Pakistan Journal of Applied Economics (1982) vol.I, no.1 (59–91).

THE DETERMINANTS OF MORTALITY IN A LOW INCOME AREA OF KARACHI

James C. KNOWLES*

Data from a low-income area of Karachi are used to show that family income and lactation behaviour are significant determinants of survival rates among pre-school children and infants. This suggests that the conventional view that only medical improvements have been responsible for the recent decline in mortality rates in developing countries should be modified.

I. Introduction

In the post World War II period most developing countries have experienced dramatic declines in mortality rates. In Pakistan it has been estimated that average life expectancy at birth increased from 33 to 52 years between 1950-52 and 1968-71 [Farooqui and Alam, (1974)]. The conventional view is that rapid declines in mortality rates have occurred not as the result of a rising standard of living in the developing countries but as the result of introducing modern public health techniques (e.g., mass immunizations, disease eradication) and, to a lesser extent, curative health services [see, for example, Stolnitz, (1955, 1974)]. Several arguments have been advanced to support this view. First, mortality rates have fallen fairly uniformly throughout the developing world, even in coutries which have not experienced extensive economic development. Second, the rate at which mortality rates have fallen in developing countries appears to have been too rapid to be explained primarily by economic growth.

There are, nevertheless, an increasing number of researchers who argue that socio-economic factors have made an important contribution toward reducing mortality rates during the post-war period. It has been pointed out, for example, that even in Sri Lanka, where mortality rates fell dramatically following a successful anti-malaria compaign, they fell dramatically in previ-

*This paper was written while the author was visiting Reasearch Professor in the Applied Economics Research Centre, University of Karachi, and Project Specialist for the Ford Foundation. He wishes to thank Professor Dr. Razia Rahimtoola of the Department of Paediatrics, Jinnah Postgraduate Medical Centre, Karachi, for supplying the data and for comments on an earlier draft.

ously non-malaria areas as well [Frederiksen, (1966, 1971)]. It has also been pointed out that gastro-intestinal infections among infants and small children, which are among the major causes of high mortality in developing countries, are not very responsive to inexpensive forms of medical treatment [Schultz, (1976)]. Finally, recently completed research on the determinants of child mortality in Kenya has produced evidence that socio-economic factors, including household income, education, toilet facilities, and use of modern (as opposed to traditional) forms of medical care, all contribute significantly to higher survival rates [Anker and Knowles, (1980)].

The purpose of the present study is to learn more about the factors which affect mortality rates in developing countries. The need for such studies arises not only from our present ignorance about the relative importance of the factors responsible for recent declines in mortality but also from a desire to learn how mortality rates are likely to change in the future in response not only to health policies but also to economic development. Further reductions in mortality rates can make an important contribution to the process of economic development where "development" is given a somewhat broader interpretation than increasing GNP per capita. The results of the present study may be useful to health planners in designing policies to accelerate the decline in mortality rates. For example, the results of the present study suggest that ante-natal care provided by traditional practitioners (dais) compares favourably with that provided by modern practitioners from the standpoint of its impact on infant and pre-school mortality rates.

The present study consists of an analysis of mortality differentials among infants and pre-school children based on data from a cross-section sample of 736 households from one low-income area of Karachi. The study utilizes multiple regression analysis in an attempt to identify the principal factors which account for variation between the sample households in the proportion of children surviving to three years of age.

The principal findings of the study are as follows. Household income per adult (and more particularly food expenditures per adult), sanitary refuse disposal, and prolonged breastfeeding are observed to have strongly significant positive effects on survival rates. Years married is observed to have a strongly significant negative effect on survival rates, implying that children born in the past had significantly lower survival rates. The results of the present study are consistent with the view that average life expectancy at birth has been rising at a rate of almost one year per annum during the post-war period and that economic development, as distinct from the introduction of public health measures, has made an important contribution to this trend.

The paper is organized into five sections. In addition to the introduction (section I), section II discusses the characteristics of the data set, section III contains a brief theoretical discussion of the factors believed to affect mortality rates among infants and pre-school children, section IV contains the regression analysis, and section V consists of a summary and tentative conclusions.

II. The Data

The data are drawn from a survey conducted by Professor Dr. R.J. Rahimtoola and her colleagues of the Department of Pediatrics, Jinnah Postgraduate Medical Centre, Karachi. The survey, which was designed as a baseline survey for the Children's Nutrition Project, was carried out in Karachi between August, 1976 and June, 1977. According to Dr. Rahimtoola (1977, pp. 1-2), the location of the survey ("Mahmoodabad Colony") is a slum area, marked by "unhealthy living conditions":

> The houses are adjoining one another with no space in between. A row of eight to ten houses have a small space of about six feet at the back and front where another row of houses is present. The lanes have filthy water and rubbish thrown by the residents as they have no proper sanitation or refuse disposal system. Most of the houses have one room approximately 8'x 12' in size and a small courtyard 6'x 8' where cooking is done. . . Some houses have a large open courtyard, about 10'x 12', where goats and fowl are kept. The children play in the courtyard and as conditions are unsanitary due to the animals, they have repeated infections. . . Half the houses are dirty inside partly due to the carelessness of the families and partly due to lack of proper water supply and toilet facilities.

The sample, which consists of 736 households, was selected on a quota basis from the proximity of a dispensary operated by the Karachi Metropolitan Corporation (KMC). The KMC dispensary served as the project's field office. Research assistants visited the dwellings of the families to collect the baseline data, while clinical examinations of the children were conducted by members of the Jinnah team at the KMC dispensary.

The data collected in the survey included information on the education of both parents, household income, family size, weekly food expenditures, monthly medical expenditures (based on an average of the preceding three months), housing and sanitation facilities (number of rooms, type of toilet facility, source of water supply, method of refuse disposal), the mother's present age, her age at marriage, and information on her past fertility (numbers of pregnancies, abortions and miscarriages, still births, and full-term deliveries), child mortality data (the number and ages of all children who died before five years of age), information about any diet changes during pregnancy (including types of any food added to the mother's diet and whether vitamins and minerals were taken), whether the mother obtained ante-natal care and the source of any care received (traditional dai, maternal/child health centre, public hospital, private practitioner), and information on the length of breastfeeding and age at which solid foods were added to children's diets.

The basic fertility and mortality data from the survey are summarized in Table 1. Women in the survey reported a total of 3,367 pregnancies (an average of 4.60 pregnancies per woman), of which 3,035 resulted in a live birth. Of the 3,035 live births, 495 (16.3 percent) died before reaching five years of age. Most of these deaths occurred in the first year of life (70.9 percent), and almost one-half of these (45.3 percent) occurred in the first month of life.[1]

The data in Table 1 indicate that there is some undercount of mortality in the survey. Part of the apparent undercount arises from the fact that the data refer to all children ever born to women in the survey, including those born within the last five years. More accurate mortality rate estimates could be obtained from data limited to children born more than five years before the survey, so that all children would have had an equal exposure to the risk of death. Unfortunately, the survey did not provide information on children's dates of birth. Nevertheless, the fact that most deaths occur in the first year of life (and almost one-third occur in the first month of life) implies that the inclusion of children under five in Table 1 is responsible to only a minor extent for the apparent undercount of mortality.

Table 2 compares the survey estimates of various survival rates (i.e., the probability of a child surviving to a given year) to corresponding survival rates from Coale-Demeny West model life tables corresponding to two different levels of average female life expectancy at birth (LX = 45 and 55 years).[2] It is clear from Table 2 that the survival rate estimates obtained from the survey data resemble very closely the model life table rates corresponding to an LX of 55 years, which compares favourably with the estimate of LX (53.6 years) for urban females obtained by Iqbal and Alam (1974, p. 341). Nevertheless, there are two reasons for believing that there is a significant undercount of mortality in the survey. First the survey data are drawn from relatively low income families, and *a priori* one would expect

[1] There was some misreporting of age at death in the survey, as evidenced by a clustering of the reported ages at 12-month intervals. The data in Table 1 are based on a smoothed age distribution of deaths.

