countries among the 53 presented: Nepal, Bangladesh, and India. Data on the under five-year-old mortality rate for these three countries was taken from the ***1998 State of the World's Children Report***. They averaged 113.0 ± 2.6. Limitations in this approach are similar to those in the first model and the small number of countries upon which this model's comparisons are based.

Without data to provide validation of any of the mortality estimates generated here, one could have little confidence in any one of the models developed above. Together, however, these four estimates, depending on three independent data sources, demonstrate a consistent trend toward increased mortality. (See Table 6 below.) These models generate a range of estimates of mortality among under five-year-olds that range from 74 to 135, averaging 104±42. This represents an estimate of more than a doubling in the mortality rate.

Table 6: Summary of Mortality Estimates from Four Models

	Description	Summary of Findings	Results or Average
Model 1	Correlation of Rates of Moderate to Severe Malnutrition Weight for Age (Underweight) Height for Age (Stunting) Weight for Height (Wasting)	 126.1 ± 67.0 131.2 ± 76.5 148.6 ± 64.2	135.3 ± 69.2
Model 2	Projecting from Baghdad	73.5 ± 26.5	73.5 ± 26.5
Model 3	Average of 8 Studies	92.9 ± 68.7	92.9 ± 68.7
Model 4	Population Attributable Risk Comparison	113.0 ± 2.6	113.0 ± 2.6

Exploratory Multiple Regressions

The above models may overestimate mortality by failing to take account of long-term social development investments which have only partly deteriorated. In other words, malnutrition is one, but not the only factor influencing mortality rates among children under five years of age. Literacy, measles immunization coverage, and access to clean water and sanitation facilities likely moderate the impact of nutritional and financial declines related to the sanctions. Other factors, including the proportion of children in primary school, the percent of mothers who are literate, and the percentage of children who are breast fed, may be independent influences or markers for independent influences on mortality rates. Fortunately, the 1996 MICS study and other recent surveys provide these data for Iraq. Data from the ***State of the World's Children Report*** (41) for these and the three nutritional indices for all of 195 other reporting countries were used to establish regression equations to predict under five-year-old mortality rates. Not all data items were complete for each variable in all 195 countries; while several variables were available for all countries, some data were available for as few as half of all countries. A complete set of all variables existed for only 48 countries.

Missing data were dealt with by the 'listwise' method, including only data for countries with complete data sets for all variables; by the 'pairwise' method, where data was imputed for missing variables if data for similar countries existed; and by the 'means' method, where simple mathematical average values were imputed wherever there were missing values. Three to six independent variables were included in each regression model as selected by the 'stepwise' procedure. The stepwise approach is the most common method for selecting variables to be included or excluded from the statistical model. A total of only seven of the nine potential independent variables were ever included in any one of these six stepwise models. Of these, the variables "Percent of infants breast fed at six months" and "GNP" showed the lowest significance and the least contribution to parameter estimates even when included by the stepwise procedure. (Table 7 below shows only combinations of up to five independent variables in the beta values column.)

Because the mortality rate among under five-year-olds appeared to rise in an exponential manner models were subsequently developed for the natural log of mortality. Unadjusted, unweighted R-square values are reported here. Table 7 shows that all of these models generated R-square values, from .73 to .84. Logarithm transformations of the dependent variable consistently generated higher R-square values and smaller standard deviation estimates. Fitting known data for Iraq in 1996 listed at the bottom of the table to the regression equations generated by these models derived estimates of under five-year-old mortality in Iraq from a low of 46 to a high of 110 per thousand.

Which of the regression models presented in table 7 is likely to be most accurate? While each model shows a rise in mortality among under five-year-olds from pre-1990 levels, the range of values generated is widely dispersed. Log transformation of the dependent variable provided lower point estimates, and had dramatically smaller confidence intervals around the point estimate for mortality levels in Iraq. The log models are thus considered better than the models for which the untransformed under five-year-old mortality is used as a dependent variable.

One of the logarithmic models generated a far lower point estimate for Iraq than any other model. This occurred because the beta value for measles immunization was strong for this, and only this, model. While measles immunization is likely an important factor influencing survival among under five-year-olds, a regression model heavily weighted to that one variable minimizes other variables, including nutritional status, which are important in their own right and serve as a marker for access to other essential goods. Indeed, the high measles immunization rate in Iraq is partly the result of a general deterioration of the health system and other social conditions; it is possible to strengthen this one program while most other influences on health are deteriorating.

Only three variables appear in nearly all of these models. With the exception of the model just mentioned, these same three variables have far larger beta values than any other. These variables are literacy among adults, stunting prevalence among under five-year-olds, and percent of population with potable water.

