

## Environmental Determinants of Child Mortality in Rural China:

### *A Competing Risks Approach*

Hanan Jacoby\* & Limin Wang\*\*<sup>1</sup>

#### Abstract

We use a competing risk model to analyze environmental determinants of child mortality using the 1992 China National Health Survey, which collects information on cause of death. Our primary question is whether taking into account of cause of death using a competing risk model, compared with a simple model of *all-cause* mortality, affects conclusions about the effectiveness of policy interventions. There are two potential analytical advantages in using cause of death information: (1) obtaining more accurate estimates and (2) validating causal relationships. Although, we do not find significant differences between estimates obtained from the competing risk model and those from simpler hazard models, we do find evidence supporting the causal interpretations of the effect of access to safe water on child mortality. Our analysis also suggests that a respondent-based health survey can be used to collect relatively reliable information on cause of death. Modifying future demographic and health survey (DHS) instruments to collect cause of death information inexpensively may be worthwhile for enhancing the analytical strength of the DHS.

World Bank Policy Research Working Paper 3241, March 2004

*The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent. Policy Research Working Papers are available online at <http://econ.worldbank.org>.*

---

\* DECRG, World Bank, \*\* ENV, World Bank.

We gratefully acknowledge the financial support from Swedish International Development Agency (SIDA), TF024884. We thank Kirk Hamilton, the task manager of this study, for suggestions and comments, Jostein Nygard and Adam Wagstaff for useful discussions. Alexandra Sears provided excellent assistance.

## Introduction

Environmental risk factors account for about one-fifth of the total burden of disease in low-income countries according to recent estimates (World Bank, 2001). WHO (2002) reports that among the 10 identified leading mortality risks in high-mortality developing countries, unsafe water, sanitation and hygiene ranked second, while indoor smoke from solid fuels ranked fourth. A number of econometric studies using household survey data from low income countries also find significant relationships between environmental factors and child morbidity and mortality (e.g., Wolfe and Behrman, 1982; Lee, et al. 1997). Such evidence suggests that public investments in health infrastructure can improve health outcomes, particularly child survival prospects.

In this paper, we evaluate alternative empirical methodologies for estimating the impact of environmental factors on child mortality in micro-data. Our primary question is whether taking into account cause of death using a competing risk model affects conclusions about the effectiveness of policy interventions. Such information is critical for prioritizing public investments in order to maximize the health benefit for given resources, particularly in the context of achieving the targets set by the Millennium Development Goals (MDG) on child mortality and environment. To our knowledge, no previous econometric study of child mortality has explicitly modeled cause of death. Indeed, distinguishing child deaths by cause may seem unwarranted. After all, the objective of policy should be to prevent child deaths from *any* cause; as long as a child's life is saved, it should be a matter of indifference whether the averted death would have been the result of diarrhea disease or respiratory infection.<sup>2</sup> By this reasoning, it is sufficient to analyze the determinants of *all-cause* mortality, which only involves estimating a simple child survival-time (or hazard) model.

Nevertheless, there are a couple reasons why accounting for cause of death using a competing risks framework might be advantageous. First, this type of survival model is more flexible than an all-cause model and might, therefore, give more accurate estimates of environmental impacts. Second, separate estimates of the effect of environmental factors by cause of death may provide a way to detect the presence of confounding

---

<sup>2</sup> This is not to deny the *epidemiological* interest in the particular pathways by which an intervention prevents a death.

factors -- in other words, of endogeneity bias. Certain environmental variables should have no effect on the probability of dying from certain causes. If it turns out that they do, then these environmental variables may be picking up unobserved attributes of households or communities that are correlated with mortality outcomes, the presence of which would invalidate any causal inferences.<sup>3</sup>

Information on cause of death is collected in only a few DHS, usually by self-reports combined with verbal autopsy methods. Aside from the well-known reliability issue, one problem is that, given the typical DHS sample size, the number of deaths from any particular cause is likely to be small. For this reason, we turn to a national health survey conducted in China in 1992, and modeled after the DHS, which does provide information on cause of child death. An advantage of the China survey is its vast sample size – more than 4,000 child deaths are recorded in rural areas – and its broad geographical coverage, which ensures substantial variation in environmental risk factors. Another advantage of looking at China, especially rural areas, is that local provision of public goods, such as piped water, is likely to be less responsive to local tastes than in a more pluralistic and market-oriented economy. As a result, the issue of purposive program placement (Rosenzweig and Wolpin, 1986) is likely to be less important in our context.

The paper is structured as follows. Section 1 provides a summary of the China health survey. In section 2, we lay out the different hazard models used in our empirical analysis and discuss the other econometric procedures. Section 3 reports results. Section 4 concludes.

## **1. Data**

The 1992 China Health Survey (CHS) was designed and implemented jointly by the China Statistical Bureau (CSB) and UNICEF. It covers all 29 provinces (except Tibet), using a sampling frame based on the 1990 census. The survey is representative both at the national and at the provincial level. Within rural areas, which are the focus of our analysis, stratification was done by geographic characteristics: hills, mountainous

---

<sup>3</sup> Galiani, et al. (2002) make a similar argument in the context of their evaluation of a water service privatization program in Argentina using aggregate data. They show that the program reduces mortality from water-borne disease, but does not reduce mortality from other causes.

areas and plateaus. The primary sampling units (PSU) are villages. A total of 380,305 rural households were interviewed during the month of June 1992. The survey questionnaire design follows closely that of the DHS. There are four basic survey instruments: (1) a household questionnaire, (2) a questionnaire for women, (3) a questionnaire for children, and (4) a village questionnaire (only for the rural sample).

All women who had given birth since 1976 are interviewed. All births between 1977-92 are recorded with the outcomes, dates, place, and type of birth assistance. For children who died before age 5 at the time of the survey, information on date, place and cause of death are collected. Studies of several DHS show evidence of downward bias in reporting child deaths; i.e., the longer the recall period, the more likely the respondents misreport the case. In order to minimize the possible recall bias on child deaths, we consider live births in the five years preceding the survey, which give a sample of 160,899 births.<sup>4</sup> The sample distribution by province is presented in Table 1 in the appendix.

