Selection into Worst Forms of Child Labor

By ERIC V. EDMONDS [*]

Little is known about why children participate in activities that are labeled worst forms of child labor. Case-control approaches common in medicine are adapted to consider the correlates of participation in worst forms in the context of two worst forms of child labor in Nepal: portering and ragpicking. Paternal disability is a strong predictor of entry into each of the worst forms, and the presence of productive assets within the child's home reduces the risk a child is observed in a worst form. We argue that our findings are consistent with a model where there are negative amenities associated with these jobs that induce the poor and those with the fewest alternative earnings options to select into these worst forms of child labor in Nepal.

JEL Classification: D10, J22, J80

Keywords: Child Labor, Exploitation, Hazardous Work, Parental Health

I. Introduction

Popular horror of the prevalence and persistence of worst forms of child labor in developing countries is nearly universal. 160 countries have ratified ILO Convention 182 "Concerning the Prohibition and Immediate Action for the Elimination of the Worst Forms of Child Labor." Commensurate with this public attention to worst forms is a literature within economics seeking to understand why children are engaged in worst forms of child labor. Policy tends to view worst forms as evidence of victimization. Children are often not free to choose their own time allocation, and one argument for the persistence and prevalence of worst forms is that they reflect parental neglect, indifference to the child's welfare, or coercion.

Empirical evidence on the determinants of selection into the worst forms of child labor is scarce (Edmonds 2007 is a review), because worst forms are difficult to capture with randomized sampling. Most of our understanding comes from research that interviews children engaged in a specific activity about their working conditions and why they participate in the work. Children often

[*] I am grateful to Salil Sharma and Maheshwor Shrestha for exceptional research assistance. My time on this project was funded by the International Child Labor Program of the Bureau of International Labor Affairs, U.S. Department of Labor and portions of this paper appear in the ICLP report: “Alternative Income Generation and Entry into Worst Forms of Child Labor.” This paper has benefited greatly from the comments of Alessandro Cigno, Deb DeGraff, Ken Swinnerton, and Sarah Gromly as well as to seminar attendants at the U.S. Department of Labor, The World Bank, the IZA/WB Conference on Employment in Development, the University of Minnesota Development Conference, and the Understanding Children's Work / University of Paris I child labor conference. Correspondence to Eric Edmonds, Department of Economics, Dartmouth College, 6106 Rockefeller Center, Hanover NH 03755 USA, eedmonds@dartmouth.edu. This version: September 2009.

respond that they are working because either they or their family need the money. However, the fact that children work in worst forms for income does not itself answer the question of why children are working in these activities. Children work in plenty of activities for income, many of which would not be considered hazardous or a worst form of child labor. More generally, it is impossible to understand why children are involved in some activity without talking to children that are not involved in that activity. In order to design policy aimed at preventing child involvement in worst forms of child labor, policy needs to know what factors are associated with entry to worst forms of child labor and whether these correlates of entry differ from correlates of entry into other types of work.

This study argues for analyzing the correlates of participation in worst forms by pooling nationally representative data and survey data from children in worst forms. The present study applies the simplest of approaches from the statistical literature on inference in contaminated samples (where sampling probabilities are correlated with treatment status) to consider the correlates of participation in worst forms of child labor. This is not a causal approach, and inference is limited to child background characteristics that are collected in the survey of children in a worst form and the nationally representative data.

This study examines what child background characteristics make it more likely that a child is observed as a short route porter or ragpicker in Nepal. Both types of work have been defined as a worst form of child labor in Nepal by the Nepali government and explicitly targeted for eradication. Survey data from children engaged in each activity are combined with estimates of the incidence of each in the population and with nationally representative data from Nepal's population census to show how this combination of data can be used to infer the correlates of selection into these worst forms of child labor in Nepal. Each activity is considered separately, and findings for children observed in worst forms are compared to results from analyzing selection into regular wage work.

Some striking patterns are observable in the data. Paternal disability appears to be strongly correlated with child participation in the worst forms considered herein. The presence of employment opportunities within the child's own household is associated with a diminished risk of entry into worst forms. Such patterns appear in the worst forms examined herein, but these correlates of entry into worst forms do not predict positively participation in wage work. Given that children rarely work for wages in Nepal, the association between participation and household employment opportunities suggests that children are more likely to be observed in a worst form when the return to the worst form is large relative to other options available to the child. These findings are consistent with Dessy and Pallage's (2005) model of compensating wage differentials in worst forms of child labor. The negative amenities associated with the worst forms of child labor are compensated so that there is sorting into worst forms along the marginal utility of income.

The next section of the study discusses the concept of a worst form of child labor and reviews the existing theoretical literature on why children participate in worst forms. Section three describes the present methods for studying entry into worst forms. Section four details the data used in this analysis, and section five presents the findings. Section six concludes with a discussion of the lessons of the empirical findings and considers the implications of this study's weaknesses for future studies of entry into worst forms of child labor.

II. Theory - Are Worst Forms Different?

ILO Convention C182 on the Worst Forms of Child Labor asks signatory countries to clarify the definition of worst forms of child labor in the signatory's country and to develop specific plans for their eradication. Article 3 of C182 contains several guidelines for what types of activities are to be considered for persons under the age of 18. These include all forms of slavery and "practices similar to slavery." This later clause is noted to include the sale and trafficking of children, debt bondage, serfdom, and forced or compulsory labor including for the purposes of armed conflict. Children in prostitution, pornography, the production or processing of drugs are also noted as being in "worst forms" of child labor. Article 3 (d) is the most ambiguous part of the convention as it allows worst forms to include "work which, by its nature or the circumstances in which it is carried out, is likely to harm the health, safety, or morals of children." Article 4 of the convention is explicit that it is up to individual countries to define what types of work are considered "worst forms" of child labor under this clause. Activities labeled "worst forms" under Article 3(d) of C182 are often labeled as "Hazardous forms of child labor." The companion recommendation document for C182, R190 Worst Forms of Child Labor Recommendation, suggests that these hazardous forms of child labor include:

> "(a) work which exposes children to physical, psychological, or sexual abuse; (b) work underground, under water, at dangerous heights, or in confined spaces; (c) work with dangerous machinery, equipment and tools, or which involves the handling or transport of heavy loads; (d) work in an unhealthy environment which may, for example, expose children to hazardous substances, agents or processes, or to temperature, noise levels, or vibrations damaging to their health; (e) work under particularly difficult conditions such as work for long hours or during the night or work where the child is unreasonably confined to the premises or the employer." (R190, Section II.3.a-e).

The ILO's SIMPOC estimates that a total of 8.4 million children are involved in child trafficking, in forced or bonded labor, are soldiers, are prostitutes or involved in pornography, or participate in illicit activities (ILO, 2002). 68 percent of these children are in bonded or forced labor. The same SIMPOC study calculates that 170.5 million children are engaged in activities that have been labeled hazardous in their home country. Altogether then, SIMPOC estimates that 178.9 million children are engaged in worst forms of child labor

A simple analytical model based on Dessy and Pallage (2005) and Rogers and Swinnerton (2008) will help fix ideas in our discussion of the determinants of entry into worst forms. The first issue that any framework of child time allocation must address is the question of who is making decisions. Agency in work decisions is important when discussing worst forms of child labor. This study does not inform about agency issues, so the present discussion of why children participate in worst forms is framed around an agent making an informed decision about job type without clarifying who the relevant agent might be. A child participates in a worst form of child labor when the decision-making agent's utility is higher than when the child does not:

$$u\left(y_c, c\right) + e_c \geq u\left(y_0, 0\right) + e_0 \qquad \text{(eq. 2.1)}$$

c is an indicator for whether the child participates in the given worst form, y_c is the agent's income when the child participates in the worst form, and y_0 is the agent's income when the child does not. e_c and e_0 are stochastic, mean zero, error terms that reflect some randomness in the agent's decisions.

