A Proportional Hazard Model to Figure out Asian Countries' Food Insecurity

Anwar Fitrianto[1,2,3], Isa bin Daud[1] & Noor Akma Ibrahim[1,2]

[1] Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, Serdang Selangor, Malaysia

[2] Institute for Mathematical Research, Universiti Putra Malaysia, Serdang Selangor, Malaysia

[3] Department of Statistcs, Faculty of Mathematics and Natural Resources, Bogor Agricultural University, Malaysia

Correspondence: Anwar Fitrianto, Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, Mathematics Building, 43400 Universiti Putra Malaysia, Serdang Selangor, Malaysia. E-mail: anwarstat@gmail.com

Abstract

A statistical approach is employed to examine the affects of covariates on food insecurity among Asian countries in the period of 40 years since 1961. We conduct a statistical technique based on information aspects of each country on human resources, environment and sustainability, land use and land resource, agricultural resource and capacity, water and sanitation, and macroeconomic indicators. We found that 22 of 32 (65.72%) countries experience insecurity food condition. The remaining are censored observations (34.38%). Stepwise Cox's regression model is used to select among the 24 independent covariates that are deemed to be significant contribution to the model. Based on the adopted model, at each time point, the West Asian region are found to be more likely to have insecurity food condition compared to those countries in the other regions. Furthermore, the occurrences of food insecurity for East Asia countries are more likely than for those in the other region. Meanwhile, it can also be seen that countries in Lower-middle income group are more likely to reach insecurity food condition than those in the other group. The analysis also shows that the high income countries have high risk of exposure to insecurity food condition.

Keywords: proportional hazard, cox regression, survival analysis, food insecurity

1. Introduction

Asia contributes the most to world population growth, at 50 million people a year, while Africa accounts for only 17 million; although at 2.36 percent, Africa's rate of growth is the highest. Two of every five people alive today are living in China or India. While 10 nations currently have populations that exceed 100 million, the number of nation is expected to rise to 19 by 2050. Half out of these 10 countries are Asian countries (United Nation Population Fund, 2003). Based on IFPRI projection on 1995-2020 (Pinstrup-Andsersen et al., 1999), world population will increase by 32% to 7.5 billion, mostly in cities in developing countries. And 85% of total food demand growth will come from developing countries.

The Food and Agriculture Organization of the United Nations (FAO) defines "food security" as a state of affairs where all people at all times have access to safe and nutritious food to maintain a healthy and productive life. Meanwhile, food is an essential requirement for every individual. The objectives of this article are adopting a mathematical model that describes the survivorship of Asian countries on food security. Then, by having the mathematical models, we can find out factors which influence difference of hazard ratio among Asian countries on food security, so that the insecurity food condition can be detected early, to classify current level of food insecurity condition of Asian countries, and to identify factors which influence the condition at each level.

2. Survival Analysis and Its Corresponding Functions

Survival analysis consists of methods of studying occurrence and timing of events (Lee, 1992). The goal is to analyse measures describing in some sense the width of the interval between an origin point and an end point. Often, the end point corresponds to death or culling and the length from the origin to the end is measured in the time units (seconds, minutes, days, etc) (Durrocq, 1997).

Let T be a nonnegative random variable representing the failure time of an individual from a population. The distribution of T can be specified in many ways, three of which are particularly useful in survival applications: the probability density function, the survivor function, and the hazard function (Chen & Wang, 1991; Lee, 1992; Deshpande & Sudha, 2001). Although these three functions are mathematically equivalent, if one of them is given, the other two can be derived. The cdf of a random variable T, denoted $F_T(t)$, is defined by

$$F_T(t) = P(T < t); \quad t \geq 0. \tag{1}$$

Several properties of a distribution function $F_T(t)$ can be listed as a consequence of the knowledge of probabilities. Because $F_T(t)$ has the probability $0 \leq F_T(t) \leq 1$, then $F_T(t)$ is a non decreasing function of t, and as t approaches ∞ then $F_T(t)$ approaches 1.