Since the 1992 CHS followed closely on the heels of the 1990 census, it is worth comparing estimates of province-level child mortality rates across the two data sources. As summarized in Table 1, this comparison is encouraging; it shows that for most provinces (22 out of 29) the mortality estimates from the CHS are roughly in line with those from the census.<sup>5</sup>

The CHS collects information on cause of child death based on the mother's report. Responses are chosen from a set of causes and symptoms questionnaire,<sup>6</sup> rather than conducting a verbal autopsy.<sup>7</sup> Given that about 77% of the rural deaths occurred at home, and the difficulties in ascertaining cause of death even under the best of circumstances, we expect considerable noise in this variable. Nevertheless, ignoring

---

<sup>4</sup> The number of child deaths in urban areas (555 over a 5-year period ) is too small to permit a separate survival analysis, hence we focus on rural areas.

<sup>5</sup> The seven provinces which have a considerably different estimate of IMR include Hanan, Jilin, Jiansu and Guanxi, Zhejiang, Hubei, Guanjia.

<sup>6</sup> Detailed survey methods are reported in "1992 National Health survey of children: proceedings on survey and research", published by the China State Statistical Bureau.

<sup>7</sup> Verbal autopsies (VA) are commonly used in the field of epidemiological research to collect information on causes of death in countries where child deaths often occurred at home without medical attention. It involves the design of a syndrome module and verification from medical personnel. Martha Anker (1997) provides an assessment of the reliability of VA.

cause of death information altogether, as has been done in much of the previous research on the determinants of child mortality, is not necessarily the best approach either.

Table 2 shows the distribution of reported cause of death by age group from the 1992 CHS along with comparable data from rural Pakistan, based on the verbal autopsy method, and a less developed country (LDC) aggregated number based on a different estimation methodology.<sup>8</sup> Results from the CHS are broadly consistent with conventional epidemiological wisdom: birth-related causes constitute the principal cause of death within the first month of birth (43%), while ARI and diarrhea diseases become more important after the first month. The overall distribution of causes of death in rural China resembles that in rural Pakistan,<sup>9</sup> except that diarrhea-caused U5 mortality is much smaller in China (6% versus 21%), probably reflecting better sanitary conditions in China. The aggregate LDC figures highlight four leading causes of death. Ranked in descending order, these are birth-related causes, other illnesses (unspecified), ARI and diarrhea. The relative importance of ARI vis a vis diarrhea related death in China is more or less in line with the LDC aggregate. We can conclude from these comparisons that the information collected on cause of death in the CHS is relatively sensible. However, it remains to be seen whether a more detailed analysis of these data will deliver sensible and credible results.

Aside from the misreporting issue, cause of death is often simply unknown. This problem crops up not only in the CHS, but, as the data from rural Pakistan reveal, in surveys that use the verbal autopsy method as well. Rather than drop observations with unknown cause of death, and thereby risk a sample selection bias in our survival analysis, we impute unknown causes to one of the five identified categories (ARI, diarrhea, tetanus, birth related and other illnesses) using the following procedure. First, we use a multinomial logit model to estimate the probabilities of the five specified causes as functions of socio-economic characteristics of the household and community, a set of province dummies, and a fifth order polynomial in the age of death of the child. Based

---

<sup>8</sup> The LDC aggregate is based on estimates from Global Burden of Disease Study 2000. The cause of death have been estimated based on data from national vital registration system, population laboratories and epidemiological studies. For most countries particularly in poor regions, vital registration systems are not available, therefore, the cause of death is extrapolated using specially developed statistical methods (Murray etc, 2001).

<sup>9</sup> The Pakistan survey collected causes of death for more than 1,000 deaths occurring in rural areas between 1990-94.

on these estimates, we predict the probabilities of dying from each of the five specified causes for every child whose actual cause of death is unknown. We then assign the child with the unknown cause to the known cause with the highest predicted probability. As should be expected, given the nature of this imputation, our procedure does not lead to major changes in the distribution of causes of death (see Table 2).

## 2. Empirical Methods

We consider three models for the length of time,  $t$ , that the child survives. The respective hazard functions  $h(t)$  are defined as follows:

**Weibull (W):** 
$$\alpha\theta t^{\alpha-1}$$

Where  $\alpha$  is the parameter, and  $t$  is the survival duration.

**Piece-wise Weibull (PWW):** 
$$\sum_k d_k \alpha_k \theta_k \tilde{t}_k^{\alpha_k-1},$$

The survival time  $t$  is divided into  $k$  intervals by the points  $c_1, c_2, \dots, c_{k-1}$

where 
$$d_k = 1 \quad \text{if } c_{k-1} < t \leq c_k \quad k = 1, 2, 3$$
  

$$= 0 \quad \text{otherwise}$$

and 
$$\tilde{t}_k = (t - c_{k-1})d_k + c_k(1 - d_k) \quad \text{and } c_0 = 0$$

**Competing Risks (CR):** 
$$\sum_j D_j \alpha_j \theta_j t^{\alpha_j-1},$$

where 
$$D_j = 1 \quad \text{if died from cause } j$$
  

$$= 0 \quad \text{otherwise}$$

The covariates  $X$  enter through the  $\theta$  parameter, so that in the CR model, for example,  $\theta_j = \exp(\beta_j' X)$ , where  $\beta_j$  is a cause-specific vector of coefficients. The PWW model allows for age-specific coefficients, based on three age intervals, as discussed below. In

each case, the survival time density is given by  $h(t)S(t)$ , where  $S(t) = e^{-\int_0^t h(\tau)d\tau}$  is the

survival probability.<sup>10</sup> The log of  $h(t)S(t)$  is the individual contribution to the likelihood function that we maximize.

While the CR model uses information on cause of death explicitly, it involves estimating a separate vector of coefficients for each cause: birth-related, diarrhea, ARI, tetanus, other illness, and accidental.<sup>11</sup> Our PWW model is more parsimonious, but also, to an extent, captures differences in the cause of death by age of death. A large body of clinical evidence shows that the determinants of mortality differ considerably for neonatal (first month of life) and postnatal mortality. Neonatal deaths are typically related to factors associated with maternal care during pregnancy and delivery (Fikree, Azam and Berendes, 2002). Socio-environmental variables are more important determinants of child survival after the first month, as infectious diseases and poor nutrition become more prominent risk factors. Diarrheal disease, in particular, becomes much more prevalent after weaning, which usually occurs by the second year of life (Black, Brown, Becker, Abdul Alim and Merson, 1982). To capture these age effects, we choose three cut-off points for the survival duration in the PWW model:

$c_1 = 1$  month;  $c_2 = 12$  months;  $c_3 = 60$  months.