[2] See Coale and Demeny (1966).

TABLE 1

Summary of fertility and mortality data from the Jinnah Survey

Pregnancies:		3,367
Miscarriages/abortions		264
Still births		68
Live births:		3,035
Infant deaths (0-11 months)		351
(0 months)	(159)	
(1-11 months)	(192)	
Child deaths (12-59 months):		144
(12-23 months)	(65)	
(24-35 months)	(40)	
(36-59 months)	(39)	
Total infant and child deaths		495

TABLE 2

Comparison of survival rates in the Jinnah Survey to corresponding rates from West Model Life Tables

Survival rates (LX)	Jinnah Survey	Model life tables	
		$\mathring{e}_o = 55$	$\mathring{e}_o = 45$
To age one (l_1)	.8843	.8974	.8408
To age two (l_2)	.8629	.8742	.8002
To age three (l_3)	.8498	.8639	.7822
To age five (l_5)	.8369	.8521	.7617

TABLE 3

Mortality differentials by years married, family income and education of husband and wife

Criterion variable	N	Survival rate to age 3
Years married:		
0-4 years	118	.950
5-9 years	225	.890
10-14 years	187	.848
15-19 years	109	.865
20+years	90	.845
	729	
Family income:		
0-249 Rs. per month	79	.838
250-499 Rs. per month	513	.872
500-749 Rs. per month	90	.884
750+Rs. per month	31	.919
	713	
Husband's education:		
Illiterate[a]	311	.844
Literate and primary	178	.883
Secondary and higher	240	.899
	729	
Wife's education:		
Illiterate[a]	596	.860
Literate and primary	91	.894
Secondary and higher	43	.979
	730	

[a]Literacy defined as being able to write one's name and read the Holy Koran.

this sub-group to exhibit considerably lower average life expectancy at birth than the urban population as a whole. Secondly, most of the births reported in the survey (about two-thirds) occurred prior to 1971. If average life expectancy has been rising as rapidly as believed, the mortality rates for the sample children should resemble more closely those of a population with an LX of about 45 years. It is apparent from the data in Table 2 that they do not.

Despite the apparent undercount of the mortality data in the survey, it should be possible to utilize this data set in a multivariate analysis of mortality differentials for the following reasons. First, although the average *level* of mortality appears to be under-reported the *differentials* by years married, family income, and husband's and wife's education appear quite reasonable (Table 3). Second, one would expect biased results in a multivariate analysis only if the degree of under-reporting were systematically related to the factors affecting mortality. Even if such systematic bias exists—for example if better educated women or women from higher income families provide more accurate mortality data—then the nature of the bias is likely to operate in the direction of biasing the observed effects of socio-economic variables toward zero. Any evidence obtained from the regression analysis which were to suggest that socio-economic factors significantly affect mortality would then be relatively conservative from a statistical standpoint.

III. Theoretical Discussion

The factors which theoretically should determine the probability of survival among infant and pre-school children include both macro-level and micro-level variables. The macro-level, or environmental factors, include variables such as disease prevalence and the availability of medical services. These variables are usually considered to be beyond the control of individual households. The proximate (or direct) micro-level determinants of mortality refer to actual practices of households with respect to such variables as nutrition, hygiene, infant feeding, and the type of medical care utilized. Alternatively, the indirect determinants of mortality refer to characteristics of households (e.g., income, parents' education, quality of housing) which may influnce or constrain actual behavior. Both the proximate and indirect micro-level determinants of mortality are clearly linked to the process of socio-economic development. Changes in the macro-level variables, on the other hand, may take place even in the absence of development. As noted above, many observers believe that recent declines in mortality rates in developing countries have been largely due to changes in macro-level variables resulting from public health measures such as disease eradication, mass immunization, and the provision of the safe drinking water.

Micro-Level Determinants of Mortality

In the present study, the variables which are hypothesized to affect the probability of survival among children refer to the mother's number of years married, parents' education, mother's nutritional status, infant feeding practices, use of medical services, and the quality of housing and sanitation.

Years married: The dependent variable in the regression analysis is the proportion of children ever born who survived to three years of age. The fact that the data include children who are not yet three years of age means that the dependent variable, as a measure of the probability of survival to age three, is biased upwards.[3] This bias is larger for women with a high proportion of their children under three years of age. Since such women tend to be those who have married fairly recently, the number of years married is included in the regression model in order to control for the bias in the dependent variable. *A priori* expectations are that years married will be inversely related to survival rates, with the relationship expected to be strongest among women who have been married for only a few years.

Parents' Education: Education is believed to affect mortality in several ways. Educated parents should be able to provide their children with a more nutritious diet at any income level, and they should be able to protect their children from many types of infectious diseases. Educated parents are also more likely to utilize modern as compared to traditional health services. The survey includes information on the education of both parents. *A priori* expectations are that parents' education will be positively related to survival rates and that mother's education should be more important in this regard than father's.[4]

Income: It is widely believed that rising income levels contribute to reducing mortality levels. The limited number of existing econometric analyses of mortality differentials have produced evidence consistent with this view.[5]

[3]The extent of the bias is not great, however, for two reasons. First, only about 20 per cent of the recorded live births were under three years of age at the time of the survey. Second, most infant and pre-school deaths occur in the first year of life (in fact, about one-third of all deaths under three years occur in the first month of life).

[4]However, Afzal, Khan and Chaudhry (1976) report a regression determining child mortality in which the estimated coefficient of father's education is larger in magnitude than that of mother's education (-.011 compared to -.002, both significant at the .10 level).

[5]See, for example, Adelman (1963), Afzal, Khan and Chaudhry (1976), and Anker and Knowles (1980).

It is not income, however, which reduces mortality; it is rather the things which income buys. From the standpoint of mortality reduction, these include more and better quality food, housing, and medical care. The role of these variables, as well as that of income, are examined in the present study. It is not income, however, which reudces mortality; it is rather the things a good proxy for the variables which are believed to affect mortality directly. It is also interesting to see whether any residual contribution is made by income when the effects of these other variables have been properly accounted for.

As an indirect measure of the constraints imposed on the consumption of food, housing and medical care, the level of family income should be adjusted by some measure of family size. Direct measures of consumption should be similarly adjusted. In the present study, the number of adults alone is used as a measure of family size. Children are thus implicitly given a zero weight in calculating family size in order to avoid the possibility that a spurious statistical relationship might be built into the relationship between child mortality and income per family member. Such a relationship would be present if relatively high per capita income occurs among smaller families due to high mortality among their children.[6]

It is important to note that the income data in the survey refer to only one period (i.e., the month of the survey). However, it is *past* levels of income, rather than current levels, which are the appropriate determinants of *past* mortality. Current income levels may be observed to have an effect on past mortality only to the extent that the current income of families is correlated with their past income levels. In other words, the cross-section differentials in mortality rates between households will be explained by the current income differentials between the households only to the extent that current income differentials are a good proxy for past income differentials. Since it is unlikely that current differentials are a perfect proxy for past differentials (this would imply at a minimum that no changes in the relative income rankings of the sample households had occurred over time), one may expect some downward bias in the coefficients relating to income per adult in the regression analysis. The same comments would, of course, apply to direct measures of food consumption, housing and medical care.