Let the decision-maker's utility when the child does not participate in the worst form be represented by an indirect utility function:

$$u\left(y_0, 0\right) = v\left(y_0, p\right)$$

The agent's relevant income when the child participates in a worst form is the agent's income absent the child's participation plus the net economic gain from having the child in the worst form:

$$y_c = y_0 + w^* \qquad \text{(eq. 2.2)}$$

An alternative to participating in a worst form (reflected in y_0) is participation in other types of work. w^* is the premium the worst form pays above those other forms of work.

For analytical clarity, treat the disutility from having the child involved in a worst form as additively separable from the utility owing to the other decisions the agent makes with their income :

$$u\left(y_c, c\right) = u\left(y_0 + w^*, c\right) = v\left(y_0 + w^*, p\right) - \tau \qquad \text{(eq 2.3)}$$

The disutility of participation in a worst form τ is known with certainty. Rogers and Swinnerton (2008) emphasize that uncertainty and poor information may be important in explaining why some children end up in worst forms. It is possible to conceptualize τ as a parameter that reflects the risk perceived by the agent and thereby interpret this set-up within their model. The functional form assumption on preferences in equation 2.3 is equivalent to assuming that the disutility that the agent gets from having a child in a worst form is independent from their income (or price) level. Poor families and rich families are made equally worse off by having a child pick through garbage. This does not imply that the marginal utility from having a child in a worst form, relative to not, will be the same for poor and rich families as the marginal utility associated with the child's net economic contribution in the worst form will differ between poor and rich.

The incidence of children involved in worst forms is then:

$$\begin{aligned}\Pr(C=1) &= \Pr\left[v\left(y_0+w^*,p\right)-\tau+e_w \geq v\left(y_0,p\right)+e_0\right]\\ &= \Pr\left[e_0-e_w \leq v\left(y_0+w^*,p\right)-\tau-v\left(y_0,p\right)\right]\end{aligned} \qquad \text{(eq. 2.4)}$$

Define $u=e_0-e_w$. u has a cdf *F(u)* and strictly positive density *f(u)*. Thus:

$$\Pr(C=1)=F\left[v\left(y_0+w^*,p\right)-\tau-v\left(y_0,p\right)\right].$$

We totally differentiate in order to organize the determinants of different risks of being observed in a worst form of child labor:

$$d\Pr(C=1)=f(u)\left(\left[\frac{\partial v_w}{\partial y}-\frac{\partial v_0}{\partial y}\right]dy_0+\left[\frac{\partial v_w}{\partial p}-\frac{\partial v_0}{\partial p}\right]dp-d\tau+\frac{\partial v_w}{\partial y}dw^*\right) \qquad \text{(eq. 2.5)}$$

With diminishing marginal utility of income, ${}^{\partial v_w}\!/_{\partial y}<{}^{\partial v_0}\!/_{\partial y}$. Declines in income opportunities open to the child absent participation in the worst forms, tend to push children towards participation in worst forms. The amount of the push depends on the curvature of the indirect utility function. Higher net income available in the worst form also pulls children towards the activity. The extent of the pull depends on the marginal utility of income. Hence, poorer families are more likely to select into worst forms, because they are poorer. The agent's disutility from the activity is also an influence as are prices.

Testing between this model of entry into worst forms against the victimization model often posited in policy work depends on a comparison of the determinants of entry into other types of child labor. The poor will always select into work, because their marginal utility of income is higher. The victimization model implies that children arbitrarily enter into worst forms because of indifferent or uncaring adult agents. This implies $\tau=0$. In equilibrium, then worst forms should not pay more than other forms of work $w^* = 0$ and the determinants of participation in worst forms should look like any other type of work. Thus, the comparison of the determinants of entry into worst forms and other forms of work is central to gauging the appropriateness of the compensating wage differential model.

III. Methodology - Estimating the Correlates of Selection

Why are children engaged in worst forms of child labor (WFCL)? Empirically, this is a hard question to answer, because WFCL are relatively rare. The probability that random sampling captures children engaged in any given WFCL is typically low. Hence, data collection can be prohibitively costly and statistical power is always a concern. Researchers have had to turn to other data sources. The most common approach is inherently qualitative. Researchers find children engaged in a worst form and interview them to find out about their circumstances. Sometimes these interviews are unstructured, but often researchers follow a survey questionnaire which can permit quantitative analysis.

It is impossible to learn about why children are in worst forms from only interviewing children in worst forms. Consider some factor D that influences selection into activity C. The researcher is interested in knowing how factor D increases the probability that a child with other characteristics X enters into activity C. When D is discrete, this is: $P\left[C=1 \middle| D=1, X\right] - P\left[C=1 \middle| D=0, X\right]$. Neither probability can be computed in the set of C=1. Put another way; let's say a child in an interview remarks that they are engaged in activity C because of factor D ("I am a ragpicker, because my dad lost his job"). There may be lots of children that experience factor D that do not select into C (lots of children have parents become unemployed without becoming ragpickers), but without data on children not in C there is no way to compute the increased chance of engaging in C with a change in D.

The problem of drawing inference about rare events is not unique to worst forms of child labor. Most observational inference in medicine is made under precisely these circumstances, and this study applies these approaches to rare events from epidemiology to the study of selection into worst forms of child labor. These techniques do not appear to have been applied to the analysis of selection into worst forms before. The present discussion draws heavily from papers such as Prentice and Pyke (1979), King and Zeng (1999), and Manski (2001).

Let C_i be an indicator that child i is involved in the given worst form of interest. D_i is the covariate of interest. In the present discussion, D_i is binary, but the discussion generalizes to when D_i takes more than two values. Our interest is in estimating the impact of D_i on the probability that child i is involved in the given worst form. Later attention will be placed on estimating this probability conditional on other confounding variables that are correlated with both C_i and D_i.

There are three main outcomes of potential interest.

1. Absolute risk. How likely is an individual with D_i to be involved in the given worst form:

$$\pi_i = \Pr\left(C=1 \middle| D_i\right). \qquad \text{(eq. 3.1)}$$

2. Relative risk. How much more likely is a child with D=1 to be observed in activity Y than a child with D=0:

$$R = \frac{\Pr\left(C=1 \middle| D=1\right)}{\Pr\left(C=1 \middle| D=0\right)}. \qquad \text{(eq. 3.2)}$$

3. Attributable risk. How much does an individual's risk of engaging in Y increase with a change in D from 0 to 1:

$$A = \Pr\left(C=1 \middle| D=1\right) - \Pr\left(C=1 \middle| D=0\right). \qquad \text{(eq. 3.3)}$$

Each of these outcomes is potentially of considerable interest to researchers and policy. For example, absolute risk is of interest to assess how likely a child with a given characteristics is to be in a WFCL. An index of vulnerability to WFCL would be constructed entirely through combining measures of absolute risk. Researchers interested in how participation in WFCL differs with variation in observable characteristics will be most concerned with relative or attributable risk. Relative risk is the most straightforward to estimate. However, relative risk can often be misleading to interpret in the context of low probability events. For example, suppose that the probability of observing a child in a WFCL is extremely low when a certain characteristic D is not present (e.g. 0.00001) and suppose the probability is higher when the characteristic is present (e.g. 0.0001) but still so small as to not be substantive. Estimates of relative risk in this case would be very large (10) even though the probabilities are so small as to not be substantive. Hence, at a minimum, relative risk should never be considered without attention to the baseline absolute risk. In contrast, attributable risk gives a direct measure of how much a child's risk of being involved in a WFCL changes with an observed characteristic. Consequently, it is the outcome of interest most often.