Unobserved heterogeneity is an issue that arises in the estimation of hazard models of any form, but one that we do not address in this paper. It is difficult to distinguish duration dependence from unobserved heterogeneity in single-spell duration (survival time) data (Heckman and Singer, 1994), and the Weibull specifications already allow for duration dependence. Thus, the results from the policy simulations are unlikely to be much affected by, in addition, accounting for unobserved heterogeneity.<sup>12</sup>

We do address the issue of the potential endogeneity of environmental risk factors. Household-level variables, such as access to piped water and sanitary housing conditions, are arguably correlated with parental preferences for and knowledge of child

---

<sup>10</sup> In the CR model, the survival probability depends not on the integral of a single cause-specific hazard, but on the sum of the integrated hazards from all causes (i.e.,  $S(t) = \exp(-\sum_j \theta_j t^{\alpha_j})$ ). This is because in order to survive to age  $t$  a child must avoid dying from *all* causes.

<sup>11</sup> In the case of accidents, we do not include any covariates, so that the hazard rate for accidental death is assumed not to depend on any household, community, or child characteristics, except the child's age.

<sup>12</sup> We do account for heterogeneity in our standard error calculations. All the variance-covariance matrices are adjusted for clustering at the PSU level using the 'sandwich' estimator.

health, factors which may exert an independent influence on mortality outcomes. To deal with this problem, we use PSU or ‘cluster’ means of these household variables, under the plausible assumption that, at this higher level of aggregation, variation in environmental risk factors primarily reflect differences in opportunities (i.e., prices and access) rather than differences in household preferences.<sup>13</sup>

A related point is that we are only interested here in the ‘reduced form’ impacts of environmental risk factors on child survival. A structural approach would allow for these factors to be mediated by such endogenous variables as nutrition and illness episodes, as well as child parity and mother’s age at first birth (see Wolpin, 1997). While a structural model sheds more light on the pathways by which exogenous environmental factors influence child survival, it is also more dependent on identifying assumptions than our reduced form approach. More importantly, a structural model yields the same overall policy conclusions as a reduced form, and these conclusions are our primary interest. In the following analysis, we use the reduced form of hazard models.

### 3. Results

Table 3 presents the variables used in our analysis along with the descriptive statistics. We focus on three environmental variables: access to safe drinking water, access to basic sanitation facilities and use of clean cooking fuels.

The full set of parameter estimates of the three survival models are reported in Appendix tables 2 and 3. Since these parameters are difficult to interpret and compare, we present our main results in terms of changes in under age-5 (U5) survival probability. In particular, for a given change in one of the  $X$  variables, say from  $X_0$  to  $X_1$ , we calculate the change in the predicted child U5 survival probability as  $S(\hat{t}; X_1) - S(\hat{t}; X_0)$ ,

---

<sup>13</sup> Rosenzweig and Wolpin (1986) argue that at the community level such aggregates also reflect average household preferences (and unobserved endowments) and that, further, the allocation of local public goods is attentive to these factors. As a consequence of this ‘purposive placement’ of social services, access to such services is endogenous. In the China context, however, purposive placement of, say, piped water infrastructure is probably less of an issue than in most countries, for reasons already mentioned. The possibility of selective migration of households to areas with better health infrastructure is also unlikely in rural China given its restrictive internal migration policies, although policies on rural-urban migration became more relaxed since the early 1990s.

where  $\hat{t}$  is equal to 60 months. In Table 4, these estimates are reported as lives saved per 1,000 births. Standard errors of the estimates are computed using the delta method.

Looking at four hypothetical policies: (1) universal private access to safe water; (2) universal access to basic sanitation facilities; (3) universal access to clean fuels; and (4) universal female primary education attainment -- we find little difference in their impact on child survival across the three models. The largest and most significant impact comes from access to safe water. Increasing this dummy variable from its sample mean value of 0.33 to universal access would save more than 3 lives out of 1,000 births, based on the CR and W models, and somewhat less than 3 lives according to the PWW model. Note that the latter model tends to under-predict survival probabilities, as well as changes in survival probabilities, relative to the other two. Policies that achieve universal female primary education attainment also have a significant impact on reducing U5 mortality (although only at the 10% level of significance for W and CR models).

Table 4 shows the results for the models estimated both with cluster means of the environmental variables and with the household-level variables directly. As discussed, any differences between these two sets of estimates can be attributed to the endogeneity of environmental health conditions at the household level.<sup>14</sup> All the models include a full set of province dummies,<sup>15</sup> so that we only exploit the cross-cluster variation in environmental variables *within* each province. For each survival model, we find little difference according to whether household variables or their cluster means are used,<sup>16</sup> except in the case of access to sanitation, and cooking fuel. Using the household-level sanitation variable produces significant effects in all models. However, using the cluster mean greatly attenuates the coefficient and raises the standard error somewhat.

The impact of interventions on child survival prospect is likely to vary across localities, as well as among households of different socio-economic background. To

---

<sup>14</sup> Cluster-level variables may also be capturing the effects of community health externalities, which would then be part of the reduced form effect of the hypothetical policy change. In the case of maternal education, we attempt to examine the externality effects on child mortality by introducing both the individual and cluster-level variables together in the same model.

<sup>15</sup> In the CR model, province dummies are not included in each cause-specific hazard. Instead, the province dummies are restricted to have the same coefficient in each hazard. This is because for certain causes there are few, or no, deaths in a number of provinces.