Finally, it should be noted that the *cross-section* differentials in income and consumption do not reflect the effects of increases in the general level

[6] For example, consider two families both receiving income of Rs. 600 per month. Suppose that both families consist of 2 adults and have each had 8 children but that one family has lost half its children while the other family's children have all survived. Considering children to be equivalent to one-half an adult, the income per equivalent adult of the first family would be Rs. 150 per month while that of the second family would be Rs. 100 per month. A positive relationship between mortality and income per family member would therefore be built-in.

of these variables *over time.* To the extent that such increases have had some effect in reducing mortality this effect should be manifested in higher survival rates for more recent births as compared to those born in the past. One would expect, therefore, that part of the effect of general rises in living standards will be reflected in the coefficient of years married, since women who have been married longer will tend to have had a higher proportion of their children born several years in the past. These considerations lead one again to expect an inverse relationship between survival rates and the number of years married.

Mother's Nutrition: It is reasonable to expect that the probability of a child's survival is affected by its mother's nutritional status. Mother's nutritional status may affect not only a child's weight at birth and the supply of certain nutrients (e.g., iron) but also the quantity and quality of her breast milk and the length of time she is able to lactate.[7] The survey contains data on both family food expenditures and family income, either of which can serve as a proxy for mother's nutritional status. In addition the survey also recorded information about any changes in the diet of mothers during pregnancy (i.e., whether any foods were added, the type of any foods added, and whether vitamins or minerals were added).[8] *A priori* expectations are that both income (or food expenditures) per adult and dietary supplements during pregnancy are positively related to survival rates.

Infant Feeding Practices: In addition to the mother's nutritional status, the actual practices followed with respect to the feeding of infants are also expected to be an important factor affecting survival rates. Breast milk has long been regarded as an ideal source of nutrition for infants.[9] Breast milk not only provides essential nutrients but also provides infants with the mother's immunity against many types of diseases. The cessation of breast-feeding exposes the infant to a wide range of health risks.

While breast feeding during the first two years of life, when the probability of death is highest, is very beneficial, first as the sole source of nutrients and later as a useful supplement to solid feeds, prolonged breastfeeding in the absence of solid foods may lead eventually to a worsened nutritional

[7]During the course of the survey the breast milk of 20 mothers with infants suffering from severe protein calorie malnutrition was collected with the aid of a mechanical breast pump. According to Dr. Rahimtoola and her colleagues (1978, Appendix A) these women were all in "poor health and varying degrees of anaemia." The quantity of breast milk was low (between 24 ml. and 120 ml. in a 24-hour period) and the milk (which was sent to Sweden for analysis) was found to be deficient in essential amino acids.

[8]In the survey, 254 mothers (34.5%) reported the addition of various types of foods during pregnancy (mostly combinations of milk, meat and eggs), while 129 mothers (17.4%) reported adding vitamins and minerals.

[9]See Berg (1973), Chapter 7.

status and may render a child more susceptible to illness. These considerations suggest that the age at which solid foods are added may also play a role in determining a child's probability of survival.

The survey provides information on both the length of breastfeeding and the age at which solid foods were added for all children under five years of age at the time of the survey. These data are available, however, only for surviving children.[10] Average values of these variables were utilized in the analysis. In the case of length of breastfeeding, the average value was limited to children who had completed their period of breastfeeding at the time of the survey, while in the case of the age at which solid foods were added, the average included only children who were eating solid foods at the time of the survey. Information on the frequency distribution of these two variables is provided below in Table 4.

Medical Services: There is very little which an individual family can do to affect the type of medical services available to it (except be prepared to travel a longer distance). However, families in the same environment differ widely in their use of available medical services. Such differences are probably closely related to differences in education and income. Nevertheless, it is interesting to investigate the consequences for infant and pre-school survival rates of decisions taken, for whatever reasons, by individual households.

In the survey area the availability of medical facilities was limited to one KMC dispensary (which did not include a maternal/child health unit), about twelve private physicians (some of whom were probably quacks), and various practitioners of traditional medicine *(dais,* or midwives without any formal training, are the traditional practitioners providing ante-natal, delivery and post-natal care). For serious problems, the residents of the survey area could go to a public hospital located a few miles away (e.g., Jinnah Postgraduate Medical Centre).

The patterns of utilization of various types of medical care by the families in the survey are summarized in Table 5. Traditional *dais* were by far the most popular source of medical care. Interestingly, the majority of women in the sample (67.3 percent) received some form of ante-natal care, although for a majority of these women (69.4 percent) this was provided by traditional *dais.* Although it is not clear what type of ante-natal care is received

[10] However, data on length of breastfeeding and the age at which solid foods were added would be of only limited use in the case of children who had passed away. If deceased children were included there would be a built-in statistical relationship between these variables and survival rates since, for example, a child who was breastfed for two years would obviously have survived that long while infants dying at an age of only a few months would be recorded as having been breastfed for only that short time.

TABLE 4

Average length of completed breastfeeding and average age at which solid foods given (limited to surviving children ages 5 and under at time of the survey)

Average age of children	Termination of breastfeeding		Solid foods added	
0–6 months	53	.091	42	.068
6–12 months	68	.117	236	.383
12–24 months	258	.445	305	.491
24+months	201	.347	36	.058
	580[a]	1.000	621[a]	1.000

[a]As is evident from the reduced number of cases in Table 4, the variables referring to average length of breastfeeding and average age at which solid foods were given could not be computed in the case of families not having at least one child who had completed breastfeeding or, in a smaller number of cases, not having any children who had yet been given solid foods.

TABLE 5

Patterns of medical care utilization among sample women

Type of care	Ante-natal care		Delivery care		Post-natal care	
None	241	.328	0	.000	642	.877
Dais	343	.467	602	.822	65	.089
Public hospital/MCH Centre	94	.128	82	.112	14	.019
Private practitioner	57	.078	48	.066	11	.015
TOTAL	735	1.000	732	1.000	732	1.000

from *dais*, it is probably limited to advice to take various dietary supplements, both food and vitamins.

From the standpoint of investigating the effect of different types of medical care on survival rates, a potential simultaneous equations problem arises from the fact that mothers experiencing difficulties in pregnancy or childbirth, or mothers with very ill children, may seek out the more modern forms of medical care. A similar problem would arise if one were to use expenditures on medical care as an independent variable. There is no way to avoid this problem completely with the data available. However, one would expect a simultaneous equations problem of this type to bias the coefficients of variables referring to modern medical care in a negative direction. Preliminary analysis revealed that this problem was most serious in the case of delivery and post-natal care, the modern components of which exhibited consistenly negative coefficients (some even statistically significant). The medical care variables utilized in the present study are therefore limited to ante-natal care and medical expenditures per adult (a three-month average). *A priori* expectations are that both variables will be positively related to survival rates and that ante-natal care received from modern practitioners is more beneficial than care received from traditional practitioners.

Housing and Sanitation: There are several characteristics of low-income housing which may affect infant and child mortality. As already noted, most of the houses in the survey area consist of only one room, with an open courtyard where cooking is done.[11] Since average family size in the sample is 6.6 persons, such housing conditions can only be described as very crowded. Crowded living conditions increase the probability of infection for all family members and should therefore contribute to higher mortality rates among infants and small children.

Only 36 houses in the survey (4.9 percent) reported having private water taps, while another 25 households (3.4 percent) reported sharing a private tap with one or more families. The balance of families in the survey (91.7 percent) reported the use of community taps. Although there is no reason to believe that the *quality* of water differs by source (all water is supplied by the Karachi Development Authority), one would expect the *quantity* of water to vary by source. Families without a private source of water typically either pay for water to be delivered to their home or stand in line to get water at fixed times from the community tap. For families without a private tap, water is usually not readily available for bathing and cleaning purposes, and the resulting poor hygiene may affect survival rates.

[11] 457 houses (62.1%) in the survey consisted of only one room, while another 216 houses (29.3%) consisted of only two rooms.

Only 47 families in the survey (6.4 percent) reported having flush toilets. Most of the other families (89.1 percent) reported having private pit latrines. It is likely that families with flush toilets experience reduced exposure to infection. Even where private pit latrines are used, the latrine itself is a possible source of contamination, especially where it is not properly protected from flies.