Estimating absolute, attributable, or relative risk requires data on both cases (subjects where C=1) and controls (C=0). When data on both cases and controls can be collected in a single randomized survey, standard cohort comparison techniques are available. However, typically the incidence of most forms of WFCL is such that a survey would need to be extremely large in scale to recover engaged children using random sampling. Thus, a more common situation is to have separately collected data on children not engaged in the WFCL (the control data) and data on children engaged in the activity (the case data). The case data do not need to be obtained through randomized sampling, but estimating absolute or attributable risk requires knowledge of the probability a child engages in the WFCL. This is most easily assessed if the case data collection is designed, in part, to estimate this parameter. Moreover, whatever sampling procedure generates the case data, sampling must be independent of the covariates *D* of interest except in as much as D is correlated with selection into the case data. Put another way, the data generation process can generate bias if it is correlated with covariates of interest for reasons other than that the covariates are correlated with selection into worst forms.

Ideally, the survey instrument used to collect data on the case and control populations will be identical. In practice, it is rare that similar case and control data exist. It will only be possible to compute any of the risk parameters of interest for covariates that appear in both the case and control data. Moreover, a common problem is that even when there are similar questions, the case and control data will be answered by different people. Often case data is collected by interviewing children while most household surveys and censuses (typical sources of control data) interview household heads or their spouses. Biases from differences in respondents can be as substantive as biases from different framing of questions, and these dissimilarities make it very challenging to assess the risk parameters of interest.

The classic case – control approach makes the rare events assumption to estimate relative risk. That is, the case and control data are pooled, and it is assumed that the probability of observing a case individual tends to zero (conditional on observed characteristics) in the limit. This assumption allows the researcher to interpret the odds ratio from a logit of participation in the WFCL on observable characteristics as an estimate of relative risk. The appeal of this approach is that it is possible to estimate relative risk without identifying absolute risk in the population. However, the rare events assumption is problematic. The existence of case data implies that the probability of observing a case is not zero, and the rare events assumption implies that attributable risk is zero.

Knowledge of the probability of observing a case in the population substantially improves estimation. Let λ denote the incidence of the worst form in the population, and let $\bar{C}$ be the fraction of the case –control pooled data that is from the case data. In order to estimate parameters such as absolute risk and attributable risk, the constant from the logit needs to be corrected to reflect the difference between λ and $\bar{C}$. Specifically, the regression's intercept in the logit β_0 needs to be adjusted as :

$$\beta_0 - \ln\left[\left(\frac{1-\lambda}{\lambda}\right)\left(\frac{\bar{C}}{1-\bar{C}}\right)\right] \qquad \text{(eq. 3.4)}$$

This result is attributable to Manski and Lerman (1977) or Prentice and Pyke (1979). The intuition behind this adjustment is that in general the ratio of case to control observations in the pooled data will not correspond to the ratio expected in the population. Hence, predicted probabilities from the pooled regression would not reflect the true, underlying population prevalence of the worst form. Rescaling the intercept as in equation 3.4 assures that the predicted probabilities match what could be observed if all of the population data existed.

Sometimes researchers will not have an estimate of the incidence rate in the population. Theoretical work such as Manski (2001) considers cases where there is no prior information about the range of plausible values of λ. However, at a minimum, researchers will have some idea of a plausible range of values for incidence in the population. Let λ_L and λ_H indicate the lower and upper values of the plausible range of λ. King and Zeng (1999) suggest computing bounds on possible values of absolute and relative risk by estimating either at both λ_L and λ_H. Because absolute and relative risk are positive monotone functions of λ, computing either at the lower and upper values of λ defines bounds on the range of possible absolute and relative risks.

Attributable risk is more difficult, because it is not a positive monotone function of λ. Define $A(\lambda_k)$ as the estimate of attributable risk associated with an estimated incidence of λ_k. King and Zeng (1999) suggest checking for whether λ_L and λ_H are in a monotone region of attributable risk by evaluating whether attributable risk appears to have the same derivative with respect to λ at both its high and low values. This can be checked by verifying that the signs of $A(\lambda_{L+\varepsilon}) - A(\lambda_L)$ and $A(\lambda_{H+\varepsilon}) - A(\lambda_H)$ are the same. When λ_L and λ_H are in a monotone region of A, then bounds can be calculated as:

$$A \in \left\{\min\left(A(\lambda_L), A(\lambda_H)\right), \max\left(A(\lambda_L), A(\lambda_H)\right)\right\}. \qquad \text{(eq. 3.5)}$$

Sometimes, population prevalence rates will be in a non-monotone region of attributable risk. In this case, King and Zeng (1999) show that bounds on attributable risk are given by:

$$A \in \left\{\min\left(A(\lambda_L), A(\lambda_0), A(\lambda_H)\right), \max\left(A(\lambda_L), A(\lambda_0), A(\lambda_H)\right)\right\}. \qquad \text{(eq. 3.6)}$$

where $A(\lambda_0) = \frac{\sqrt{\omega}-1}{\sqrt{\omega}+1}$ and ω is the odds ratio:

$$\omega = \frac{\Pr(C=1|D=1)\Pr(C=0|D=0)}{\Pr(C=0|D=1)\Pr(C=1|D=0)}. \qquad \text{(eq. 3.7).}$$

In what follows, econometric work focuses on presenting estimates of attributable risk. Attributable risk is the focus of this study, because the primary aim of the empirical work is to isolate indicators associated with increased risk of participating in a worst form. Hence, the size of the increased risk associated with a given factor of interest is important. In the present case, attributable risk is computed using prior correction (eq. 3.4) to compute absolute risks and in turn compute the difference. All empirical work is implemented using the regression code available freely from King and Zeng (2001).

IV. Data

There are an estimated 127,143 children engaged in the worst forms of child labor in Nepal (ILO 2001). There are approximately 8 million children below the age of 16 in Nepal, and the ILO estimates that 1.5 percent of these children work in these worst forms of child labor. This study uses a survey of short-route porters, ragpickers, and the population census to consider the correlates of selection into worst forms of child labor in Nepal for children age 10-14.

A. Porters

For many areas of Nepal, porters are critical for transporting consumer goods, getting business output for market, and delivering construction materials to remote areas. Porters are typically classified as long route and short route porters, and the two types of porters appear to be somewhat segmented. This study focuses on short route porters. Short route porters are typically contracted at spot markets in local markets and bus parks.

Portering is considered a worst form of child labor, because children often carry heavy loads, across difficult terrain, for long hours. In the data used in this study, short route porters report working approximately 10 hours per day for 6 days a week on average (KC et al 2001a). Two thirds of short route porters report averaging roughly 10 routes per day that range in weight from 10 to 50 kilograms (although one has to be suspect about self-reported load weights). Sixty percent report not wearing protective gear such as boots, gloves, or pads on the head.

The short route porter (SRP) survey was conducted in urban areas of Nepal as that is where short route portering is concentrated. The SRP survey sampled work sites: markets and bus parks.