Probability density function (pdf) is also very useful in describing the continuous probability distribution of a random variable. According to Lee (1992) the pdf of a random variable T, denoted $f_T(t)$, is defined by:

$$f_T(t) = \lim_{\Delta t \to 0} \frac{P(\text{an event is in interval } [t, t + \Delta t))}{\Delta t}$$
$$= \lim_{\partial t \to 0} \frac{P(t \leq T < t + \partial t)}{\partial t} = \frac{d\, F_T(t)}{dt}. \tag{2}$$

Like any other continuous random variable, the survival time T has a probability that an individual fails in the interval t to $t + \Delta t$ per unit width Δt. Let $T > 0$ have a pdf, $f_T(t)$, and cdf, $F_T(t)$. Then the survival function takes on the following form:

$$S_T(t) = P(\text{an individual surpass longer than or equal to } t)$$
$$= P(T \geq t) = 1 - F_T(t). \tag{3}$$

It is defined as $S_T(0) = 1$ and $\lim_{t \to \infty} S_T(t) = 0$. The function $S_T(t)$ is also known as the cumulative survival rate.

Related to survival function, there is hazard function. It can be computed among those who are at risk (uncensored) at each moment in time. Further, if the censoring time and survival time are independent, then the hazard function among uncensored subjects equals the hazard function among all subjects, which is defined by:

$$h_T(t) = \lim_{\partial t \to 0} \frac{P(t \leq T < t + \partial t \mid T \geq t)}{\partial t}, \tag{4}$$

where, $h(t)$ measures the "risk of dying" at time t, among individual alive at that time. All these distributions are interrelated. In particular:

$$h_T(t) = \frac{f_T(t)}{S_T(t)} = -\frac{d\, \log\, S_T(t)}{dt}, \tag{5}$$

and

$$S_T(t) = \exp\left(-\int_0^t h_T(u)\, du\right). \tag{6}$$

Then, the hazard function on Equation 4 can be written as follows:

$$h_T(t) = P(t \leq T < t + \Delta t \mid T \geq t)$$
$$= \frac{f_T(t)}{(1 - F_T(t))} = \frac{f_T(t)}{S_T(t)}. \tag{7}$$

By some mathematical manipulation, we will get

$$S_i(t \mid x_i) = S_0(t) \exp(\boldsymbol{\beta} \mathbf{x}_i), \tag{8}$$

where $S_0(t) = \exp(-h_0(t))$ is a baseline survival function and $\boldsymbol{\beta}$ is the vector of regression coefficients. Thus, the effect of the covariate values $\mathbf{x}_i$ on the survivor function is to raise it to a power given by the relative risk $\exp(\boldsymbol{\beta} \mathbf{x}_i)$.

3. Cox's Proportional Hazard Model

Let $x_1, x_2, \ldots, x_p$ be the possible prognostic variables. For the ith individual, observed values of the p variables

are $x_{1i}, x_{2i}, \ldots, x_{pi}$. In the multiple Cox's regression approach, the survival time of the ith individual, t_i, is independent variable (Lee, 1992). In a sample of size n, it consists of $(t_i, \delta_i, \mathbf{x}_i), i = 1, 2, \ldots, n$ where t_i is the time on study for the i th individual, δ_i is the event indicator ($\delta_i = 1$ if the event has occurred and $\delta_i = 0$ if the lifetime is censored) and $\mathbf{x}_j$ is the vector of covariates or risk factors for the i th individual ($\mathbf{x}_i$ may be a function of time) which may affect the survival distribution of T, the time to event.

The relation between the distribution of event time and the covariates or risk factors $\mathbf{x}$ ($\mathbf{x}$ is a $1 \times p$ vector) can be described in terms of a model according to Cox (1972), in which the hazard rate at time t for an individual is

$$h(t; \mathbf{x}) = h_0(t) \exp(\mathbf{x}\boldsymbol{\beta}), \tag{9}$$

where $h_0(t)$ is the baseline hazard rate, an unknown (arbitrary) function giving the hazard function for the standard set of condition $\mathbf{x} = \mathbf{0}$. The $\mathbf{x}$ is a $1 \times p$ vector of explanatory variables (covariates) measured at $t = 0$ with associated parameter vector $\boldsymbol{\beta}$ which is a $p \times 1$ vector of unknown parameters. The factor $\exp(\mathbf{x}\boldsymbol{\beta})$ describes the hazard for an individual with covariates $\mathbf{x}$ relative to the hazard at a standard $\mathbf{x} = \mathbf{0}$.