Table 7: Exploratory Multiple Regression Models

Dependent Variable	Independent Variables	Beta Values	Missing Data Handling Method	Number of Data Points	R-Square Value	Under Five-Year-Old Mortality Estimate (per thousand)	95% Confidence Interval
Under Five-Year-Old Mortality	Stunting Adult Literacy Pop with Water	1.013 -0.1488 -0.957	Listwise	47	.73	107	5.6
Log of Under Five-Year-Old Mortality	Per Capita GNP Stunting Pop with Water Sanitation	8.92^{5} 6.963^{3} 5.20^{3} -3.57^{3}	Listwise	47	.82	104	1
Under Five-Year-Old Mortality	Stunting Sanitation Primary School Pop with Water	1.968 -0.552 -0.615 -0.876	Pairwise	75-191	.76	110	8.8
Log of Under Five-Year-Old Mortality	Stunting Adult Literacy Per Capita GNP Pop with Water	6.031^{3} -0.809^{3} -2.24^{5} -6.08^{3}	Pairwise	75-191	.84	83	1
Under Five-Year-Old Mortality	Stunting Measles Pop with Water Adult Literacy Per Capita GNP	0.809 -0.596 -0.809 -1.446 -2.25^{3}	Means	195	.72	112	4.5
Log of Under Five-Year-Old Mortality	Adult Literacy Stunting Pop with Water Measles Per Capita GNP	7.38^{3} 4.40^{3} -2.94^{3} -5.06^{3} -3.405^{5}	Means	195	.77	46	1

Data Used for Above Rates:	(in percentages):	Sources:
Adult Literacy:	68	(34,49)
Measles Immunization Coverage	80	(35,36)
Stunting Prevalence, Under Five-Year-Olds	31.2	(35,36)
Population with Potable Water:	52 (79% with water, 1/3 est. contaminated)	
Percent Households with Adequate Sanitation:	74.6	(35,36) (35,36)
Percent Attending Primary School:	74.2	(37)
Exclusive Breast Feeding, 4 months	30	(35,36)
Per Capita GNP (in dollars):	761 (in 1994)	(37)

Imputing Values for Missing Variables

Since the countries with the most complete data sets are likely a biased sample of all countries, there is interest in an improved approach to imputing values for missing variables. The choice of listwise, pairwise, or means methods to impute missing variables is arbitrary, and fails to provide enough variation to the imputed values.

The "bootstrap" procedure provides a theory-based method to estimate missing data through multiple imputation of missing values. This procedure involves multiple sampling with replacement from the universe of known values to estimate the unknown values.

We use the multiple imputation method (73) to impute data for the missing values in the database from which regression models are developed. The multiple imputation approach uses data items which are nearly complete in the database to create linear regression estimates for the missing values of interest. The value being estimated is thus a dependent variable with some known data values; the known three variables are independent variables used in these models. From this database, those variables include the under five-year-old mortality rate, the under one-year-old mortality rate, and GNP per capita. The model for estimating each missing value is:

$$Y = \beta_0 + \beta_1 V_1 + \beta_2 V_2 + \beta_3 V_3 + \epsilon \qquad \epsilon \sim (\theta, \sigma^2)$$

where Y is the variable to be estimated, β_0 is the intercept value, $\beta_1 V_1$ is the value for the first independent variable, etc. . . ., and ϵ is an error term, error resulting both from variance of the data input, (θ), and the estimation procedure, (σ^2). The procedure for imputing the missing values is done five times, creating five data sets with estimated values. Stochastic variations for the imputed values permits us to fit the following logistic regression model for under five-year-old mortality with a binomial distributed error term.

The binomial distribution, which provides for a logistic transformation of regression models is appropriate for variables that can have either one or another expression, as in the result of death or not death, of interest for this model. This error term includes variance from only the regression model:

$$\log p/1\text{-}p = \beta_0 + \beta_1 V_4 + \beta_2 V_{10} + \beta_3 V_{11} + \epsilon$$

With coefficients and error terms from the above procedure, known data for the model independent variables are included in the five data sets. The mean value of the dependent variable predicted by these five models provides our estimate of under five-year-old mortality in Iraq. The variance of these five estimates provides an additional error term for the imputed variables, by the formula (73):

$$Var(p)= 1/5 \sum_{I=1}^{5} \hat{Var}(p_i) + (1+1/5)\cdot 4\cdot \sum_{I=1}^{5} (\hat{p} - \bar{p})^2$$

The R-square values and the point estimates for Iraq generated by the regression models, using a data base which includes values which were imputed where original values were missing, are very similar to those generated by the exploratory regression above (See table 8). Using the log of mortality still consistently generated lower point estimates for Iraq than regressions using untransformed under five-year-old mortality. As predicted, little power is lost by including the three most important independent variables rather than the entire variable set.

Table 8: Untransformed and Log Regressions Using Imputed Values for Missing Data

Dependent Variable	Independent Variables	Missing Data Handling	Number of Data Points	R-Square Value	Under Five-Year-Old Mortality Estimate
Under Five-Year-Old Mortality	Adult Literacy Stunting Pop With Water	Multiple Imputation	191	.77	110
Under Five-Year-Old Mortality	Adult Literacy Stunting Pop with Water Per Capita GNP Breast feeding Sanitation	Multiple Imputation	191	.77	105
Log of Under Five-Year-Old Mortality	Adult Literacy Stunting Pop with Water	Multiple Imputation	191	.72	78
Log of Under Five-Year-Old Mortality	Adult Literacy Stunting Pop with Water Per Capita GNP Breast feeding Sanitation	Multiple Imputation	191	.82	76

Logistic regression including all variables provided a point estimate for under five-year-old mortality in Iraq midway between estimates using untransformed and log transformed values of under five-year-old mortality (See table 9). Beta coefficients for the three main variables were even more important in the logistic model. Eliminating the three other variables used in the full