<sup>16</sup> The correlation coefficients between the household- and cluster-level variables are 0.82, 0.88 and 0.88 for access to safe water, access to sanitation and use of clean fuel for cooking, respectively.

examine the differential impact of policy changes on child mortality, we use the same parameter estimates to also calculate changes in the predicted U5 survival probability evaluated at the sample means of poor households and households residing in poor counties.<sup>17</sup> Table 5 summarize the estimation of U5 deaths averted from policy changes targeted at poor households and poor localities. For the sake of comparability with earlier results, we calculate the elasticity of numbers of U5 deaths saved with respect to changes in access to safe water. The simulation results illustrate that the health benefits are larger when policies are targeted at poor localities or poor households, in comparison to that from untargeted policies (i.e. for all rural areas), with elasticity for access to piped water being 6.5 and 6.2 for poor counties and households respectively, compared with 5.6 for all rural areas. The same is true for all policy changes (sanitation, clean fuel and female education), although the numbers of lives saved are not statistically significant for sanitation and clean fuel.

To help validate the findings, particularly in regards to the impact of safe water, we consider the PWW and CR estimates more closely. In the PWW, we find, as expected, that the safe water variable has the smallest impact in the first month of life, when most infants are exclusively breastfed, as compared to the two later periods (see Appendix table 2). This result suggests that access to safe water is not just capturing general socioeconomic conditions that influence infant and child mortality at any age.

In a similar vein, the CR model allows us to calculate the (unconditional) probability that a child U5 dies from a specific cause, as well as how a given change in access to safe water influences each cause-specific probability. Note that it is not correct simply to compare the estimated cause-specific hazard rate coefficients in Appendix tables 2 and 3. In general, the cause-specific mortality probabilities depend on all the model parameter by virtue of the formula  $\Pr(\text{death from cause } j) = \int_0^{\infty} h_j(t)S(t)dt$  where  $h_j(t)$  is the cause-specific hazard (see Thomas, 1996). Table 6 reports the results of this

---

<sup>17</sup> Poor counties are defined as those with average per capita income in the bottom two quintiles across all counties. We define poor households as those whose per capita income are in the bottom two quintiles of the distribution.

calculation, showing the predicted cause-specific probabilities and the changes in the probabilities induced by the hypothetical policy of universal access to safe water.

The first point to note in Table 6 is that the predicted changes in cause-specific mortality probabilities do indeed differ in relative terms compared to the corresponding hazard rate coefficient estimates. Nonetheless, it is still true that safe water has a negligible impact on the risk of dying from tetanus and birth-related causes, which is encouraging from the point of view of model validation. Universal access to safe water reduces the odds of dying from diarrhea, acute respiratory infection (ARI), and other illness by about the same absolute amount. In relative terms, however, improved access to safe water reduces the death rate from diarrhea the most, followed by ARI and other illnesses.

It should be noted that the prevalence of fever/ARI is particularly high in China. Using the 1992 CHS, the estimated ARI incidence among children U5 is 17.5% compared to the LDC average of 9.4% (using all low-income countries with DHS data excluding China). Although epidemiological studies often suggest that fever/ARI is caused mainly by exposure to air pollution resulting from using solid fuels for cooking and heating, or combustion of fossil fuels for transport and power generation<sup>18</sup>, other pathways through which the ARI virus can be transmitted are also possible. The recent outbreak of severe acute respiratory syndromes (SARS) in China and Hong Kong, has shown that such a virus can be transmitted through piped water. While perhaps surprising, in China at least, some ARIs may be due to water-borne pathogens.

To sum up, what have we learned by estimating a competing risk model versus a simpler all-cause model of child mortality? First, how reliable is the cause of death information? Establishing causes of child death accurately is difficult, especially when children died at home as is often the case in most LDCs. We assess the reliability of the China survey by comparing the distribution of cause of death with that in a Pakistan survey which uses the much more thorough (and expensive) methodology of verbal autopsy. The overall distribution of cause of death in rural China resembles that in rural

---

<sup>18</sup> Rudan and Aambell (2002) reviewed 39 community-based studies to estimate the incidence of ARI in children U5 in LDCs. They conclude cautions should be taken on the empirical findings, and only studies with explicit diagnostics criteria for ARI, and diagnoses were made by specially trained field workers who were involved surveillance over at least one year, can be regarded as methodologically credible.

Pakistan, except that diarrhea-caused U5 mortality is much smaller in China (6% versus 21%), probably reflecting better general sanitary conditions in China. We also impute unknown causes into one of the five identified causes based on predicted probabilities using a multinomial logit model. The ranking of the five causes from the imputed data is similar to that using the 2000 data for all LDCs estimated by WHO. The above comparison provides encouraging evidence suggesting that respondent-based surveys can be used to collect relatively reliable information on cause of death.

Second, does the CR approach affect policy conclusions? Using the three different hazard models, we simulate policy interventions that aim to achieve universal access to safe water, basic sanitation, use of clean cooking fuel, and universal female primary education in rural China. The comparison of results from these different models show no significant differences between the estimated environmental effects on child mortality from the CR model and those from simpler hazard models (using all-cause mortality) that do not explicitly account for cause of death. Hence, we conclude that the accuracy of overall policy conclusions about the effectiveness of environmental interventions would not be improved if one had access to cause of death information.

Third is the identification of policy interventions. The analysis shows that interventions targeted at improving access to safe water in rural China have a statistically significant impact on reducing U5 mortality probability. To achieve universal female primary education attainment can improve child survival prospect, but the education impact is significant only in one model specification (the PW model). We do not find evidence suggesting that to improve access to sanitation or clean cooking fuels in rural China can significantly reduce child mortality risk. The policy simulation also shows that targeting environmental policies (in particular private access to safe water) in poor localities or poor households can avert more U5 deaths than untargeted interventions, with elasticity for access to safe water being 6.5 and 6.2 for poor counties and households respectively, compared with 5.6 for all rural areas.

Fourth is validating causal effects. One of the potential analytical advantages of using the CR model lies in the validating causal relationships using information on cause of death. Using the CR model, we find that the probability of death from causes that should not be related to safe water, i.e., birth related deaths and neonatal tetanus, in fact

are unrelated to access to safe water. Moreover, the probability of dying from diarrhea disease is the most responsive, at least in relative terms, to interventions that improve access to safe water. These findings lend support to our causal interpretation of the impact of access to safe water on reducing child mortality risk.