When we makw special assumption about $h_0(t)$, it leads to parametric models. But the advantage of Cox's model is the fact that such assumptions can be avoided. The approach is said to be semi parametric. The Cox model is called a *proportional hazards model* since the ratio of the hazard rates of two individuals with covariate values $\mathbf{x}_1$ and $\mathbf{x}_2$ is

$$\frac{h(t \mid \mathbf{x}_1)}{h(t \mid \mathbf{x}_2)} = \exp(\mathbf{x}_1 - \mathbf{x}_2)\boldsymbol{\beta}, \tag{10}$$

an expression which does not depends on t. Writing the covariate vector of the ith individual $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$, further rearrangement shows

$$\log\left(\frac{h_1(t)}{h_0(t)}\right) = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}, \tag{11}$$

so that the proportional hazards model is a linear model of the log hazard ratio (Shoukri & Cheryl, 1999). Estimates of the unknowns $h_0(t)$ and $\boldsymbol{\beta}$ are obtained in the following way.

Let $t_1 < t_2 < \ldots < t_r$ denote the ordered distinct event times (suppose there are no ties) and let x_{ik} be the k th covariate associated with the individual whose failure time is t_i, $k = 1, \ldots, p$. Further, define the risk set at time t_i, $R_{(t_i)}$, as the set of all individuals who are still under study at a time just prior to t_i. The partial likelihood according to Cox's, based on the hazard function, is expressed by

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{r} \frac{\exp\left(\sum_{k=1}^{p} \mathbf{x}_{ik}\boldsymbol{\beta}\right)}{\sum_{j \in R(t_i)} \exp(\mathbf{x}_{ij}\boldsymbol{\beta})}. \tag{12}$$

Then, its partial maximum likelihood estimates are found by maximizing the natural logarithm of the likelihood equation above:

$$LL(\boldsymbol{\beta}) = \sum_{i=1}^{r}\sum_{k=1}^{p} \beta_k x_{ik} - \sum_{i=1}^{r} \ln\left[\sum_{j \in R(t_i)} \exp\left(\sum_{k=1}^{p} \beta_k x_{jk}\right)\right]. \tag{13}$$

The efficient score equations are found by taking partial derivatives of the equation with respect to the β 's as follows:

Let

$$U_h(\boldsymbol{\beta}) = \frac{\delta LL(\boldsymbol{\beta})}{\delta \beta_h}, \quad h = 1, 2, \ldots, p, \tag{14}$$

then

$$\frac{\delta LL(\boldsymbol{\beta})}{\delta \beta_h} = U_h(\boldsymbol{\beta}) = \sum_{i=1}^{r} x_{ih} - \sum_{i=1}^{r} \frac{\sum\limits_{j \in R(t_i)} x_{jh} \exp\left[\sum\limits_{k=1}^{p} \beta_k x_{jk}\right]}{\sum\limits_{j \in R(t_i)} \exp\left[\sum\limits_{k=1}^{p} \beta_k x_{jk}\right]}.$$ (15)

4. Data and Methodology

4.1 Data

The data source is from FAO, specifically Statistical Year Book of FAO and Compendium of Food and Agriculture Indicators. The Food and Agricultural Indicators are based on data available in the FAO statistical database (FAOSTAT) as of mid November 2001 and Compendium of Food and Agricultural Indicators 2001 (FAO, 2001). The macro-economic indicators are mainly based on information extracted from World Development Indicators 2001 on the online World Bank database (World Bank, 2004).

The data set to be used in the study consists of food and agricultural indicators and macro-economic indicators. The food agricultural indicators are obtained in the FAO statistical database (FAOSTAT) as of mid November 2001 and Compendium of Food and Agricultural Indicators 2001 (FAO, 2001). Meanwhile, the macro-economic indicators are based on information extracted mainly from World Development Indicators 2001 on the online World Bank database (World Bank, 2004). According to FAO, there are 36 countries of Asia, however, a total of 32 countries are considered to be analysed based on the complete availability of the dataset. The datasets contain information aspects of each country on human resources, environment and sustainability, land use and land resource, agricultural resource and capacity, water and sanitation, and macroeconomic indicators.

There is also information about Agricultural Production Index (API). According to (FAO, 2001), the index is the relative level of the aggregate volume of agricultural production for each year in comparison with the base period 1989. To obtain the index, the aggregate for a given year is divided by the average aggregate for the base period. The API dataset of each country is recorded by FAO since 1961 until 2001. The value changes or assumed remain in a certain condition over the years. Generally, two different countries with the same API values do not necessarily have the same of agricultural condition or food security conditions. This is due to different population sizes.