Most families in the survey (61.6 percent) reported that they dispose of refuse by throwing it into the street. Of the remaining families, 204 (27.7 percent) reported disposing of refuse through the municipal refuse disposal service, while the other 79 families (10.7 percent) reported dumping their refuse in the open (rather than in the streets). It is likely that families dumping refuse in the street expose themselves and their children to a greater risk of disease, if only because the refuse attracts flies. *A priori* expectations, therefore, are that survival rates are lower for families who dump their refuse in the streets.

Macro-Level Determinants of Mortality

The macro-level variables affecting mortality include disease prevalence, the state of medical technology, the availability of medical services, and public health measures such as mass immunization and the provision of safe water. Due to the nature of the data set used in the present study (a cross-section sample from one locality), no significant variation in these macro characteristics can be expected to exist among the families in the sample. Nevertheless, despite the absence of variation at a moment in time, most of these factors have probably changed considerably over time.[12] Thus, for example, there has been significant progress in controlling malaria during the past thirty years, and mass immunization compaigns have been directed against certain diseases (e.g., tuberculosis, small pox). The quality of municipal water has also undoubtedly improved. During this period, the state of medical technology has also advanced considerably, and many previously serious diseases can now be controlled by the simple administration of drugs. It is also likely that the availability in Karachi of both private and public medical services has increased dramatically in recent years, although the supply of such services may not have increased much relative to the effective demand.

[12]Since some of the families in the survey are likely to have come from other parts of the Indian subcontinent, some variation between families in macro characteristics may have been present in the past. Unfortunately, no data on geographical origins were available in the survey.

All of these considerations lead one to suspect that the various macro-level factors affecting mortality have changed significantly over time, at least in Karachi, in such a way as to raise significantly the probability of survival over what it was only thirty years ago. If this is true, then it implies that children born several years in the past should exhibit lower survival rates than children born in recent years. Unfortunately, no data are available in the survey on the date of birth. However, years married should once again serve as a good proxy–this time, for the average year of birth of children born to a particular woman–and the expectation that survival probabilities have increased significantly over time provides yet another basis for expecting a strong inverse relationship between survial rates and years married.

IV. Regression Analysis

The following regression model is utilized:

$$\text{SURVIVE3} = B_o + B_1 \text{ LITH} + B_2 \text{ LITW} + B_3 \text{ SECW} + B_4 \text{ LYERS} + B_5 \text{ LINC} + B_6 \text{ DIET} + B_7 \text{ MILK} + B_8 \text{ MILK2} + B_9 \text{ SOLID} + B_{10} \text{ WATER} + B_{11} \text{ TOILET} + B_{12} \text{ REFUSE} + B_{13} \text{ PRIVATE} + B_{14} \text{ PUBLIC} + B_{15} \text{ DAI} + u$$

SURVIVE3 Proportion of children ever born surviving to the age of three years (the dependent variable).

LITH binary variable: 1 if the father is literate, 0 otherwise.

LITW binary variable: 1 if the mother is literate or has completed primary (but not secondary) schooling, 0 otherwise.

SECW binary variable: 1 if the mother has cmpleted secondary schooling, 0 otherwise.

LYEARS number of years married (natural log)

LINC family income per adult, in thousands of Rupees per month (natural log).

DIET binary variable: 1 if the mother added some type of food or vitamins and minerals to her diet during pregnancy, 0 otherwise.

MILK (MILK2) average length of breastfeeding (in months) among surviving children (MILK2 = milk squared).

SOLID average age (in months) at which solid foods given to surviving children.

WATER binary variable: 1 if the family has a private water tap, 0 otherwise.

TOILET binary variable: 1 if the family has a flush toilet, 0 otherwise.

REFUSE binary variable: 1 if the family uses the municipal service to dispose of refuse (or dumps it in the open), 0 if family throws refuse in the street.

PRIVATE binary variable: 1 if the mother used a private physician for ante-natal care, 0 otherwise.

PUBLIC binary variable: 1 if the mother used a public hospital or maternal/child health centre for ante-natal care, 0 otherwise.

DAI binary variable: 1 if the mother used a traditional *dai* for ante-natal care, 0 otherwise (excluded class: no ante-natal care).

Three other variables were also employed in some of the regressions as substitutes for one or more of the above variables. They were:

LFOOD family food expenditures per adult, in hundreds of Rupees per week (natural log).

MED family expenditures on medical care per adult, in hundreds of rupees per month.

ROOMS number of rooms per adult in the family's dwelling.

The means and standard deviations of all variables are provided in Appendix A, while the correlation matrix of the independent variables is provided in Appendix B.

The functional form of the regression model was selected after some initial experimentation. For example, in the case of the average length of breastfeeding (MILK), the following functional forms were tried:

(i) a simple linear form (MILK);
(ii) a linear and quadratic form (MILK, MILK2);
(iii) a log form (LMILK); and
(iv) a set of binary variables referring to varying lengths of breastfeeding (BMILK1, BMILK2, BMILK3).

Similar forms were tried in the case of the other continuous variables. The functional form selected was in each case the one which yielded the highest adjusted coefficient of multiple determination ($\bar{R}^2$).[13]

The final functional form which emerged from this process accords fairly well with prior expectations. Thus, a log form for family income (LINC) seems reasonable because it implies that the marginal effect of income on survival rates declines with the level of income. One would expect the satisfaction of basic needs to have a greater effect on survival rates than the consumption of luxuries.[14] The quadratic relationship between survival rates and MILK may reflect a failure on the part of some women to supplement breastmilk with necessary solid food. Reasons have already been given for expecting the inverse relationship between survival rates and years married to be strongest among recently married women.[15] A log form for years married (LYEARS) is therefore also consistent with theoretical expectations.

The fact that the dependent variable (SURVIVE3) is bounded between zero and one creates certain econometric problems. First, a problem may arise from the fact that the estimated regression function, being linear, can yield predicted values which lie outside the zero-one bounds. Even if this problem does not arise in practice one might prefer on prior grounds a functional form which approached its limits asymptotically rather than linearly. Both probit and logit models have this property.[16] Some experimentation

[13] It is recognized that this is only one of several possible criteria for selecting a functional form; in particular, no examination of the residuals was carried out during the selection process because of the computation cost this would have entailed with 700 observations. In the case of the binary (dummy) variables, two or more adjacent categories were combined into one category whenever this produced a higher $\bar{R}^2$. Thus, for example, husband's education was finally represented by only two clssifications (literate, illiterate) whereas wife's education was represented by 3 classifications (secondary and above, literate/primary, and illiterate).

[14] One interesting implication of the nonlinear effect of income on mortality is that greater income inequality leads to higher mortality.

[15] It may be recalled that recently married women are more likely to have a higher proportion of children who have not yet completed three years of age and who have therefore not been exposed to the same degree of mortality risk as have the children of women married for a longer period.

[16] For a discussion of these models see Theil (1971).

with a modified logit model was carried out (see Appendix C). The results, however, were quite similar to those obtained with the regression model. Accordingly, it was decided to confine the discussion in the text to the more familiar and easier to interpret regression model.

A second econometric problem may arise from the sample proportion nature of the dependent variable. This suggests that its variance may be systematically related to the number of live births and that heteroskedasticity may therefore be present.[17] For this reason conventional tests for the presence of heteroskedasticity were performed.[18] No evidence of heteroskedasticity was encountered, however, and this appeared to justify the use of ordinary least-squares (OLS) as the estimation technique.[19,20]

The estimated regressions are reported in Table 6. There are five regressions in all. The first two regressions (columns 1, 2) are limited to a "core list" of variables: husband's and wife's education (LITH, LITW, SECW), income (LINC) and years married (LYEARS). The first regression was estimated with the full set of observations containing information on the "core list" of variables (N = 686), whereas the second regression (and all other regressions reported in Table 6) was estimated with a smaller sample consisting of observation with information on all of the independent variable (N = 505).[21] The fact that the estimated regressions in columns one and two are quite similar suggests that the results are not too sensitive to the use of the smaller sample.

