	180	20.7	118	18.7	107	13.2	76	572
5+	21.0	338	18.8	303	19.6	316	8.9	143	1,611
All parities observed	32.6		31.6		35.1		27.7		
Expected N		1873		1737		1602		1005	4,928
Expected (Standardized %)		38.0		35.2		32.5		20.4	

It will be observed that the standardized percentages are much more variable than the standardized ones. This is because parity suppresses the effect of education on never-use. The net effect of high education is to increase use, i.e., to decrease never-use; but the indirect effect is to increase never-use, because better educated women will have lower parity, and women with lower parity are less likely to feel a need for contraception. Therefore the direct effects of education are considerably greater than one might have thought at first.

It is essential to note that in Fiji, as in most developing countries, contraception is used mainly to terminate childbearing, rather than to control birth intervals. Therefore, as a rule, current parity is prior to contraceptive use or non-use. If there were substantial reciprocal effects between the dependent variable and either a predictor or a control, then standardization would not be advisable.

APPENDIX II

APPROXIMATE TESTS FOR INTERACTION

CASE 1: CELL ENTRY IS A MEAN

A complete test for the type of interaction being considered here would require knowledge of the sum of squares within each cell, and in our context this information is simply not available. Fortunately, for most applications in Country Report No.1, a good estimate can be made. If the cell entry is mean parity (or mean additional number of children wanted, or mean total number of children desired) then the individual response can be assumed to have approximately a Poisson distribution within each cell. If the mean in row i and column j is referred to as $\bar{x}_{ij}$, and it is based on n_{ij} cases, then the variance of the response will be estimated to be $\bar{x}_{ij}$ as well, the within-cell sum of squares will be estimated to be $\bar{x}_{ij}n_{ij}$, and the estimated variance of the sample mean will be $\bar{x}_{ij}/n_{ij}$.

We shall suggest that checks be made on 2 x 2 subtables corresponding to pairs of vertically displaced lines in the graphical check (see, for example, Figures 1, 2, and 3). That is, arbitrarily numbering the rows and columns of such a subtable by the numbers 1 and 2, we check a subtable such as the following:

		Predictor	
		1	2
Control Variable	1	$\bar{x}_{11}$ (n_{11})	$\bar{x}_{12}$ (n_{12})
	2	$\bar{x}_{21}$ (n_{21})	$\bar{x}_{22}$ (n_{22})

A test of whether the population difference between the means in the first row is equal to the population difference between the means in the second row may be based on the statistic

$$U = \frac{\bar{x}_{11} - \bar{x}_{12} - \bar{x}_{21} + \bar{x}_{22}}{\sqrt{\frac{\bar{x}_{11}}{n_{11}} + \frac{\bar{x}_{12}}{n_{12}} + \frac{\bar{x}_{21}}{n_{21}} + \frac{\bar{x}_{22}}{n_{22}}}}$$

If the entire table has only two rows and two columns, then U will be approximately a unit normal deviate, and the hypothesis will be rejected at the α level if

$$|U| > Z_{1 - \alpha/2}$$

where Z_p is the cumulative $100p$ percentile level of the tabulated unit normal variate.

If there are R rows and C columns, then $(R - 1)(C - 1)$ independent subtables may be formed. If the statistics U were calculated for each of these tables, and squared and added, the resulting sum would have approximately a χ^2 distribution with $(R - 1)(C - 1)$ degrees of freedom.

Rather than testing the table as a whole, however, we prefer to consider the subtables themselves. So long as there is additivity in most of the table, and the commentary is limited to that part of the table, standardization will be valid. When a 2 x 2 subtable is checked because there is a suspicion of interaction, there is the complication that the tabulated levels of the unit normal deviate will not apply. Nevertheless, we shall take these as an approximation. Thus, for example, a .01 level test in the subtable will lead to rejection of the hypothesis of no interaction if $|U| > 2.58$.

CASE 2. CELL ENTRY IS A PROPORTION

Say that the dependent variable is a dichotomy and P_{ij} is the proportion of observed cases in cell (i, j) which have the criterion characteristic.

The within-cells sum of squares will be $P_{ij}(1 - P_{ij})n_{ij}$ and the estimated variance of the sample proportion will be $P_{ij}(1 - P_{ij})/n_{ij}$.