Basically, API value can be written as:

$$y_{ij} = \frac{p_{i,j}}{p_{i,1989}} \times 100; \left(i = 1, 2, \ldots, 32; j = 1961, 1962, \ldots, 2001\right),$$ (16)

where, y_{ij} is API value of country i at year j, $p_{i,j}$ is the agriculture production quantity of country i at year j, and $p_{i,1989}$ is the agriculture production quantity of country i at base year 1989 (FAO, 2001).

In this study, the unit of observations is an Asian country. There are several continuous variables in the study. Those variables are agricultural production index (1961-2001), population total (1961-2001), CO_2 emissions in metric tons per capita (V5), life expectancy at birth, year 2000 (V6), age dependency ratio (V7), percentage of urban population (V8), average irrigation for agricultural area (V9), average land use for agricultural area (V10), average tractor use (V11), average land use for arable & permanent crops year 1961 to 2001 (V12), average ratio agricultural to total area year 1961 to 2001 (V13), average precipitation year 1961 to 1990 (V14), total renewable water resources (V15), per capita GDP year 2001 in USD (V16), annual population growth between 1996 to 2000 in percentage (V17), Ratio rural/total population, in percentage (V18), Ratio agricultural/total labor force, in percentage (V19), agricultural GDP as share of total GDP, in percentage (V20), total population/arable land (pop/ha) (V21), fertilizer use/arable land (kg nutrs./ha) (V22), and annual growth rate of agricultural GDP, year 1990 to 1999 (%) (V23). Meanwhile, categorical factors considered in the study are human development indicator, region, income group, and indebtness.

For the categorical factors, the corresponding dummy variables are introduced for the analysis. Typically such variables are coded as 1 if the variable has the characteristic. The coding of each categorical variable is presented in Table 1. The 'event' that matters in defining time variable is the first sign of food insecurity problem. The food security condition of each country is represented by API value.

Table 1. Categorical variable and the coding

Variables	Categories	Code	Dummy Variables		
Human Development Indicator			Dummy1 *(D1)*	Dummy2 *(D2)*	
	High	1	1	0	
	Medium	2	0	0	
	Low	3	0	1	
Region			Dummy3 *(D3)*	Dummy4 *(D4)*	Dummy5 *(D5)*
	South East Asia	1	1	0	0
	West Asia	2	0	1	0
	East Asia	3	0	0	1
	South Asia	4	0	0	0
Income Group			Dummy6 *(D6)*	Dummy7 *(D7)*	Dummy8 *(D8)*
	Low	1	1	0	0
	Lower-Middle	2	0	1	0
	Upper-Middle	3	0	0	0
	High	4	0	0	1
Indebtness			Dummy9 *(D9)*	Dummy10 *(D10)*	Dummy11 *(D11)*
	Severely	1	1	0	0
	Moderately	2	0	1	0
	Less	3	0	0	0
	No debt	4	0	0	1

4.2 Methodology

As it was suggested by Therneau & Grambsch (2000), an initial time point of the first observation of each subject (country) is assumed to begin at $t = 0$. We will define the starting point of survival analysis based on Agricultural Production Index. Agricultural Production Index among Asian countries increase over time. On the other hand, as Malthusian theory has mentioned, the population grows faster. As the result, the Agricultural Production Index that is weighted by the population for individual country mostly decreases over time, although some countries have constant or low-slope increasing as Figure 1 displayed. We use this fact to create time variable. First, we assume that 1961 is the starting point at which all countries under study have the same percentage values of Weighted Agricultural Production Index (WAPI). By using right censoring mechanism, a country is said be experiencing of the event if the value of its WAPI decrease by a certain percentage.

Let w_{ij} is WAPI value of country *i* year *j*, and *n* is the number of Asian countries:

$$w_{ij} = \frac{y_{ij}}{z_{ij}}, \quad \left(i = 1,2,\ldots,32; j = 1961,1962,\ldots,2001 \right), \tag{17}$$

where, y_{ij} and z_{ij} are API value of country i at year j and population size of country i at year j, respectively. As a consequence, the average value of WAPI among Asian countries is equal to:

$$\overline{w}_j = \frac{1}{n} \sum_{i=1}^{n} \frac{y_{ij}}{z_{ij}}.$$

Figure 1. Trend of weighted Agricultural Production Index of some selected asian countries under study during 1961-2001

Not all countries in the study will reach the terminal event during the study time frame of 1961 to 2001. These represent the censored time variable. However, if during that period of WAPI in the forth coming years after the starting point decreases less than a certain value, called the cut-off value, then the country is said to experience the 'event'. The number of years until the event occurs is the uncensored time variable.