The third regression (column 3) includes, in addition to the core variables, variables representing the mother's nutritional status (DIET) and infant feeding practices (MILK, MILK2, SOLID), while the fourth regression (column 4) adds variables referring to housing and sanitation (WATER, TOILET, REFUSE) and the type of medical care utilized (PRIVATE, PUBLIC, DAI).

[17] If P is the probability of survival to age three, then the variance of SURVIVE3 should be equal to P (1-P)/N, where N (analogous to sample size) is the number of live births.

[18] The tests included the Goldfeld-Quandt test and a regression of the absolute values of the least-squares residuals on the number of live births (N). For a discussion of these tests, see Johnston (1972, pp. 218-21).

[19] The absence of heteroskedasticity in this case may be due to the fact that the probability of survival itself is not a constant but rather declines with the number of live births (due to the correlation between the average year of birth and the number of live births). In other words, both the numerator [P(1-P)] and the denominator (N) in the expression for the variance increase with the number of live births (N).

[20] The fact that the dependent variable is limited implies that the assumption of strict normality cannot be made in the present model. Nevertheless, it is still likely that conventional t-tests and F-tests are approximately correct. See Malinvaud (1966, pp. 251-54).

[21] Most of the observations with missing data involved women who had not yet completed the breastfeeding of at least one child and for whom, therefore, the variable MILK was undefined.

TABLE 6

Regression analysis of factors determining survival rates to age three

Independent variables	Estimated Regression Coefficients (t– statistics)				
	(1)	(2)	(3)	(4)	(5)
LITH	.0271	.0272	.0267	.0283	.0288
	(1.78)*	(1.60)	(1.57)	(1.66)*	(1.69)*
LITW	.0085	.0084	.0103	.0025	.0030
	(0.38)	(0.35)	(0.43)	(0.10)	(0.12)
SECW	.0547	.0434	.0703	.0464	.0595
	(1.72)*	(1.14)	(1.82)*	(1.10)	(1.42)
LYEARS	-.0630	-.0667	-.0750	-.0727	-.0791
	(5.72)***	(4.64)***	(5.15)***	(4.96)***	(5.29)***
LINC	.0554	.0650	.0613	.0585	
	(4.43)***	(4.67)***	(4.41)***	(4.19)***	
LFOOD					.0693
					(4.23)***
DIET			.0061	-.0066	-.0053
			(0.35)	(0.37)	(0.30)
MILK			.0130	.0143	.0147
			(3.41)***	(3.75)***	(3.86)***
MILK2			.0003	-.0003	-.0003
			(2.69)***	(3.09)***	(3.19)***
SOLID			.0003	.0004	.0000
			(0.20)	(0.31)	(0.01)
WATER				-.0101	-.0045
				(0.23)	(0.10)
TOILET				.0073	.0054
				(0.19)	(0.14)
REFUSE				.0381	.0395
				(2.19)**	(2.28)**
PRIVATE				.0621	.0486
				(1.76)*	(1.37)
PUBLIC				.0071	.0001
				(0.26)	(0.01)
DAI				.0344	.0213
				(1.84)*	(1.13)
Constant	1.0973	1.1373	1.0185	0.9683	0.9708
$\bar{R}^2$	.079	.081	.104	.112	.113
F	12.69***	9.87***	7.47***	5.23***	5.26***
N	686	505	505	505	505

***Statistically significant at the .01 level.
**Statistically significant at the .05 level.
*Statistically significant at the .10 level.

The last regression in Table 6 (column 5) utilizes the full list of explanatory variables (as in column 4) but with the log of food expenditures per adult (LFOOD) substituted for the log of income per adult (LINC). The successive "stepping-in" of the explanatory variables in groups allows one to see the extent to which the effects of income and education on survival rates operate through such proximate determinants as infant feeding practices, housing, and type of medical services utilized.

The results in Table 6 suggest that the number of years married (LYEARS), family income (LINC), and the length of breastfeeding (MILK, MILK2) are the principal determinants of infant and pre-school survival rates. All three relationships are non-linear and all three are statistically significant at the .01 level in all regressions in which they are included. The estimated coefficient of LYEARS in the core regression (column 1) implies that SURVIVE3 is lower by .101 for women married twenty-five years as compared to those married only five years. Since this difference corresponds to a change in average life expectancy at birth (LX) of about 12.7 years,[22] these results provide strong support for the view that mortality rates have been declining rapidly over time.[23]

The estimated effect of LINC on survival rates is similarly strong, implying (column 1) that a doubling of INC at the sample mean (.167 thousand rupees per month) increases SURVIVE3 by .038 (or raises LX by about five years). Moreover, the estimated coefficient of LINC remains statistically significant (and almost as large in magnitude) even after the inclusion of variables referring to infant feeding practices, housing and medical care (columns 3, 4). This suggests that income does not affect mortality primarily through these intermediate variables. On the other hand, when LFOOD is substituted for LINC (column 5), the R^2 actually increases, suggesting that it is family food expenditures which affects survival rates (through its effect on the nutritional status of mothers and children) rather than income *per se*.[24,25]

The effect of adding MILK and MILK2 to the list of independent variables (column 3) is to raise somewhat the magnitude of the negative coeffi-

[22]In interpreting the regression coefficients it is helpful to keep in mind that, according to West Model Life Tables, a change in SURVIVE3 of .04 corresponds to a change in LX of about five years.

[23]It is necessary to qualify this conclusion somewhat, however, in light of the fact that part of the observed effect of LYEARS is likely to be due to the fact that a higher proportion of the children of women in the sample who have been married only a few years have not yet completed three years of life.

[24]When both LFOOD and LINC are included (unreported regression), high collinearity between the two variables causes both coefficients to become statistically insignificant.

[25]The variable MED and ROOMS are statistically insignificant when they are included in any regression containing LINC or LFOOD (unreported regressions). Moreover, their inclusion has little effect on either the magnitudes or significance levels of the coefficients of LINC or LFOOD.

cient of LYEARS. This change is due to a positive correlation between MILK and LYEARS (r = .25). The coefficient of LINC is changed very little. The estimated coefficients of MILK and MILK2, which are statistically significant at the .01 level in all regressions, indicate that survival rates reach a maximum when breastfeeding is carried on for 22 months (as compared to the sample mean of 17.5 months). The magnitude of the estimated effect is quite large: the results in column 4 suggest, for example, that breastfeeding for 22 months as compared to only six months increases SURVIVE3 by .094 (an increase in LX of approximately 12 years).

While the results of the present study offer strong evidence of the effectiveness of prolonged breastfeeding in reducing infant and child mortality, there is no reason to belive that breastfeeding habits have been changing over time in such a way as to have contributed to the recent declines in mortality. In fact, the evidence seems to imply just the opposite. A regression of MILK (i.e., the average length of breastfeeding) on LITH, LITW, SECW, DIET, INC, SOLID, PRIVATE, PUBLIC, DAI, AGEMARRY (mother's age at marriage), and AGEWIFE (mother's current age) produced the following results (with t-statistics in parentheses):

$$
\begin{aligned}
\text{MILK} = 13.86 &- \underset{(0.75)}{.517\ \text{LITH}} - \underset{(0.33)}{.317\ \text{LITW}} - \underset{(1.69)^{*}}{2.710\ \text{SECW}} - \underset{(3.88)^{***}}{.464\ \text{AGEMARRY}} \\
&+ \underset{(4.31)^{***}}{.251\ \text{AGEWIFE}} + \underset{(2.20)^{**}}{6.848\ \text{INC}} + \underset{(2.43)^{**}}{1.703\ \text{DIET}} + \underset{(2.89)^{***}}{.156\ \text{SOLID}} \\
&- \underset{(0.26)}{.361\ \text{PRIVATE}} + \underset{(0.50)}{.540\ \text{PUBLIC}} + \underset{(1.30)}{.979\ \text{DAI}}
\end{aligned}
$$

$R^2 = .12$
N = 531

These results show that younger women tend to breastfeed their children for shorter periods than older women (about 2.5 months for each ten years' difference in age), and that women who marry at a later age tend to breastfeed for shorter periods (almost one-half month less for each additional year of delay before marriage).[26] Since it is likely that average age at marriage has been rising,[27] both of these results suggest that the average length of

[26]While this result suggests that prolonged breastfeeding is a traditional practice rather than a modern practice, it may also reflect the desire of women who marry later to catch up in their fertility with that of other women. It is well known that prolonged breastfeeding lengthens the period of infertility following a birth. See Bongaarts (1978) and Berg (1973), Chapter 7.