Out of an estimated 423 market centers and bus stops, a random sample of porters was interviewed in 97 randomly selected market centers and 15 randomly selected bus parks. When appropriately weighted, the SRP survey suggests a total of 5,087 short route child porters age 6-17 in Nepal in 2003. A total of 30 of these are below the age of 10, and most are age 14 or more. In the present study, we focus on short route porters age 10-14. There are an estimated 1,404 short haul porters age 10-14 in urban Nepal in 2003.

Column 1 of Tables 1 - 4 provide summary statistics on the 164 porters age 10-14 captured in the SRP. Table 1 summarizes child characteristics. Table 2 describes the child's background (information about where the child's family lives). Table 3 contains information on the child's father. Table 4 details the information on the child's mother. All information in the SRP is collected from interviewing the child porter. Means and standard errors are reported. Both are corrected for sample design and weighted to be nationally representative.

Porters are mostly boys. Less than one in five attend school. Less than 70 percent say they are literate. They come from the Hills and the Terai and from the Central and Western areas. Their parents are relatively old, illiterate, and much more likely to be disabled than other populations.

B. Ragpickers

Ragpickers collect rags and other used goods to be recycled and reused. As an activity, ragpicking is primarily an urban activity. Adult and child ragpickers collect plastics, polyethylene, bottles, metals, and tins from dumping sites, streets, river banks, etc. These collected materials are sold to junkyards and shops which in turn sell these materials to suppliers for recycling. Ragpicking is nearly universally viewed as a worst form because of the extremely hazardous work environment (KC et al 2001b).

The ragpickers survey (RAG) was conducted in urban areas of Nepal. The original survey design was to sample sites where ragpickers worked. However, researchers found it difficult to interview children in dumping areas, garbage disposal and refuse areas, slums, and river banks and faced additional difficulties associated with the mobility of ragpickers. Thus, while the survey was being fielded, enumerators abandoned the original sample frame and interviewed children in junkyard shops or locations where they spend their leisure time (Mhukherjee 2003).

The nonrandom nature of the survey and this disconnect between sample design and survey implementation creates an unknowable array of problems for inference and makes it impossible to know whether estimates of the incidence of ragpicking from this data are accurate. If one is willing to treat the RAG data as if it were based on random sampling of job sites, it is possible to make inferences about the scope of ragpicking in Nepal. That said, the survey suggests that there are 3, 695 child ragpickers age 6-18 in Nepal in 2002. 974 of these ragpickers are age 10-14.

Column 2 of tables 1 - 4 provides summary statistics for the 372 children 10-14 interviewed in the RAG survey. They are mostly boys. One in ten are in school. Less than half say they can read and write. They come from the central hills and Terai. Their parents are more likely to be literate than Porters, but their parents are also far more likely to be disabled.

C. The Census

The population and housing census of 2001 is used for the control sample. It precedes RAG by a year and SRP by two, but it contains considerable overlap in content. In using the earlier sample as a control, the analysis herein is based on the assumption that there are not substantive changes over time in any of the characteristics that are the focus of the analysis. For example, if the civil war in Nepal lead to a substantive rise in disability between 2001 and 2002, then one might misattribute the time trend in disability in Nepal as a correlate of selection into ragpicking. The close proximity in time between the census and the two targeted survey mitigates this concern, but it would be ideal to have the control data simultaneous with the targeted surveys.

There are several substantive issues that arise from using the Census as a control sample beyond timing. First, it is only possible to study children 10-14. Information on economic activities is not collected for children below age 10. Education data are incomplete on children above age 14 in the census because of an odd skip pattern in the questionnaire.

Second, it is only possible to discern familial relationships for children of the household head. This introduces non-random selection bias into the control sample by eliminating children who live in households where a parent is not codified as the head. This could be a problem for inference if whether a parent is coded as a household head in the census is correlated with selection into portering or ragpicking and other observable household characteristics. Parental death is one potential concern. A child who has experienced a parental death is less likely to have a parent coded as the household head (mechanically, they have one fewer parent who could be a household head). If parental death is associated with other background characteristics and selection into a worst form, any estimates of attributable risk associated with the background characteristics could be severely biased. Unfortunately, there appears to be no obvious solution to this problem.

Third, it is not possible to identify porters and ragpickers in the census. Occupation and Industry of employment are only reported at the 1 digit level. It is technically possible that our analysis is biased by miscoding some "case" observations as "control". This would diminish our ability to identify correlates of entry into these worst forms. Given that the public use microdata is a (11 percent) subsample of the entire census and that portering and ragpicking are rare, it is unclear how substantive this bias is likely to be.

Fourth, the census data is collected from the head of the household. Thus, the respondent in the census is different than the respondent in the SRP and RAG, who are children 10-14. It is possible that this difference in respondent's creates a source of bias. For example, in tables 3 and 4, all porters respond that their parents have some school while only 18 percent of wage working children have parents with some school. This may reflect differences in interviewer instructions, the exact wording of the questionnaire, or differences in the respondent. Ultimately, it is impossible to discern when differences in data between surveys reflects substantive differences versus any of these other issues.

Columns 3 - 5 of tables 1 - 4 provide summary statistics for children 10-14 of the household head in the public use census micro-data. Census children are trifurcated into children involved primarily in wage work, primary in home enterprise work, or no form of work. This latter category includes children who work in domestic service in their own home, children who are inactive, and children are primarily students. For each category of activity, means and standard errors are reported. Both are corrected for sample design and weighted to be nationally representative.

The analysis in this study will compare the correlates of entry into worst forms with the correlates of entry into wage work. The choice of wage work for benchmarking is driven by the fact that, like portering and ragpicking, the activity takes place outside of the child's home. Other research on child labor has suggested that work inside and outside of the home can be treated very differently (e.g. Edmonds 2008). Hence, it is more interesting to compare findings on ragpickers and porters to wage workers rather than home enterprise workers, for example. In fact, though magnitudes differ substantively, the flavor of the basic comparisons would be similar if all working children were used as a benchmark rather than just wage workers. The correlates of selection into worst forms looks distinct compared to both wage working children and children in family businesses or farms.

V. Main Findings

A. Porters

A comparison of the descriptive statistics of porters to wage works in table 1 -4 highlight many similarities and differences. Table 1 summarizes child characteristics. Short route porters differ from wage workers in that they are more likely to be high status, more likely to speak Nepali, and less likely to be Muslim. These differences likely reflect that short route porters are only interviewed in urban areas, and tend to be from those some areas. That is, most short route porters in the survey are not in-migrants to urban areas. Thus, populations of rural origin (such as Muslims or non-Nepali speaking populations) are less present in the SRP. The reported completed schooling of porters is also higher than other wage workers. Short route porters differ from wage workers in that they are more likely to be high status, more likely to speak Nepali, and less likely to be Muslim. These differences likely reflect that short route porters are only interviewed in urban areas, and tend to be from those some areas. That is, most short route porters in the survey are not in-migrants to urban areas. Thus, populations of rural origin (such as Muslims or non-Nepali speaking populations) are less present in the SRP. The reported completed schooling of porters is also higher than other wage workers.

Background characteristics are in Table 2. Porters are more likely to be from hill areas than wage workers. It is not surprising that porters are more prominent in hill areas, less prominent in plains, as the road infrastructure around the Terai's mid sized cities are generally better than in the hill areas. Moreover, porters are more active in the western development region of Nepal than are wage-workers. This likely reflects the fact that short haul porters often work around bus stations and larger markets, which are more prevalent in the central and west regions of Nepal than elsewhere.