Table 2. Illustration of event and time definition based on WAPI data

| Country | WAPI (in percentage) | | | | | |
	1961 (starting point of event)	1962	1963	1964	...	2001
A	100	99	98	97	...	
B	100	100	100	99	...	
C	100	100	102	100	...	101

Table 2 presents an artificial WAPI data from 1961 until 2001 as an illustration. For example, if we define cut-off value of 1% decreasing, country A will have the value of the time variable as 1 year, while Country B is 3 years. Country C is right censored since no event occurs during the period. The value of censoring time is 40 years. We may change the cut-off value arbitrarily since the statistical definition of food insecurity condition based on Agricultural Production Index does not exist. Figure 2 presents an illustration of the study design where the observation times start at a consistent point in time ($t = 0$). All computations will be done using SAS software.

Figure 2. Illustration of study design where the ($t = 0$) is a starting point of event

5. Results

5.1 Cox's Model Building

In this section, we investigate the Cox's model building for certain definition about time variable and censoring. We use the censoring variable by using 45% reduction of WAPI. It means that if WAPI of an individual (a country) decreases until or less than 45% compared to the WAPI value at point of event (1961), then the individual is said to have an event. We found that 21 of 32 (65.62%) countries experience insecurity food condition. The remaining were censored observations (34.38%). We use stepwise Cox's regression to select among 24 independent variables which significaly contribute to the model.

Employing a SAS code of PHREG procedure involving only the significant variables at $\alpha = 0.05$ based on stepwise procedures. We discovered 6 variables have significant contribution to the response variable at $\alpha = 0.05$. Those variables are D4, D5, D7, D8, V10, and V12. The summary of the output is displayed in Table 3.

Table 3. Parameter estimation of the Cox's model of Asian countries food security

Variables	Parameter estimate, $\hat{\beta}$	Standard error, $SE(\hat{\beta})$	Hazards ratio, $\exp(\hat{\beta})$	*p value*
D4	1.37210	0.64611	3.944	0.0337
D5	-4.04918	1.54674	0.017	0.0088
D7	1.64775	0.70185	5.195	0.0189
D8	3.62154	1.17778	37.395	0.0021
V10	0.0000317	0.0000113	1.000	0.0051
V12	-0.0002014	0.0000618	1.000	0.0011

The parameter estimates, $\hat{\beta}$, of the D4 is 1.37210 with a *p value* of 0.0337. The positive sign indicates a positive association between the hazard of being insecure and the D4 variable. Furthermore the ratio of the hazards is given by

$$\frac{\lambda_2(t)}{\lambda_1(t)} = \frac{\exp^{Dummy4=1}}{\exp^{Dummy4=0}} = \exp^{1.37210} = 3.944 \, .$$

It means that at each time point, the models predict that countries in West Asia region are approximately 3.9 times more likely to have insecurity food condition than those of other regions. This fact is also shown in Figure 3. Furthermore, the hazard ratio of the D5 is equal to 0.017, which mean that occurrence of food insecurity for others Asian countries is approximately $1000/17 = 58.82$ times more likely to have insecurity food condition than for those in the East Asia regions.

Figure 3. Comparison of stimated hazard function $\hat{\lambda}(t)$ of time to insecurity food condition for West Asia and the other countries

In the Figure 3, the curve is quite short because no events in this group beyond $t = 5$. Meanwhile, the coefficient

for D5 is negative, but the hazard ratio is less than 1. Meanwhile, we also found that countries of Lower-middle income group were approximately 5.2 times more likely to reach insecurity food condition than those in the other group, since the value of $\exp^\beta = 5.195$.

In the case of variable D8, one immediately notices an extremely high risk of exposure to the state of food insecurity condition. As we can see in Figure 4 (also Table 3), high income countries have more risk to experience the insecurity food. The high income countries have about 37.395 likely to have insecurity condition. That fact occurs in countries, which have low capacity of agricultural production. Although, it is widely believed that high-income countries should have higher opportunities to give their citizen more food due to higher income they have, their risk to have insecurity food is very high due to very low value of API. Although they can overcome their need of food by their power of foreign trade, they have to reduce the risk by increasing the agricultural production.