[27]See Afzal and Iftikhar (1974) and Afzal, Khan and Chaudhry (1976).

breastfeeding has been falling over time. The other results do not provide any basis for altering this conclusion. Thus, the estimated coefficients of the education variables (LITH, LITW, SECW) are negative, while that of SOLID is significantly positive (indicating that earlier introduction of solid foods shortens the average length of breastfeeding). The other significant variables (INC, DIET), while pointing to a positive relationship between modernization and length of breastfeeding, exhibit coefficients too small to offset the negative effects of the other variables (e.g., a doubling of INC at the sample mean increase MILK by only about one month).

The regression results in Table 6 do not point to parents' education as important determinants of survival rates. These results are surprising, particularly with respect to mother's education. The estimated coefficients of both LITW and SECW are positive but the former is insignificant in all reported regressions, while the latter is statistically significant in only two out of five regressions (and then only at the .10 level of significance). Nevertheless, the estimated coefficient of SECW is relatively large in magnitude (ranging from .04 to .07), suggesting that secondary education on the part of the wife is associated with an increase in LX of between five and nine years. The magnitude of SECW's coefficient is quite sensitive, however, to the inclusion of certain of the other explanatory variables. Thus, it increases in magnitude when MILK and MILK2 are added (column 3), while it falls in magnitude with the addition of variables referring to housing and medical care (columns 4, 5). Given that a still very small proportion of Pakistan's women have received secondary education and that education at less than a secondary level is seen to have almost no impact on survival rates, the results of the present study suggest that rising levels of female education have probably played a relatively minor role in the recently observed declines in mortality.

A similar conclusion is suggested with respect to husband's education. Although the estimated coefficient of LITH is consistently positive and is statistically significant (at the .10 level) in three regressions, the magnitude of the estimated coefficient is rather small (less than .03). These results together with the fact that recent gains in male literacy have been something less than dramatic in Pakistan, suggest that rising levels of male education have probably also played a relatively minor role in reducing mortality rates.[28]

[28] To consider the possibility that the statistical insignificance of the education variables was due to collinearity between husband's and wife's levels of education, regressions were also estimated in which either husband's or wife's education was deleted (unreported regressions). No significant differences were observed in the coefficients of wife's (husband's) education when husband's (wife's) education was deleted.

The results indicate that REFUSE has a statistically significant positive effect on survival rates while WATER and TOILET do not. Households which dispose of refuse by some method other than by throwing it into the street are seen to have a higher survival rate by about .04 (or an increase in LX of five years). This result is somewhat surprising since one would not expect such a variable to have any effect on an individual household. It may be that disposing of refuse in a manner other than throwing it in the street is simply a proxy for other individual practices conducive to good health. Nevertheless, the results as they stand suggest that more research is needed into the possible role which improved sanitation may play in reducing the incidence of disease and mortality.

The statistical insignificance of WATER is not very surprising, given the fact that all households in the sample are using the same basic water supply (i.e. municipal water). The statistical insignificance of TOILET suggests that the type of toilet facility utilized by a given household does not significantly affect its own health status. These results do not rule out, however, the possible existence of beneficial community-level effects from the use of proper toilet facilities.

The results indicate that ante-natal care, whether it is received from private physicians or from traditional *dais,* has a positive effect on survival rates. The fact that the estimated effect of ante-natal care received from public hospitals (PUBLIC), while positive, is very small suggests either that public hospitals fail to provide care which is even up to traditional standards or, more probably, that public hospitals get cases more frequently involving serious complications. It is interesting that ante-natal care received from traditional dais (DAI) is seen to have a significant positive effect on survival rates when compared to no care at all. This finding supports the view that some minimal form of medical care is better than none at all. Nevertheless, the evidence for this relationship is quite weak. When LFOOD is substituted for LINC (column 5), the estimated coefficients of both PRIVATE and DAI become statistically insignificant at even the .10 level.

The estimated coefficients of DIET and SOLID are statistically insignificant in all of the reported regressions. The insignificance of DIET suggests that changes in the mother's diet during pregnancy do not have any appreciable effect on a child's probability of survival. This conclusion requires some qualification, however. DIET was seen above to be a statistically significant determinant of the length of breastfeeding (MILK): women who reported altering their diets during pregnancy had an average length of breastfeeding 1.7 months longer than women reporting no change in diet. At the sample mean average length of breastfeeding (17.5 months), this implies an *indirect effect* on survival rates of .0065 (or an increase of about .8 years in LX).

The statistical insignificance of SOLID in Table 6 may reflect the presence of two offsetting factors: while, on the one hand, giving solid food at an early age may improve an infant's nutritional status, it may, on the other hand, expose him to an appreciably greater risk of infection from contaminated foods. Despite the absence of any significant *direct effect* of *SOLID* on survival rates, this variable may, as in the case of DIET, have a significant indirect effect via its effect on the average length of breastfeeding. Thus, women who add solid foods to their children's diet at a younger age are seen to breastfeed for a significantly shorter period. Women, for example, who give solid foods at only 6 months of age instead of at 14.5 months (the sample mean), tend to breastfeed about 1.3 months less. This in turn implies (at the sample mean length of breastfeeding) a reduction in survival rates of .005 (or a decrease in LX of about .6 years). Thus, we have the unexpected result that an improvement in infant feeding practices has the unfortunate *indirect effect* of lowering survival rates. This relationship probably reflects the low standards of hygeine which predominate in poverty groups and the tendency of the sample women to cease breastfeeding soon after solid foods are given.

V. Conclusions

This paper has analyzed mortality differentials among infants and preschool children from one low-income area of Karachi. The results of the analysis indicate that rising levels of family income have played a major role in the dramatic decline in mortality rates observed in Pakistan since 1950. If one assumes that real family income has approximately doubled during this period, then the estimates obtained in the present study suggest that this factor alone would account for an increase in average life expectancy at birth (LX) of somewhere between 4.8 and 5.6 years (about 25% of the total estimated increase in LX during this period). Since the cross-section effect of income on mortality probably understates the true effect (as argued above), these results make a strong case for those who believe that socio-economic change has contributed significantly to the rapid post-war declines in mortality rates.

Other interesting results include the finding that prolonged breastfeeding can significantly increase a child's probability of survival, while parents' education apparently has little effect on mortality. In the case of breastfeeding, the regression results indicate that breastfeeding for up to two years can increase a child's probability of survival by almost 10 percent. Other results of the study suggest, however, that the average length of breastfeeding has been falling in recent years, so it is unlikely that the relationship between length of breastfeeding and probability of survival offers any explanation for the recent declines in mortality rates. The finding that education does not have a strong impact on mortality is somewhat surprising.[29] Nevertheless, given the still very high rates of illiteracy prevailing in Pakistan, it would be difficult to argue that the rapid declines in mortality which Pakistan has experienced in recent years are the result of progress in education.