Compared to wage workers, porters seem to come from relatively disadvantaged backgrounds. Literacy among both fathers and mothers is lower for porters than wage workers. Porters report higher levels of paternal schooling completion than paternal literacy, so either there is data error on the coding of paternal education or there are lots of illiterate fathers of porters who have completed primary school. The maternal education data are consistent with the observed lower maternal literacy rates for porters than wage workers. Porters are also more likely to report both a father or a mother that is disabled, and to report a mother who is working.

Paternal and maternal disability and maternal wage work stand out as strong predictors of selection into portering. The first column of table 5 contains estimates of attributable risk for each listed row characteristic separately. Attributable risk is computed as described in the methodology section, following King and Zeng (2001). Specifically, the census and SRP data are pooled. An indicator that a child is a porter is regressed (using a logit) on age, gender, ethnicity, language, belt, development region, and, in column 1, the variable indicated by the row. The logit is estimated using prior correction (eq. 3.4) with a bias correction for small samples. Attributable risk is then computed by estimating the differences in absolute risk level as computed with equation 3.3. Standard errors are corrected for clustering owing to sample design. In column 4, the conditional attributable risk estimates are computed by including all of the listed controls in the logit and holding all listed observable characteristics constant (at their mean) except for the variable specified by the row.

Paternal disability is the largest predictor of selection into portering. Paternal disability raises the probability a child is observed portering by more than a tenth of a percent. Maternal disability has a similar positive association with portering. Another strong indicator factor associated with an elevated risk of being a porter is having a mother working for wages. Female wage work is relatively rare in Nepal, so that this observation might reflect something about the geographic location of the control population relative to the portering population. It is also consistent with the idea that women only enter the labor market when the family's marginal utility of income is very high. Hence, the wage work observation might be consistent with a view that poverty is critical in explaining selection into portering.

The attributable risk estimates in table 5 are not causal estimates of how selection into portering will be affected by changes in any of the listed observable characteristics. Rather, they describe how the likelihood of observing a child porter varies with changes in maternal or paternal characteristics. Table 6 contains estimates of attributable risk for becoming a porter associated with changing several of the covariates from Table 5simultaneously. For example, a mother who is disabled and not working raises the probability a child is observed as a porter by nearly 0.2 percentage points for a landless household (nearly double the risk observed in a household with land). In general, landless households are more likely to be observed sending children to porter in the context of a paternal or maternal disability or if both mother and father are observed working.

A comparison of attributable risk estimates in table 6 to that observed for wage work in table 7 is illustrative. The patterns observed with disability and literacy are similar for portering and other types of wage work. The main difference with portering is that the presence of self employment in the household lowers the risk of portering (while it raises the risk of observing a child in wage work). Hence, the portering data at least contain some suggestion that the availability of employment within the household may be associated with a diminished risk of seeking work in portering outside of the family. It is important to note, however, that the magnitudes of the observed changes in attributable risk with home enterprises are very small.

B. Ragpickers

Ragpickers also appear distinct from wage workers. Among the child characteristics described in table 1, ragpickers tend to be younger than wage workers, and they are much less likely to be ethnic Tharu. The fact that ragpickers are younger is consistent with a role for employment opportunities in selection into ragpicking as young children have fewer formal wage earning opportunities. Ragpickers also appear to be relatively more educated although it seems likely that this difference with the census might reflect biases owing to who responds to the questionnaire. Ragpickers are less likely to be higher caste than the general population, and less likely to be Tharu. The low incidence of Tharu ragpickers is interesting. Two possible explanations seem obvious. First, ragpicking may be more common in places where the Tharu are less prevalent. Second, desperate Tharu may have better options than ragpicking. Bonded labor is common in the Tharu population, and one interpretation of their lower incidence of ragpicking is that the disamenities associated with accepting bondage are not as bad as those associated with a life of ragpicking.

In table 2, ragpickers appear more likely to be from hill areas than are wage workers and are more likely to be from central Nepal. The concentration of ragpickers is consistent with the location of the large recycling centers which are especially prevalent in the Kathmandu Valley (central-hill). However, this is also where trash is especially concentrated because of the population density. Hence, one should not infer that the presence of the recycling industry is the reason why there are ragpickers in the Valley. Of course, if there was no market for their output, it seems unlikely children would pick through trash except to help meet basic needs.

Ragpickers are also less likely to come from households that own farmland. This observation is consistent with the view that a lack of alternative income generating strategies may play an important role in selection into ragpickers. To some extent this seems obvious as its hard to imagine that picking through trash and debris is ever someone's first choice for income. However, it is easy to over interpret this correlation between farmland and ragpicking. Children working for wages are less likely to own farmland than children who work in family enterprises (like farms). Moreover, a lack of land may be correlated with fewer at home employment opportunities but it also may be correlated with a lack of income.

Several parental background characteristics in tables 3 and 4 suggest that selection into ragpicking is correlated with having a relatively disadvantaged background. Maternal literacy is lower than wage workers and both mothers and fathers of ragpickers are less likely to have some post primary education. Moreover, parental disability is a strong correlate of ragpicking (as has been observed with porters as well). Four percent of ragpickers have a disabled father, and 1 percent of ragpickers have a disabled mother. For contrast, less than one tenth of one percent of the general population has a disabled father.

Also, ragpickers are less likely to have a parent who owns a small business or is employed in agriculture. While 63 percent of children in wage work have a father who works in agriculture, less than 9 percent of ragpickers do. Forty-eight percent of wage earning children have a mother in agriculture. Less than 8 percent of ragpickers have a mother engaged in agriculture. It is impossible to discern whether this reflects the employment opportunities open to the children, the family's disadvantaged background, or something transitory in the child's family's economic environment. However, the differences in the means are not present in other activities.

Table 8 provides estimates of attributable risk by observable background and family characteristic. It is constructed identically to table5. Paternal disability stands out as the largest predictor of selection into ragpicking. Less than three hundredths of a percent of children 10-14 are engaged in ragpicking, but paternal disability raises the probability that a child is observed in ragpicking by nearly two tenths of a percent. While no other characteristic is as strong a predictor as paternal disability, the observation that the child's family's employment background is an important risk factor persists in the attributable risk estimates. Either owning agricultural land or maternal or paternal work in agriculture substantially lowers the odds of observing a child in ragpicking. This may reflect differences in location rather than the household's employment opportunities, but the fact that maternal self employment also is associated with a diminished risk of observing a child as a ragpicker suggests that at least some part of why these are risk factors may owe to employment opportunities.

Estimates of changes in attributable risk are generally uninformative in the conditional specification. The one exception is with regards to paternal disability, because that is such a large predictor of selection into ragpicking. In table 9, observing a disabled father significantly increases the risk that a child is observed ragpicking, and this increased risk of ragpicking is larger for the landless than for children who come from families with land. The larger magnitudes estimated for landless families are consistent with the descriptive data which also suggest a link between selection into ragpicking and employment opportunities. However, in general, there are few observable characteristics other than paternal disability which can predict a risk of ragpicking. This suggests that most of the determinants of selection into ragpicking are outside the scope of the available data.