#### **4 Conclusions**

In this paper, we have analyzed the impact of environmental factors on child mortality in rural China using a competing risk model that takes into account the cause of death. We have argued that, given the ultimate objective of predicting the total number of child deaths that would be averted by a specific policy change, information on cause of death may, in principal, be superfluous. However, we have also identified two reasons why such information might be useful in practice: first, to obtain more accurate estimates of policy impacts and, second, to validate that these estimated impacts are indeed causal.

What have we learned by estimating a competing risk mode versus a simpler *all-cause* model of child mortality? First, we find that taking into account cause of death information does not affect conclusions about the effectiveness of policy interventions on reducing child mortality risks. Second, knowledge on cause of death is particularly useful for validating causal interpretation of the effect of access to safe water on child mortality, hence increasing our confidence that we are not picking up spurious effects when modeling using *all-cause* mortality framework. Third, the policy simulations show that policy interventions targeted at poor localities or poor households avert more U5 deaths than untargeted interventions. Fourth, the findings from this study suggest that modifying future DHS survey instruments to incorporate the collection of simple cause of death information, especially for high mortality countries, may be worthwhile.

If one accepts the analytical benefits of knowing cause of death, then it is proper to ask about the costs of accurately collecting such information. In the China survey that we analyze, cause of death is based on the mother's report alone, prompted by a questionnaire listing the main causes and symptoms of fatal events. Verbal autopsies, though undoubtedly more accurate, are also much more expensive to implement, and

would be prohibitively costly on a survey of such vast scale as the CHS. Thus, it is encouraging that the distribution of death by reported cause looks quite sensible in the CHS data. This suggests that modifying future DHS survey instruments to incorporate the collection of simple cause of death information, especially for high mortality countries, may be worthwhile. Indeed, a particularly useful experiment for evaluating this accuracy-cost tradeoff would be to collect respondent-based information alongside verbal autopsy data in the same DHS.

## References

Black, R, K Brown, S Becker, A Alim and M Merson (1982), "Contamination of weaning foods and transmission of enterotoxigenic *Escherichia coli* diarrhea in children in rural Bangladesh", *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 76: 259-264.

China State Statistical Bureau (1993), *National Final Report : 1992 National Health Survey of Children*, China Statistics Publishing House.

China State Statistical Bureau (1993), *Proceeding on Survey and Research: 1992 National Health Survey of Children*, China Statistics Publishing House.

Fikree, F , S Il Azma and H Berendes (2002), "Time to focus child survival programs on the newborn: assessment of levels and causes of infant mortality in rural Pakistan", *Bulletin of the World health Organization*, 80 (4), pp271-276.

Galiani S, P Gertler and E Schargrotsky (2002), "Water for life: the impact of the privatization of water services on child mortality", Center for Research on Economic Development and Policy Reform, working paper No. 154.

Heckman, J.J and Singer, B, (1984), "A method for minimizing the impact of distributional assumptions in econometrics models for duration data" *Econometrics* 52, 271-320.

Huang, R and L Yan, (1995), *Mortality Data of China Population*, China Population Publishing House.

Lee, L, M Rosenzweig and M Pitt, (1997), "The effects of improved nutrition, sanitation, and water quality on child health in high-mortality populations", *Journal of Econometrics*, 77 p209-235.

Marth Anker (1997), "The effect of misclassification error on reported cause-specific mortality fractions from verbal autopsy", *International Journal of Epidemiology*, vol 26, no 5, pp 1090-1096.

Murray, C, A Lopez, C Mathers and C Stein, 2001, "The global burden of disease 2000 project: aims, methods and data source", Global program on evidence for health policy discussion paper No 36.

Rosenzweig, M. R and K.I. Wolpin (1986), "Evaluating the effects of optimally distributed public programs: child health and family planning interventions", *American Economic Review*, 76: 470-482.

Thomas, D, V Lavy and J Strauss (1996), "Public policy and anthropometrics outcomes in Cote d'Ivoire", *Journal of Public Economics* 61: 155-92.

Thomas, J (1996), "On the interpretation of covariate estimates in independent competing risks models", *Bulletin of Economic Research*, 48 volume 1, p27-39.

Wolpin, K.I. (1997) "Determinants and consequences of the mortality and health of infants and children" in M.R. Rosenzweig and O Stark (eds.), *Handbook of Population and Family Economics*, Vol 1A, Elsevier, Amsterdam.

Wolfe, B. L. and J. R. Behrman (1982), "Determinants of Child Mortality, Health, and Nutrition in a Developing Country." *Journal of Development Economics*, vol. 11 : 163-193.

World Bank (2001), "Health and Environment". Background paper for the World Bank Environment Strategy. Washington, D.C.

World Health Organization (2002), *The World Health Report 2002:Reducing Risks, Promoting Healthy Life*, WHO, Geneva.

## **Main Result Tables**

**Table 1: Infant mortality rates in rural China**

<i>Province</i>	<i>1992 survey</i>	<i>1990 census</i>
Beijing	13.2	8.1
Tianjin	11.7	8.8
Hebei	10.1	9.4
Shan(1)xi	28	20.8
Neimong	31.1	34.8
Liaoning	22.3	20
Jilin	19.4	28.7
HeilongJiang	24.8	22.3
Shanghai	10.7	14.9
Jiansu	32	15.7
Zhejiang	28.8	19
Anhui	27.8	27.5
Fujian	26.2	23.8
Jaingqi	51.9	45.7
Shandong	17.6	14
Henan	17.5	19.4
Hubei	40.2	27.8
Hunan	62.5	40.4
Guangdong	11.8	16.8
Guanqi	29.2	46.5
Hanan	7.5	31.1
Sichun	41.2	39.7
Guizhou	75.3	55.3
Yunan	68.3	69.7

Note: IMR from the 1992 survey are estimated using life table approach, and adjusted using sample weights. The IMR should be interpreted as mortality rates between 1987-92, five years before the survey. The IMR estimates from 1990 census are also based on life table method (Huang R and L Yan, 1995). They are the estimates of IMR between 1989-90.