A number of variables were found to be insignificantly related to mortality. These include: changes in the mother's diet during pregnancy, the age at which solid foods are first given to children, the use of a private water tap or flush toilet, and ante-natal care received at a public hospital. Some evidence of a relationship between survival rates and the method of refuse disposal was also encountered. However, since one would not expect the practices of individual families in this regard to have any significant impact on individual survival rates, further investigation of this relationship seems warranted.

One should be very cautious about drawing policy conclusions from a study which is so limited, as is the present one, in its geographical and socio-economic coverage. Nevertheless, the present study's findings do point to a number of policy-relevant conclusions, however tentative. First, one may expect mortality rates to continue declining in Pakistan as income levels rise. This suggests in turn an acceleration of Pakistan's already high rate of population growth, unless the Family Planning Program is more successful in lowering fertility levels sufficiently to offset the expected declines in mortality. Accelerated population growth will make it more difficult to achieve self-sufficiency in food production and to raise the population's currently very low standards of education and health. It is also likely to contribute to higher rates of rural-urban migration and unemployment.

If, despite its implications for accelerating the rate of population growth, a more rapid rate of decline in mortality is desired as a policy objective, the present study's findings suggest that appropriate policies to achieve such a result should include income redistribution to the poor and encouraging extended periods of breastfeeding (two years or more). The results of the present study do not suggest that increased education would contribute much in this regard (apart from secondary education among females), at least until education standards for the urban poor improve significantly over their current levels.

[29]It is interesting to note, however, that Afzal, Khan and Chaudhry (1976) obtained similar results in a study of low-income families from Lahore.

Appendix A

Means and standard deviations of the variables

(N = 505)

Variable	Mean	Standard deviation
SURVIVE3	0.878	0.190
LITH	0.572	0.495
LITW	0.141	0.348
SECW	0.052	0.221
YEARS	10.986	6.524
INC	0.167	0.109
FOOD	0.291	0.169
DIET	0.390	0.488
MILK	17.280	7.840
SOLID	14.691	6.122
WATER	0.046	0.209
TOILET	0.061	0.240
REFUSE	0.388	0.488
PRIVATE	0.077	0.267
PUBLIC	0.131	0.337
DAI	0.455	0.499

Appendix B

Correlation matrix of the independent variables

(N = 505)

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
LITH	(1)	1.00	.14	.18	-.14	.06	.03	.22	-.06	-.05	.11	.07	.08	.12	.10	-.02
LITW	(2)		1.00	-.09	-.06	.09	.06	.18	-.03	-.11	.08	.09	.17	.12	.11	-.11
SECW	(3)			1.00	-.15	.05	-.04	.14	-.13	-.11	.29	.28	.15	.30	.07	-.07
LYEARS	(4)				1.00	.17	.26	-.04	.25	-.01	-.02	-.02	.04	-.14	-.07	-.00
LINC	(5)					1.00	.82	.14	.11	-.07	.10	.09	.10	.08	-.10	.01
LFOOD	(6)						1.00	.12	.14	.02	.04	.07	.06	.06	-.13	.16
DIET	(7)							1.00	.07	-.11	.08	.10	.20	.21	.04	.04
MILK	(8)								1.00	.14	-.02	-.05	.03	-.09	-.05	.11
SOLID	(9)									1.00	-.03	-.05	-.16	-.17	-.12	.21
WATER	(10)										1.00	.42	.16	.19	-.00	-.07
TOILET	(11)											1.00	.17	.14	-.05	.01
REFUSE	(12)												1.00	.13	.06	-.05
PRIVATE	(13)													1.00	-.11	-.26
PUBLIC	(14)														1.00	-.35
DAI	(15)															1.00

Appendix C

Modified Logit Analysis

One problem with the use of a linear regression model in the present study is the fact that the dependent variable is inherently bounded between zero and one while the regression function is inherently unbounded. This implies that an estimated regression function can yield predicted values of the dependent variable which lie outside the upper or lower bounds of the dependent variable. The linear functional form may also be inappropriate for sample values close to the upper and lower bounds since it implies that the effect of income on the probability of survival is the same regardless of how close the probability is to its upper or lower bound. *A priori* a given functional form would be more appealing if it implied that the effect of an independent variable on the dependent variable were to diminish in magnitude as the dependent variable approached its upper or lower bound (i.e., the function should approach its limits asymptotically).

These considerations suggest some transformation of the dependent variable such that the functional relationship between it and the explanatory variables becomes inherently bounded. In principle there are many possible transformations of this type. One which has been wiaely used in the biological sciences is the logit transformation:[30]

$$\log \frac{P_i}{1-P_i} = \beta_o + \beta \log X_i + \epsilon_i \qquad (1)$$

where P is a variable bounded between zero and one (typically interpreted as the probability of some occurrence) so that the ratio P/(1-P) can be interpreted as the "odds" that some event will occur. The natural logarithm of the "odds" then becomes a variable bounded between negative infinity and positive infinity with a value of zero when the "odds" are even (i.e. when P = 0.5). The terms on the right-hand side of equation (1) include a vector of independent variables (X),[31] a set of constants (β_0, β), and a random disturbance term (ϵ_i).

[30]See, for example, Theil (1971), p.632.

[31]The independent variables may or may not be expressed as natural logarithms. See, for example, Kmenta (1971), pp. 461-462 and Goldfeld and Quandt (1972), p. 127.

Equation (1) is called a logit model and its relationship to the ligistic curve may be seen by solving for P_i:

$$P_i = \frac{1}{1+e^{-\beta_o-\beta \log x_i-\epsilon_i}} = \frac{1}{1+e^{\beta_o-\epsilon_i} X_i^{-\beta}} \qquad (2)$$

Equation (2) shows that the logistic transformation of P_i achieves the necessary results: P_i is bounded between zero and one, and it approaches its limits asymptotically.

In order to estimate the unknown parameters (β_o,β) it is usually convenient to group the data according to ascending values of one of the explanatory variables and then apply weighted least-squares (GLS) to equation (1).[32] In the present case, the dependent variable (SURVIVE3) already represents a grouping of observations on individual births by families. Nevertheless, a problem would arise in applying least-squares directly to equation (1) with SURVIVE3 substituted for P_i. The fact that the sample values of SURVIVE3 are frequently zero or one would render the left-hand side of equation 1 undefined for a large number of observations. This problem can be circumvented by adding an arbitrary constant (α) to both the numerator and denominator of P/(1-P) to give:

$$\log \frac{P_i + \alpha}{(1-P_i) + \alpha} = \beta_o + \beta \log X_i + \epsilon_i \qquad (3)$$

For small values of α, equation (3) preserves the essential characteristics of the logit transformation.[33,34] In the present case, the actual specification used was as follows:

[32]Weighted least-squares estimation is more efficient than OLS when the number of observations in each group differs. See Kmenta (1971), pp. 322-336.

[33]In equation (3), P is still bounded but the bounds are only approximately equal to zero and one (depending on the size of α). The function still intersects the origin at "equal odds" but is only approximately symmetric on either side of the origin. An asymptotic-like appraoch to the limits of P is preserved; but the left-hand side of equation (3) is bounded between log $[\alpha/(1+\alpha)]$ and log $[(1+\alpha)/\alpha]$, instead of ranging between negative and positve infinity.