Another important reason why the attributable risk of ragpicking is so small is that ragpicking is estimated to be extremely rare (less than three hundredths of a percent of children 10-14). Section 3 discussed how to estimate bounds on attributable risk when the incidence of a worst form is uncertain. Table 10 implements this methodology. The incidence of ragpicking is assumed to vary between 0.03 percent and 0.3 percent. Thus, the estimates from table 9 are used for one bound and attributable risks are recalculated assuming an incidence of three tenths of a percent to form the other bound. The data pass the test for positive monotonicity suggested in section 3. Table 10 contains bounds on attributable risk for landless households.

Contrasting table 9 and Table 10 highlights how important estimates of baseline incidence are for computing attributable risk. In very low probability events, it is a challenge to capture covariates that substantially increase the risk of the child entering the worst form simply because the event itself is rare. In general, the patterns recovered by the bounds estimates in table 10 suggest risk factors for entry into ragpicking that are similar to that observed for portering and different with regards to self employment from what was observed in table 7 for wage work.

V. Conclusion

This study illustrates an approach to study the correlates of participation in a worst form of child labor. Survey data on the background characteristics of children engaged in worst forms of child labor are combined with nationally representative data on those same background characteristics. With this combination of data, it is possible to calculate what characteristics are associated with an increased risk of engaging in a worst form of child labor. When combined with data on the incidence of the worst form in the population, it is possible to compute how large of an increased risk of involvement in a worst form can be attributed to variation in a characteristic. This simple, descriptive comparison sheds some light on how the background characteristics of children engaged in portering and ragpicking in Nepal differ from the general population of children in Nepal.

Are worst forms different than other more common forms of employment from the perspective of the agent who makes decisions about child time allocation? The data are consistent with the view that worst forms are different. Most theoretical treatments of entry into worst forms posit that children are more likely to enter worst forms when their alternative employment opportunities are limited. A child is more likely to participate in a worst form when the net economic return is larger. The data suggest that children are more likely to be involved in wage work when there is a family enterprise. This could reflect a causal impact of the family's business, or it might reveal that family's are more apt to own businesses in locations with more active labor markets. However, children are less likely to engage in work as ragpickers and porters when there is a family business at home. This association could reflect something about the impact of a family enterprise on entry into worst forms through the value of child time in the family business or the enterprise's correlation with family incomes. Alternatively, the association between family enterprises and entry into worst forms might owe to an association between family enterprises and the overall local labor market (as speculated with regards to wage work). Most porters and ragpickers are working in the same geographic location as their parents. Thus, if omitted labor market characteristics were driving the finding that home enterprises are associated with a reduced risk of participation in a worst form, it is surprising that the ragpicker and porter patterns would differ from that observed for wage work. The idea that the association between home enterprises and entry is driven by either the potential economic contribution of the child to its household or household living standards is more compelling.

There are some further associations in the data that are consistent with the idea that the child's employment opportunities in their household cast an important influence on entry into worst forms. Households with porters and ragpickers are less likely to own agricultural land, although this association is not particularly robust for these two populations. Portering is most prevalent in areas where there is the most need for porters as ragpickers are most prevalent in areas where there is trash and a recycling industry. Maternal wage work also seems to predict portering, and self-employment is negatively correlated with ragpicking. However, all of these characteristics predict only a small amount of the observed prevalence of each worst form.

Parental, especially paternal, disability stands out as a strong predictor of observing a child in a worst forms in Nepal. Relative to wage working children, porters are 5 times and ragpickers are 4 times more likely to report that their father is disabled. This association between paternal disability and entry into worst forms could reflect that children are more vulnerable to victimization when their father is disabled, but their father is still living and there is little correlation between paternal and maternal disability in the data. Paternal disability also does not appear to be strongly associated with some particular source location for the child; it is not likely to be capturing omitted geographic factors. Moreover, the magnitudes are so much larger than what is observed for any individual measure of parental self-employment or other household economic activity, it seems likely that paternal disability reflects more than an association between paternal disability and employment opportunities open to the child within its own household (which are conditioned on in the empirical work). It seems most plausible that the strong association between paternal disability and entry into worst forms reflects that paternal disability is strongly correlated with the child's family being substantially poorer. If this interpretation is correct, then the data support Dessy and Pallage's (2005) model of partially compensated wage differentials for worst forms of child labor.

The methodology used to assess the correlates of selection into worst forms is general, but its data requirements are not trivial. Namely, four conditions must be met:

1. The type of work that qualifies as a worst form is explicitly identified
2. Reasonable estimates of the incidence of that worst form exist in the population

3. There are individual level data on background characteristics of children engaged in the worst form available
4. There are nationally representative data on the same set of background characteristics available for the general population.

Unfortunately, the data on children in worst forms and the representative data used in this study are not perfectly consistent in how they collect information, and there is limited information that is in common in the targeted surveys and the nationally representative data. This problem is easily resolved if future survey work on children in worst forms would merely be attentive to existing data resources, and design their survey work to be in part consistent with nationally representative data. Even better of course, would be to integrate target surveys into a broader national survey program and to combine that effort with scientific evaluation of interventions aimed at children engaged in worst forms of child labor..

VI. References

Baland, J.M. and Robinson, J.A. (2000), "Is Child Labor Inefficient?" Journal of Political Economy. August, Vol. 108 No. 4, pp. 663-79.

Basu, K., and Van, P. H. (1998), "The Economics of Child Labor." American Economic Review, Vol. 88 No. 3, pp.412-27.

Dessy, S. and Pallage, S. (2005), "A Theory of the Worst Forms of Child Labour", Economic Journal, Vol. 500 No. 1, pp 68-87.

Edmonds, E. (2007), "Child Labor", in Shultz, T.P. and Strauss, J. (Eds.), Handbook of Development Economics Volume 4, North-Holland, Amsterdam, pp. 3607-3709.

Edmonds, E. (2008). Defining Child Labour: A Review of the Definitions of Child Labour in Policy Research, International Labour Organization - IPEC, Geneva.

International Labour Organization (2001), The Time Bound Program in Nepal, International Labour Organization, Kathmandu Nepal.

KC, B.K., Adhikari, K.P, Subedi, G. and Gurun, Y.B. (2001a), "Situation of Child Porters: A Rapid Assessment", Investigating the Worst Forms of Child Labour Series, No.6. ILO-IPEC, Geneva.

KC, Bal Kumar, Gurung, Y.B., Adhikari, K.P. and Subedi,G. (2001b), "Nepal, Situation of Child Ragpickers: A Rapid Assessment", Investigating the Worst Forms of Child Labour, No.4. ILO-IPEC, Geneva

King, G. and Zeng, L. (2001), "Logistic Regression in Rare Events Data", Political Analysis, Vol. 9 No. 2, pp. 137-163.

Lancaster, T. and Imbens, G. (1996): "Case Control Studies with Contaminated Controls", Journal of Econometrics, Vol. 71 No. 1 , pp. 145-160.

Manski, C. (2001), “Nonparametric Identification Under Response-Based Sampling”, in Hsiao, C., Morimune, K. and Powel, J. (Eds.), Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya, Cambridge University Press, New York, pp. 241-258.

Manski, C. and Lerman, R. (1977), "The Estimation of Choice Probabilities from Choice-Based Samples," Econometrica, Vol. 45 No. 8, pp. 1977-1988.

Mhukherjee, S. (2003). Child Ragpickers in Nepal: A Report on the 2002-2003 Baseline Survey, International Labour Organization, Kathmandu Nepal.

Prentice, R. L., and Pyke, R. (1979), “Logistic Disease Incidence Models and Case-Control Studies”, Biometrika, Vol. 66 No. 3, pp. 403–411.