Most of the preceding discussion will be unchanged except that U should be redefined to be

$$U = \frac{\log\left[\dfrac{P_{11}\ P_{22}}{P_{12}\ P_{21}}\right]}{\sqrt{\dfrac{1}{P_{11}(1 - P_{11})n_{11}} + \ldots + \dfrac{1}{P_{22}(1 - P_{22})n_{22}}}}$$

Those parts of a table which show non-additive effects should be de-emphasized in the part of the discussion based on standardization. However, they should certainly not be ignored otherwise. The identification of these effects and their nature forms an important part of the analysis.

APPENDIX III

TREATMENT OF EMPTY CELLS

It may well happen that some cells in the initial cross-classification will be empty. That is, for some combinations of predictors and controls there may be no base frequency on which to compute a cell mean or proportion. The normal default in such cases would simply be to insert a value of 0.0 for the unavailable mean or proportion.

Empty cells are essentially of two types. In the first instance, they may be empty in the entire population from which the sample was drawn, perhaps for logical reasons (e.g., it is impossible for a woman aged 15-19 to have been married more than twenty years) or empirical reasons (e.g., there may be no women aged 15-19 with 10 children).

In the second instance, the empty cell may result from an inadequate sample size. The observed cell mean or proportion is an estimate of a population quantity, and as the base frequency is reduced, the quality of the estimate deteriorates; in the extreme case of an empty cell, the estimate is simply at its worst.

With respect to the technique of standardization, no distinction can in fact be made between the circumstances outlined above. With more sophisticated techniques some distinctions can be made, but we see no easy way of doing so with the present method.

The more empty cells there are in a table, the less acceptable is the technique of standardization. One should attempt to keep the number of such cells at a minimum, either by reducing the number of predictors and controls, or else by combining categories. Suppose, for example, that age and parity are used to construct a joint control variable in examining the impact of education upon desire for more children. Say that age has four categories (<20, 20-29, 30-39, 40+) and parity has six categories (0, 1, 2, 3, 4, 5+). The joint control would have 4 x 6 = 24 categories. If it were found that there were several empty cells involving 5+ children (within categories of the predictor, education) then the 5+ category might be combined with parity 4. This would reduce parity to five categories and the joint control to 4 x 5 = 20 categories. Another strategy would be to combine several of the age x parity categories by themselves. Thus, if it were only the young women who rarely had high parities, one could group the 24 basic categories into, say, 19, as indicated in the following table (cell entires are the category numbers of the joint control):

<table>
<tr><th></th><th></th><th colspan="6">Parity</th></tr>
<tr><th></th><th></th><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5+</th></tr>
<tr><td rowspan="4">Current Age</td><td><20</td><td>1</td><td>2</td><td>3</td><td colspan="3">4</td></tr>
<tr><td>20-29</td><td>5</td><td>6</td><td>7</td><td>8</td><td colspan="2">9</td></tr>
<tr><td>30-39</td><td colspan="2">10</td><td>11</td><td>12</td><td>13</td><td>14</td></tr>
<tr><td>40+</td><td colspan="2">15</td><td>16</td><td>17</td><td>18</td><td>19</td></tr>
</table>

Similarly, one may combine categories of a predictor or joint predictor.

The question still remains of what to do with the cells which are empty even if there are very few of them. If they are, say, fewer than one cell in twenty, then one may as well ignore them, and proceed with the usual default of a zero mean or proportion. A substituted estimate of the population mean or proportion in these cells would have little effect on the conclusions, particularly since the cell entries are generally small anyway. (If, say, one were studying income differentials, and the cell entry were mean income, then a default of zero in an empty cell could result in distortions.)

Otherwise, an estimated entry, even if not a particularly sophisticated one, will be preferable. We recommend the following procedure. It is phrased in terms of a mean, but would also apply to a proportion (or logit, etc.).

a. Compute the overall mean and the total effects for each predictor (or joint predictor) and each control (or joint control).
b. Estimate the mean in an empty cell by adding the mean and all of the total effects pertaining to that cell.
c. Use this estimate (with a zero base frequency) in the actual standardization.