**Table 2: Distribution of cause of death by survival duration: An International Comparison**

<b>Causes</b>	Birth related	Tetanus	ARI	Diarrhea	Oth. illness	accidence	unknown	total (%) No. deaths
<b>(a) China Survey</b>								
	<b><u>From the data</u></b>							
<1 month (%)	43.37	15.92	6.47	1.22	14.01	4.2	14.8	100
								<b>2783</b>
1 month - 1 year (%)	10.79	3.51	21.59	13.31	31.55	7.95	11.3	100
								<b>1195</b>
1 year - <5 years (%)	6.92	0.97	10.79	13.28	31.26	31.4	5.39	100
								<b>723</b>
All deaths (%)	29.48	10.47	10.98	6.15	21.12	9.34	12.47	100
No. deaths	<b>1386</b>	<b>492</b>	<b>516</b>	<b>289</b>	<b>993</b>	<b>439</b>	<b>586</b>	<b>4701</b>
	<b><u>Imputation</u></b>							
<1 month (%)	55.48	16.92	6.72	1.33	4.2	15.34		100
								<b>2783</b>
1 month - 1 year (%)	12.55	3.6	23.26	13.97	38.66	7.95		100
								<b>1195</b>
1 year - <5 years (%)	7.19	0.97	11.07	13.69	35.68	31.4		100
								<b>723</b>
All deaths (%)	37.14	11.08	11.59	6.45	24.4	9.34		100
No. deaths	<b>1746</b>	<b>521</b>	<b>545</b>	<b>303</b>	<b>1147</b>	<b>439</b>		<b>4701</b>
<b>(b) Rural Pakistan 1990-93</b>								
Infant deaths (%)	21.7	11.7	11.6	21.6	11.7		21.8	100
No. deaths	<b>248</b>	<b>133</b>	<b>132</b>	<b>246</b>	<b>133</b>		<b>249</b>	<b>1141</b>
<b>(d) All LDC 2000</b>								
Age 0-4	27.4	2	20	12.1	0.7	26.2		100
No. deaths								<b>10.9 (million)</b>

Note: The figures for China are unweighted. We impute unknown causes to the five specified causes of death (birth related, tetanus, ARI and diarrhea, and other illnesses). Birth related in CHS cause include suffocation caused by umbilical cord, prematurity, and congenital anomaly. In the Pakistan survey, birth related causes cover low birth weight, small size for gestational age, birth injury, congenital anomaly and prematurity. For all LDCs, birth related causes include conditions arising in the perinatal period and congenital anomalies, and accidence includes only motor vehicle related accidents

All LDC data are quoted from a report, "The global burden of disease 2000 project: aims, methods and data source", C Murray, A Lopez, c Mothers and C Stein, 2001, Global program on evidence for health policy discussion paper No 36. The figures for Pakistan are based on a survey of 54,834 households and 1141 infant deaths in two provinces (Balochistan and North-West Frontier province), Fikree, F,S Azam and H Berendes (2002).

**Table 3: Variable summary**

<b>Variable</b>	<b>Estimation</b>	<b>Definition</b>	<b>(Mean)</b>	<b>(S dev)</b>
<b>Individual-Household-level</b>				
Child's gender	<i>Male</i>	male=1, female=0	0.54	0.50
Mother high school	<i>M high</i>	mother with high school and above education (=1)	0.06	0.23
Mother middle school	<i>M middle</i>	mother with middle school education(=1)	0.31	0.46
Mother primary school	<i>M primary</i>	mother with primary school education(=1)	0.40	0.49
log(income-ph)	<i>Log(inc-ph)</i>	household income per capita	6.98	0.65
Living space per head	<i>Lspa-ph</i>	living space per capita	21.00	14.26
<b>Cluster-level</b>				
Safe water*	<i>Safe water</i>	covers water sources from private tap water and deep well	0.33	0.39
Has sanitation *	<i>Has sanitation</i>	includes flush and latrine toilets, either private or shared	0.85	0.31
Clean cooking fuel I*	<i>Clean fuel I</i>	includes electricity, liquidized petroleum gas, bio gas, coal and oil	0.29	0.40
Clean cooking fuel II*	<i>Clean fuel II</i>	same as fuel I, but exclude coal	0.02	0.14
Mother's education	<i>M mid&amp;hgh</i>	% of households with female of middle and above education attainment	0.52	0.10
<b>County-level</b>				
Access to buses	<i>Village Bus</i>	% villages in the county with access to buses	0.26	0.44
Access to clinics	<i>Village Clinic</i>	% villages in the county with access to health clinics	0.70	0.46

\* household level values are used in some specifications ( see Table 4 and appendix Table 3)

**Table 4: Predicted U5 deaths averted per 1,000 Births from achieving universal access: Policy simulations**

	<b>Model</b>	<b>W</b>	<b>PWW</b>	<b>CR</b>
<b>Policy change</b>	Current access	U5 Deaths	U5 deaths	U5 deaths
	(% HH)	(s.e)	(s.e)	(s.e)
<b>Cluster-variable</b>				
Access to safe water	33.2	3.3 (1.1)	2.6 (1.1)	3.5 (1.4)
Access to sanitation	85.3	0.4 (0.2)	0.3 (0.3)	0.5 (0.3)
Use clean cooking fuels I	29.2	0.6 (1.2)	0.4 (0.9)	0.8 (1.1)
Use clean cooking fuels II	1.84	1.2 (0.2)	0.9 (0.9)	7.6 (12.0)
Universal female primary education	76.9	0.6 (0.4)	0.6 (0.2)	0.5 (0.3)
<b>HH-variable</b>				
Access to safe water	32.7	3.5 (0.8)	2.7 (0.8)	3.4 (1.0)
Access to sanitation	85.2	0.8 (0.2)	0.6 (0.2)	0.7 (0.3)
Use clean cooking fuels I	30.1	0.4 (0.9)	0.3 (0.7)	0.6 (0.9)
Use clean cooking fuels II	0.8	0.7 (0.1)	5.2 (3.9)	4.8 (5.1)
Universal female primary education	76.9	0.6 (0.3)	0.6 (0.2)	0.9 (0.3)
U5MR (at sample mean) (1000 births)	K-M method 33.2	W 30.7	PW 24.3	CR 30.9

Note: W, PWW and CR refer to weibull, piece-wise weibull, and competing risk models, respectively. The definition of clean fuel I includes coal, while clean fuel II treats coal as dirty fuel. K-M method refers to Kaplan-Meier method which is a nonparametric approach to estimate survival probability. Standard errors are adjusted for clustering on PSU in parentheses.