[34]Alternatively, maximum likelihood estimation can be applied directly to equation (1) but this involves the solution of a system of non-linear equations and is thus much more expensive and difficult to carry out. See Goldfeld and Quandt (1972), p. 127.

$$\log \frac{\text{SURVIVE 3} + \alpha}{1 - \text{SURVIVE 3} + \alpha} = \beta^o + \beta_1 \text{ LITH} + \beta_2 \text{ LITW} + \beta_3 \text{ SECW} + \beta_4 \text{ LYEARS} + \beta_5 \text{ LINC} + \beta_6 \text{ DIET} + \beta_7 \text{ LMILK} + \beta_8 \text{ LSOLID} + \beta_9 \text{ WATER} + \beta_{10} \text{ TOILET} + \beta_{11} \text{ REFUSE} + \beta_{12} \text{ PRIVATE} + \beta_{13} \text{ PUBLIC} + \beta_{14} \text{ DAI} + u \quad (4)$$

where LMILK and LSOLID denote the natural logarithms of MILK and SOLID and all other variables are defined as in the text.[35] The parameters of equation (4) were estimated by ordinary least-squares (OLS). Because there was no prior basis for selecting a particular value for α, three different values were tried (α = .001, α = .01, α = .1). For each of these three values of α, predicted values of SURVIVE3 were obtained from the estimated logit function, and these were used to obtain an estimate of the R^2 (based on SURVIVE3). The highest R^2 (.138) was obtained for α = .01.[36] This is the same value of R^2 obtained in the regression analysis (see Table 6). The estimation results are reported in Table C-1.

The results of the logit analysis are quite similar to those of the regression analysis (see Table 6 for purposes of comparison). The three variables which were statistically significant at the .01 level in Table 6 (LINC, LYEARS and MILK) are also statistically significant at the .01 level in Table C-1 (although LMILK is specified instead of MILK and MILK2 in the logit analysis). LITH is positive and statistically significant at the .01 level in both Tables 6 and C-1. REFUSE is positive and statistically significant at the .10 level in Table C-1, instead of at the .05 level as in Table 6.

It is interesting to compare the estimated magnitude of the effect of a given change in one of the independent variables on SURVIVE3 from the logit analysis to that from the regression analysis. In general, this depends on the initial level of SURVIVE3. If, for example, SURVIVE3 is at an initial level of .70 (corresponding roughly to an LX of 35 years) then, according to the estimated logit function (α = .01) a doubling of income per adult from 83.5 Rupees per month to 167 Rupees per month would cause SURVIVE3 to increase by .067 while the estimated increase, based on the regression function (Table 6, Col. 5), would be .041. On the other hand, if the initial

[35] No experimentation with alternative functional forms was carried out in the logit analysis. All of the continuous variables were entered as natural logarithms while the qualitative varibles were entered as binary varibles.

[36] The other values of R^2 were .124 (α = .001) and .137 (α = .1). The significance level of only one variable (REFUSE) was affected by the choice of α and only one change in sign was encountered (the coefficient of PUBLIC became positive (t = 0.13 for α = .1).

TABLE C - 1

Modified logit analysis of factors determining survival rates to age three

Independent variables	$\alpha = .001$	$\alpha = .01$	$\alpha = .1$
LITH	.4705	.3216	.1711
	(1.69)*	(1.82)*	(2.00)*
LITW	.1725	.0959	.0325
	(0.44)	(0.38)	(0.27)
SECW	.8756	.5575	.2621
	(1.28)	(1.28)	(1.25)
LYEARS	-1.5954	-.9531	-.3975
	(6.73)***	(6.31)***	(5.44)***
LINC	.7063	.4829	.2605
	(3.10)***	(3.33)***	(3.71)***
DIET	-.0366	-.0448	-.0402
	(0.13)	(0.24)	(0.45)
LMILK	.6962	.4518	.2216
	(3.37)***	(3.43)***	(3.48)***
LSOLID	.1128	-.0518	.0063
	(0.40)	(0.29)	(0.07)
WATER	-.1311	-.0937	-.0556
	(0.18)	(0.21)	(0.25)
TOILET	.0277	.0266	.0220
	(0.04)	(0.07)	(0.12)
REFUSE	.4354	.3018	.1653
	(1.54)	(1.67)*	(1.90)*
PRIVATE	.8765	.5606	.2686
	(1.52)	(1.53)	(1.51)
PUBLIC	-.1945	-.0672	.0180
	(0.44)	(0.24)	(0.13)
DAI	.3463	.2437	.1365
	(1.15)	(1.27)	(1.47)
Constant	6.7297	4.4577	2.2950
R^2	.145	.141	.131
R^2 (based on SURVIVE3)	.121	.138	.106
F	5.97***	5.74***	5.29***
N	506	506	506

*Statistically significant at the .10 level.
**Statistically significant at the .05 level.
***Statistically significant at the .01 level.

value of SURVIVE3 were .90 instead of .70 (i.e. corresponding to an LX of about 60 years), the regression function would still predict a change of .041 with a doubling of income while the logit function would predict an increase of only .029.[37] This difference reflects the non-linearity of the logit as compared to the essential linearity of the regression function. The fact that the calculated R^2's are equal (to three decimal places) means that the results of the present study do not provide any conclusive evidence in favour of one functional form over the other.

The Research Triangle Institute
North Carolina

References

Adelman, L., 1963, An econometric analysis of population growth, American Economic Review, 53.

Afzal, M., and N. Iftikhar, 1974, Net nuptiality tables for males and females in Pakistan, 1962-1965, Pakistan Development Review, 13, no. 1.

Afzal, M., Z. Khan and N. Chaudhry, 1976, Age at marriage, fertility and infant-child mortality in a Lahore suburb, Pakistan Development Review, 15, no. 1.

Anker, R., and J. Knowles, 1980, An empirical analysis of mortality differentials in Kenya at the macro and micro levels, Economic Development and Cultural Change, 29, no. 1, October, 1980.

Berg, A., 1973, The nutrition factor, Washington: Brookings Institution.

Bongaarts, J., 1978, A framework for analyzing the proximate determinants of fertility, Population and Development Review, 4, no. 1.

Coale, A., and P. Demeny, 1966, Regional model life tables and stable populations, Princeton: Princeton University Press.

Farooqui, N., and I. Alam, 1974, Provisional abridged life tables for urban and rural areas in Pakistan, Based on PGS 1968 and 1971, Pakistan Development Review, 13, no. 3.

Frederiksen, H., 1966, Dynamic equilibrium of economic and demographic transition, Economic Development and Cultural Change, 14.

Frederiksen, H., 1971, Determinants and consequences of mortality trends in Ceylon, Public Health Reports, 76, no. 8.

[37] If α = .1 the predicted changes from the estimated logit function are .047 (initial value of SURVIVE3 = .70) and .028 (initial value = .90), while for α = .001 they are .092 (initial value = .70) and .037 (initial value = .90).

Goldfeld S., and R. Quandt, 1972, Nonlinear methods in econometrics, Amsterdam: North Holland Publishing Company.

Iqbal, A., 1974, Relationship of infant-child mortality to fertility: Some theoretical and empirical considerations relevant to Pakistan, Pakistan Development Review, 13, no. 4.

Johnston, J., 1972, Econometric methods, second edition, New York: McGraw-Hill Book Company.

Kmenta, J., 1971, Elements of econometrics, New York: The MacMillan Company.

Malinvaud E., 1966, Statistical methods of econometrics, Chicago: Rand McNally and Company.

Rahimtoola, R. J., 1977, Annual report: epidemiological study of nutrition disorders in pre-school children in an urban community, Unpublished Mimeo, Department of Paediatrics, Jinnah Post-graduate Medical Centre, Karachi.

Rahimtoola, R. J., 1978, Epidemiological study of nutrition disorders in pre-school children in an urban community, Unpublished Mimeo, Department and School of Paediatrics, Jinnah Post-graduate Medical Centre, Karachi.

Schultz, T. P., 1976, Inter-relationships between mortality and fertility, in: Ronald Ridker, ed., Population and development, Baltimore: The John Hopkins University Press.

Stolnitz, G., 1955, A century of international mortality trends, Population Studies, 9, no. 1.

Stolnitz, G., 1974, International mortality trends: Some main facts and implications, in The population debate, 1. (Papers of the World Population Conference, Bucharest, 1974), New York: United Nations.

Theil, H., 1971, Textbook of econometrics, New York: John Wiley and Sons, Inc.