Rogers, C. and Swinnerton, K (2008), "A Theory of Exploitative Child Labor", Oxford Economic Papers, Vol. 60 No. 1, pp. 20-41.

Tomz, M, King, G., and Zeng, L. (2003), "ReLogit: Rare Events Logistic Regression", Journal of Statistical Software, Vol. 8 No. 2, pp. 1-27.

Table 1: Child Characteristics in Porters Survey, Ragpickers Survey, and Census

Mean/SE	Short Route Porters Survey (1)	Ragpickers Survey (2)	2001 Population and Housing Census: Wage Work (3)	2001 Population and Housing Census: Home Enterprise Work (4)	2001 Population and Housing Census: Not Work (5)
# of observations	164	372	6,900	25,390	297,506
Estimated population size	1,404	974	63,143	254,290	2,592,568
Age	13.0	12.0	12.4	12.3	11.8
	(0.107)	(0.096)	(0.020)	(0.010)	(0.003)
Female	0.282	0.198	0.370	0.612	0.470
	(0.058)	(0.043)	(0.008)	(0.004)	(0.001)
Ethnicity					
High Status Hindu Caste	0.192	0.163	0.094	0.253	0.351
	(0.036)	(0.049)	(0.005)	(0.005)	(0.003)
Tharu	0.126	0.006	0.151	0.062	0.076
	(0.042)	(0.003)	(0.007)	(0.003)	(0.002)
Newar	0.013	0.026	0.027	0.025	0.057
	(0.008)	(0.014)	(0.003)	(0.002)	(0.002)
Dalit	0.285	0.339	0.302	0.202	0.145
	(0.068)	(0.104)	(0.009)	(0.004)	(0.002)
Muslim	0.037	0.080	0.100	0.047	0.034
	(0.023)	(0.059)	(0.007)	(0.003)	(0.001)
Other	0.348	0.384	0.325	0.411	0.336
	(0.058)	(0.071)	(0.008)	(0.006)	(0.002)
Native Language					
Nepali	0.588	0.394	0.222	0.484	0.520
	(0.067)	(0.090)	(0.008)	(0.006)	(0.003)
Tharu	0.109	0.007	0.133	0.052	0.058
	(0.042)	(0.004)	(0.007)	(0.003)	(0.002)
Other	0.303	0.599	0.644	0.464	0.422
	(0.059)	(0.089)	(0.010)	(0.006)	(0.003)
In School	0.190	0.100	0.159	0.271	0.864
	(0.054)	(0.026)	(0.006)	(0.005)	(0.001)
Can read and write	0.687	0.450	0.272	0.381	0.875
	(0.046)	(0.068)	(0.008)	(0.005)	(0.001)
Completed Some School	0.869	0.944	0.185	0.291	0.822
	(0.056)	(0.019)	(0.007)	(0.005)	(0.002)
Completed Std. 5	0.159	0.067	0.063	0.105	0.346
	(0.041)	(0.020)	(0.004)	(0.003)	(0.002)
Completed Post Primary	0.085	0.011	0.025	0.049	0.191
	(0.034)	(0.008)	(0.002)	(0.002)	(0.002)

Sample restricted to children age 10-14.

Table 2: Background Characteristics in Porters Survey, Ragpickers Survey, and Census

			2001 Population and Housing Census		
Mean/SE	Short Route Porters Survey (1)	Ragpickers Survey (2)	Wage Work (3)	Home Enterprise Work (4)	Not Work (5)
Belt					
Hill	0.503	0.540	0.191	0.495	0.462
	(0.086)	(0.190)	(0.011)	(0.007)	(0.005)
Terai	0.497	0.460	0.789	0.368	0.478
	(0.086)	(0.190)	(0.011)	(0.006)	(0.005)
Region					
East	0.131	0.165	0.312	0.192	0.230
	(0.054)	(0.120)	(0.011)	(0.005)	(0.004)
Central	0.300	0.616	0.406	0.306	0.335
	(0.066)	(0.175)	(0.012)	(0.006)	(0.005)
West	0.363	0.177	0.107	0.155	0.223
	(0.087)	(0.122)	(0.006)	(0.005)	(0.004)
Mid-West	0.172	n/a	0.112	0.188	0.117
	(0.078)		(0.007)	(0.005)	(0.003)
Far-West	0.034	0.041	0.063	0.158	0.094
	(0.018)	(0.044)	(0.006)	(0.005)	(0.003)
Household Background					
Owns Farmland	0.671	0.397	0.508	0.934	0.821
	(0.075)	(0.077)	(0.010)	(0.002)	(0.005)

Sample restricted to children age 10-14.

Table 3: Paternal Characteristics in Porters Survey, Ragpickers Survey, and Census

			2001 Population and Housing Census		
	Short Route Porters Survey	Ragpickers Survey	Wage Work	Home Enterprise Work	Not Work
Mean/SE	(1)	(2)	(3)	(4)	(5)
Reports Characteristics	0.847	0.871	0.910	0.906	0.888
	(0.044)	(0.020)	(0.004)	(0.002)	(0.001)
Age	48.779	44.110	43.855	45.175	44.717
	(1.696)	(0.735)	(0.196)	(0.099)	(0.040)
Can Read and Write	0.263	0.299	0.290	0.335	0.573
	(0.054)	(0.043)	(0.011)	(0.004)	(0.003)
Completed Some School	n/a^	0.166	0.179	0.162	0.334
		(0.028)	(0.010)	(0.003)	(0.003)
Completed Std. 5	0.342	0.105	0.148	0.101	0.269
	(0.136)	(0.022)	(0.009)	(0.003)	(0.003)
Completed Post Primary	0.251	0.082	0.124	0.068	0.214
	(0.120)	(0.018)	(0.009)	(0.002)	(0.003)
Disabled	0.005	0.035	0.001	0.002	0.001
	(0.004)	(0.013)	(0.001)	(0.000)	(0.000)
Not Work	0.059	0.087	0.057	0.036	0.066
	(0.023)	(0.017)	(0.004)	(0.002)	(0.001)
Owns Small Business	0.053	0.095	0.108	0.059	0.109
	(0.021)	(0.042)	(0.006)	(0.002)	(0.002)
Works for Wages	0.425	0.762	0.571	0.095	0.225
	(0.065)	(0.057)	(0.009)	(0.003)	(0.003)
Employed in Agriculture	0.545	0.088	0.629	0.862	0.682
	(0.067)	(0.024)	(0.011)	(0.003)	(0.005)

Sample restricted to children age 10-14.
^All children report parent completing at least grade 1

Table 4: Maternal Characteristics in Porters Survey, Ragpickers Survey, and Census

			2001 Population and Housing Census		
	Short Route Porters Survey	Ragpickers Survey	Wage Work	Home Enterprise Work	Not Work
Mean/SE	(1)	(2)	(3)	(4)	(5)
Reports Characteristics	0.842	0.828	0.919	0.928	0.953
	(0.036)	(0.026)	(0.004)	(0.002)	(0.001)
Age	40.512	36.914	39.468	40.337	39.484
	(1.872)	(0.786)	(0.172)	(0.084)	(0.035)
Can Read and Write	0.065	0.127	0.151	0.075	0.232
	(0.025)	(0.036)	(0.010)	(0.003)	(0.003)
Completed Some School	n/a^	0.088	0.088	0.026	0.117
		(0.024)	(0.008)	(0.001)	(0.003)
Completed Std. 5	0.042	0.042	0.079	0.017	0.090
	(0.051)	(0.017)	(0.007)	(0.001)	(0.002)
Completed Post Primary	0.042	0.021	0.072	0.010	0.068
	(0.051)	(0.012)	(0.007)	(0.001)	(0.002)
Disabled	0.018	0.013	0.007	0.006	0.004
	(0.017)	(0.007)	(0.002)	(0.001)	(0.000)
Not Work	0.284	0.413	0.376	0.161	0.366
	(0.048)	(0.050)	(0.010)	(0.004)	(0.003)
Owns Small Business	0.007	0.032	0.098	0.098	0.102
	(0.004)	(0.013)	(0.006)	(0.003)	(0.001)
Works for Wages	0.301	0.496	0.350	0.027	0.053
	(0.057)	(0.074)	(0.009)	(0.002)	(0.001)
Employed in Agriculture	0.503	0.076	0.477	0.780	0.556
	(0.071)	(0.023)	(0.011)	(0.004)	(0.004)

Sample restricted to children age 10-14.