Figure 4. Comparison for the High income and the other countries of the estimated hazard function $\hat{\lambda}(t)$ of time to insecurity food condition

5.2 Assessment of Food Insecurity Categories

We define time variable from 1% until 45% reduction of WAPI, as has been mentioned in the Methodology. It means that there will be 45 time variables. We analyze the dataset, contains covariates, which we used in the previous Cox' model building. As consequences, there will be 45 created models. But, we are only interested in significance variable (at $\alpha = 0.05$) appropriate to each created model. Then, those variables are categorized into three categories based on interval of reduction percentages; 1%-15%, 16%-30%, and 31%-45% respectively for Low, Medium and High condition of food insecurity.

Related to regions, Table 4 presents cross tabulation of the stage of insecurity food condition among Asian countries. We need to construct the cross tabulation to see whether there exist association between the level of food security and region.

Table 4. Cross tabulation between region of Asian countries and stage of food insecurity

Food Insecurity Stage	Region				Total
	South East Asia	South Asia	West Asia	East Asia	
High	5	6	12	1	24
Medium	4	0	0	1	5
Low	0	0	1	2	3
Total	9	6	13	4	32

Table 5. Chi-square tests of stage of Asian countries food insecurity in association with region

Statistical test	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	19.271	6	0.004
Likelihood Ratio	18.840	6	0.004
N of Valid Cases	32		

Based on the table, we roughly say that there exists association between those variables. But, formally, we need to test that association by using Chi-square test. The result of the statistical analysis is displayed on Table 5. Both Pearson Chi-square and Likelihood Ratio test give *p value* which is equal to 0.004. We reject the null hypotheses of no association between those two categorical variables.

6. Discussion

The goal of this study is to develop an analysis of the effect on Asian countries food security through statistical vision. Survival analysis is analytical technique that uses the counting process of data input for managing a series of observations of time intervals with a defined starting time point, stopping time point and event status, recurrence or censored (Cox & Oakes, 1984). The initial time point of the first observation of each subject is assumed to begin at zero (Therneau & Grambsch, 2000). We analyse the food insecurity as a single event. The 'event' that matters in defining time variable is the first sign of food insecurity problem and the time variable is the time (in years) from 1961 until the event occurs. The food security condition of each country is represented by API value. The API that is weighted by the population for individual country (WAPI) mostly decreases over time, although some countries have constant or low-slope increase as in Figure 3. This means that it is not impossible that one country has non monotonic WAPI. As a consequence, further research and follow-up study may focus on the non monotonic WAPI, where we have to consider not just the single event, but also follow-up event in defining the food insecurity problem.

References

Chen, C. H., & Wang, P. C. (1991). Diagnostic plots in Cox's regression model. *Biometrics, 47*, 841-850. http://dx.doi.org/10.2307/2532643

Cox, D. R. (1972). Regression Models and Life Tables. *Journal of Royal Statistical Society, B*(34), 187-20.

Cox, D. R., & Oakes, D. (1984). *Analysis of Survival Data*. London: Chapman and Hall.

Deshpande, J. V., & Sudha, G. P. (2001). Survival, hazard, and competing risk. *Journal of Current Science, 80*, 1191-1202.

Durrocq, V. (1997). Survival analysis, a statistical tool for longevity data. 48th Annual *Meeting of the European Association for Animal Production* 25-28 August, 1997, Vienna, Austria.

FAO. (2001). *Compendium of Food and Agricultural Indicators 2001*. Rome: FAO.

Lee, E. T. (1992). *Statistical Methods for Survival Data Analysis*. New York: John Wiley & Sons.

Pinstrup-Andsersen, Per., Pandya-Lorch, R., & Mark, W. R. (1999). World Food Prospects: Critical issues for the early twenty-first century. *Research Report No. 29*. International Food Policy Research Institute. Washington.

Shoukri, M. M., & Cheryl, A. P. (1999). *Statistical Methods for Health Sciences* (2nd ed.). CRC Press, Boca Raton.

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. New York: Springer.

United Nations Population Fund (UNPFA). (2003). Global Population and Water: access and sustainability. *Research Report Number 6*. United Nations, New York.

World Bank. (2004). Online data base. World Development Indicators 2001-2001. Retrieved from http://www.worldbank.org/data/countryclass/countryclass.html.