**Table 5: Predicted U5 deaths averted: Policies targeted at the poor  
(per 1000 Births)**

<b>Policy change</b>	<b>Model</b>	<b>W</b>	<b>PWW</b>	<b>CR</b>
	Current access (% HH)	U5 Deaths (s.e)	U5 deaths (s.e)	U5 deaths (s.e)
<b>Poor Households (bottom 2 quintiles)</b>				
Access to safe water	25.3	4.5 (1.5)	3.5 (1.5)	5.1 (1.9)
Access to sanitation	84.0	0.6 (0.3)	0.4 (0.4)	0.6 (0.5)
Use clean cooking fuels I	29.8	0.7 (1.4)	0.4 (1.1)	1.1 (1.4)
Use clean cooking fuels II	0.8	7.5 (1.4)	9.2 (14.4)	5.9 (6.1)
Universal female primary education	38.7	1.0 (0.6)	1.0 (0.4)	0.8 (0.4)
<b>Poor Counties (bottom 2 quintiles)</b>				
Access to safe water	25.0	4.1 (1.4)	3.3 (1.4)	4.7 (1.8)
Access to sanitation	84.9	0.5 (0.3)	0.4 (0.3)	0.6 (0.4)
Use clean cooking fuels I	32.4	0.6 (1.3)	0.4 (1.0)	0.9 (1.2)
Use clean cooking fuels II	0.6	13.2 (1.8)	10.2 (10.2)	9.2 (14.4)
Universal female primary education	38.0	0.9 (0.6)	0.9 (0.3)	0.7 (0.4)

**Table 6: The impact of access to safe water on U5 mortality probability (U5MP)**

Cause of death	Safe Water Coef (s.e.)	U5MP (per 1000 births)			
		Actual	Predicted	Change in U5MP	
				Absolute	Percentage (%)
Diarrhea	-1.09 (-0.27)	1.88	1.86	-0.96	-51.5
ARI	-0.76 (-0.19)	3.39	3.55	-1.41	-39.9
Tetanus	-0.12 (-0.17)	3.23	2.88	-0.21	-7.3
Birth related	0.01 (-0.09)	10.85	11.14	0.11	1.0
Other illnesses	-0.23 (-0.12)	7.13	7.67	-1.06	-13.9

## Appendix Tables

**Table 1: Summary of sample statistics**

Province	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	population		No. county	No. cluster	No. HH	No. births	No. deaths
	Total	Rural China					
	(unit 10,000)	(unit 1000)					
Beijing	1094	1644	8	105	1388	1585	18
Tianjin	909	2248	7	98	1338	1531	15
Hebei	6220	33040	42	363	6143	7560	84
Shan(1)xi	2942	11312	35	289	4633	6266	140
neimong	2184	10542	33	256	3706	4598	106
Liaoning	3990	10548	24	291	3866	4170	84
Jilin	2509	6358	20	285	4017	4634	72
HeilongJiang	3575	10129	27	257	3842	4619	93
Shanghai	1340	4213	6	166	1370	1442	12
Jiansu	6844	30513	34	369	4783	5810	137
Zhejiang	4202	15165	27	438	5277	5824	135
Anhui	5761	36814	31	271	4041	5503	110
Fujian	3079	13806	26	285	4739	6687	135
Jiangqi	3865	22433	34	235	4149	6118	259
Shandong	8570	30100	41	357	5216	6258	100
Henan	8763	54646	42	289	4875	6531	96
Hubei	5512	10577	32	231	4300	6063	181
Hunan	6209	33699	37	304	4768	6477	304
Guangdong	6439	5301	30	231	4134	5974	87
Guanqi	4324	22742	33	300	5177	7466	158
Hanan	674	1912	5	25	460	702	5
Sichun	10897	71311	54	414	5882	6912	251
Guizhou	3315	21508	31	252	4620	6631	368
Yunan	3782	21603	38	263	5006	6968	385
<b>Total</b>	<b>106999</b>	<b>482164</b>	<b>697</b>	<b>6374</b>	<b>97730</b>	<b>126329</b>	<b>4701</b>

Note : columns (1)-(2) are from China Statistical Year Book, 1992. Columns (3)-(7) are calculated from the 1992 CHS.

**Table 2. Estimation Results from Three Hazard Models: Cluster-level Environmental Variables**

	(Mean) (S dev)	W	PWW			CR				
		coef (s.e)	<1 mth coef (s.e)	1-12 mth coef (s.e)	12-60 mth coef (s.e)	Diarrhea coef (s.e)	ARI coef (s.e)	Tetanus coef (s.e)	Birth related coef (s.e)	Oth illnesses Coef (s.e)
Safe water	0.33 (0.39)	-0.17 (0.05)	-0.12 (0.09)	-0.30 (0.12)	-0.21 (0.16)	-1.09 (0.27)	-0.76 (0.19)	-0.12 (0.17)	0.01 (0.09)	-0.23 (0.12)
Has sanitation	0.85 (0.31)	-0.10 (0.07)	-0.07 (0.10)	-0.21 (0.14)	0.04 (0.18)	0.05 (0.20)	-0.04 (0.17)	0.01 (0.17)	-0.28 (0.11)	-0.07 (0.12)
Clean fuel I	0.29 (0.40)	-0.03 (0.04)	-0.02 (0.07)	-0.01 (0.09)	-0.04 (0.12)	0.04 (0.17)	0.01 (0.13)	-0.12 (0.13)	-0.04 (0.08)	-0.07 (0.09)
M mid& above	0.52 (0.10)	-0.78 (0.19)	-0.61 (0.28)	-1.18 (0.43)	-0.81 (0.52)	-1.35 (0.80)	-0.79 (0.61)	-0.92 (0.59)	-0.45 (0.34)	-1.34 (0.38)
Male	0.54 (0.50)	0.00 (0.03)	0.04 (0.04)	-0.07 (0.06)	-0.08 (0.08)	-0.07 (0.11)	0.04 (0.09)	-0.04 (0.09)	0.07 (0.05)	-0.07 (0.06)
M high	0.06 (0.23)	-0.22 (0.08)	-0.08 (0.11)	-0.48 (0.19)	-0.57 (0.25)	-0.41 (0.34)	-0.13 (0.26)	-0.19 (0.25)	0.10 (0.13)	-0.70 (0.22)
M Middle	0.31 (0.46)	-0.17 (0.05)	-0.10 (0.07)	-0.26 (0.09)	-0.56 (0.12)	-0.29 (0.18)	0.03 (0.15)	-0.38 (0.15)	-0.07 (.08)	-0.20 (0.10)
M primary	0.40 (0.49)	-0.09 (0.04)	-0.08 (0.05)	-0.11 (0.08)	-0.28 (0.10)	-0.21 (0.14)	0.14 (0.12)	-0.09 (0.11)	-0.06 (0.07)	-0.17 (0.08)
Log (inc-ph)	6.98 (0.65)	-0.28 (0.02)	-0.26 (0.04)	-0.17 (0.05)	-0.27 (0.07)	-0.18 (0.10)	-0.30 (0.07)	-0.46 (0.07)	-0.24 (0.04)	-0.36 (0.05)
Lspa-ph	21.00 (14.26)	0.00 (0.00)	0.00 (0.00)	-0.01 (0.00)	0.00 (0.00)	0.00 (0.01)	-0.01 (0.00)	-0.01 (0.00)	0.00 (0.00)	0.00 (0.00)
Village Bus	0.26 (0.44)	0.06 (0.04)	0.04 (0.06)	0.09 (0.08)	0.11 (0.10)	0.14 (0.16)	-0.04 (0.11)	0.08 (0.12)	-0.01 (0.07)	0.17 (0.08)
Village Clinic	0.70 (0.46)	-0.04 (0.04)	-0.04 (0.05)	-0.05 (0.08)	-0.01 (0.09)	-0.54 (0.14)	-0.11 (0.10)	-0.05 (0.11)	0.06 (0.07)	-0.01 (0.08)
LogL		-31369.4		-29154				-36669.6		
No obs		160899		160899				160899		
No failure		4701	2783	1195	723	1746	521	545	303	1147

Note: location and province dummy variables are included in all three models.

**Table 3. Estimation Results from Three Hazard Models: HH-level Environmental Variables**

	(Mean) (S dev)	W	PWW			CR				
		coef (s.e)	<1 mth coef (s.e)	1-12 mth coef (s.e)	12-60 mth coef (s.e)	Diarrhea coef (s.e)	ARI coef (s.e)	Tetanus coef (s.e)	Birth related coef (s.e)	Oth illnesses coef (s.e)
Safe water	0.33 (0.47)	-0.18 (0.04)	-0.09 (0.06)	-0.36 (0.09)	-0.34 (0.12)	-0.85 (0.22)	-0.57 (0.13)	-0.12 (0.14)	-0.04 (0.07)	-0.21 (0.09)
Has sanitation	0.85 (0.36)	-0.17 (0.05)	-0.11 (0.08)	-0.36 (0.10)	-0.05 (0.14)	-0.02 (0.17)	-0.12 (0.13)	-0.07 (0.13)	-0.31 (0.08)	-0.13 (0.10)
Clean fuel I	0.30 (0.46)	-0.02 (0.04)	-0.01 (0.05)	0.01 (0.08)	-0.12 (0.10)	0.05 (0.15)	0.00 (0.11)	-0.11 (0.11)	-0.01 (0.06)	-0.08 (0.08)
M mid& above	0.52 (0.10)	-0.78 (0.19)	-0.64 (0.27)	-1.17 (0.420)	-0.69 (0.51)	-1.53 (0.80)	-0.93 (0.59)	-0.93 (0.58)	-0.45 (0.33)	-1.36 (0.38)
Male	0.54 (0.50)	0.00 (0.03)	0.04 (0.04)	-0.07 (0.06)	-0.08 (0.08)	-0.07 (0.11)	0.04 (0.09)	-0.04 (0.09)	0.07 (0.05)	-0.07 (0.06)
M high	0.06 (0.23)	-0.22 (0.08)	-0.08 (0.11)	-0.47 (0.19)	-0.55 (0.25)	-0.42 (0.34)	-0.13 (0.26)	-0.18 (0.25)	0.10 (0.13)	-0.69 (0.22)
M Middle	0.31 (0.46)	-0.16 (0.05)	-0.09 (0.07)	-0.26 (0.09)	-0.54 (0.12)	-0.29 (0.18)	0.03 (0.15)	-0.38 (0.15)	-0.07 (0.08)	-0.20 (0.10)
M primary	0.40 (0.49)	-0.09 (0.04)	-0.08 (0.05)	-0.11 (0.08)	-0.27 (0.10)	-0.21 (0.14)	0.14 (0.12)	-0.08 (0.11)	-0.06 (0.07)	-0.17 (0.08)
Log (inc-ph)	6.98 (0.65)	-0.28 (0.02)	-0.27 (0.04)	-0.17 (0.05)	-0.27 (0.07)	-0.18 (0.10)	-0.30 (0.07)	-0.46 (0.07)	-0.23 (0.04)	-0.36 (0.05)
Lspa-ph	21.00 (14.26)	0.00 (0.00)	0.00 (0.00)	-0.01 (0.00)	0.00 (0.00)	0.00 (0.01)	-0.01 (0.00)	-0.01 (0.00)	0.00 (0.00)	0.00 (0.00)
Village Bus	0.26 (0.44)	0.06 (0.04)	0.04 (0.06)	0.09 (0.08)	0.12 (0.09)	0.13 (0.16)	-0.05 (0.11)	0.09 (0.12)	-0.01 (0.07)	0.17 (0.08)
Village Clinic	0.70 (0.460)	-0.04 (0.04)	-0.04 (0.05)	-0.05 (0.08)	-0.01 (0.09)	-0.54 (0.14)	-0.11 (0.10)	-0.05 (0.11)	0.07 (0.07)	-0.01 (0.08)
LogL		-31359		-29138				-36665		
No obs		160899		160899				160899		
No failure		4701	2783	1195	723	1746	521	545	303	1147

Note: location and province dummy variables are included in all three models.