^All children report parent completing at least grade 1

Table 5: Attibutable Risk Estimates for Background Characteristics in Short Route Porters Survey

	Unconditional Attributable Risk			Conditional Attributable Risk		
		95% Confidence Interval			95% Confidence Interval	
	Estimate	Lower	Upper	Estimate	Lower	Upper
Household Background						
Owns Farmland	-0.00033	-0.00078	-0.00003	-0.00009	-0.00028	0.00005
Father Characteristics						
Reports Paternal Char.	-0.00009	-0.00036	0.00008			
Can Read and Write	-0.00024	-0.00039	-0.00012	-0.00012	-0.00023	-0.00004
Disabled	0.00118	0.00006	0.00503	0.00065	0.00000	0.00315
Not Working	0.00000	-0.00016	0.00025	-0.00010	-0.00018	-0.00003
Owns Small Business	-0.00012	-0.00024	0.00002	-0.00009	-0.00017	-0.00002
Works for Wages	0.00030	0.00010	0.00065	-0.00003	-0.00011	0.00006
Employed in Agriculture	-0.00018	-0.00043	-0.00003	-0.00017	-0.00038	-0.00004
Mother Characteristics						
Can Read and Write	-0.00020	-0.00031	-0.00009	-0.00010	-0.00016	-0.00004
Disabled	0.00143	-0.00009	0.00773	0.00054	-0.00007	0.00236
Not Working	-0.00006	-0.00016	0.00006	-0.00001	-0.00012	0.00010
Owns Small Business	-0.00025	-0.00039	-0.00015	-0.00014	-0.00025	-0.00007
Works for Wages	0.00127	0.00051	0.00257	0.00042	0.00011	0.00105
Employed in Agriculture	-0.00010	-0.00034	0.00005	-0.00005	-0.00024	0.00004

All regressions include controls for child age, gender, ethnicity, language, belt, and development region. All standard errors corrected for clustering at the block level (primary sampling unit). Estimates computed using King and Zeng's relogit code with prior correction: http://gking.harvard.edu/stats.shtml#relogit. Each estimate of attributable risk in the "unconditional" column is from a separate regression. Each estimate in the "conditional" column is from one regresion, including all of the listed covariates. All estimates assume an incidence of short route porters of 0.05 percent. Attributable risks are computed for a change in the row variable from 0 to 1 at the mean of all other covariates except all "conditional" estimates are computed at father and mother reports characteristics =1.

Table 6: Attibutable Risk Estimates for various scenarios in Short Route Porters Survey

	At Mean Landholding Rate			Landless		
	Attributable Risk	95% Confidence Interval		Attributable Risk	95% Confidence Interval	
	Estimate	Lower	Upper	Estimate	Lower	Upper
Disability						
Dad is disabled & cannot work (1)	0.0004	0.0000	0.0017	0.0006	-0.0001	0.0029
Mom is disabled & cannot work (2)	0.0008	0.0000	0.0043	0.0016	-0.0001	0.0102
Literacy						
Literate dad (avg sch) & illiterate mom (no sch.) to literate mom (3)	-0.0001	-0.0001	0.0000	-0.0001	-0.0003	0.0000
Illiterate mom & dad to literate dad (no schooling) (4)	-0.0001	-0.0003	-0.0001	-0.0002	-0.0005	-0.0001
Illiterate mom & dad to literate mom & dad (no schooling) (5)	-0.0002	-0.0003	-0.0001	-0.0003	-0.0007	-0.0002
Home Enterprises						
Household without any self employment to Mom self employment (6)	-0.0002	-0.0003	-0.0001	-0.0003	-0.0006	-0.0001
Household w/o self emp. to dad self emp. (7)	-0.0001	-0.0002	0.0000	-0.0002	-0.0004	0.0000
Household w/o self emp. to mom & dad self emp (8)	-0.0002	-0.0003	-0.0001	-0.0003	-0.0006	-0.0001
Wage Labor						
Household w/ no wage work to dad (9)	0.0000	-0.0001	0.0001	0.0000	-0.0002	0.0001
Household w/ no wage work to mom & dad (10)	0.0003	0.0001	0.0010	0.0005	0.0001	0.0015

Attributable risks computed using results from the "conditional regression" results in table 5. The first columns compute probabilities for households with mean probability of holding land. The second column computes probabilities for household without landholdings.

(1) - Change in probability that child is a porter if father moves from not disabled and mean work to disabled and no work (any catagory).

(2) - same as (1) for mother

(3) - change in probability that child is a porter if dad is literate with average schooling and mom moves drom illiterate to literate (with no schooling)

(4) - change in probability that child is a porter if illiterate mom and dad shifts to a illiterate mom with literate dad (no schooling)

(5) - change in probability that a child is a porter if illiterate mom and dad shifts to literate mom and dad (no schooling)

(6) - change in probability that child is porter if household moves from no self employment to mom self employment

(7) - same as (6) only for father

(8) - change in probability that child is a porter if household moves from no self-employment to both mom and dad in self-employment

(9) - change in probability that child is a porter if household moves from no wage work to father wage work

(10) - same as (9) except mom & dad in wage work

Table 7: Attibutable Risk Estimates for various scenarios, census wage workers

	Average Land Holdings			Landless		
	Attributable Risk	95% Confidence Interval		Attributable Risk	95% Confidence Interval	
	Estimate	Lower	Upper	Estimate	Lower	Upper
Disability						
Dad is disabled & cannot work (1)	0.002	-0.004	0.012	0.003	-0.007	0.021
Mom is disabled & cannot work (2)	0.013	0.004	0.024	0.021	0.008	0.041
Literacy						
Literate dad (avg sch) & illiterate mom (no sch.) to literate mom (3)	-0.001	-0.002	0.000	-0.002	-0.003	0.000
Illiterate mom & dad to literate dad (no schooling) (4)	-0.006	-0.007	-0.005	-0.010	-0.012	-0.008
Illiterate mom & dad to literate mom & dad (no schooling) (5)	-0.008	-0.009	-0.006	-0.013	-0.015	-0.010
Home Enterprises						
Household without any self employment to Mom self employment (6)	0.001	0.000	0.002	0.002	0.000	0.004
Household w/o self emp. to dad self emp. (7)	0.007	0.005	0.009	0.011	0.008	0.015
Household w/o self emp. to mom & dad self emp (8)	0.009	0.006	0.012	0.014	0.010	0.020

The first columns compute probabilities for households with mean landholdings. The second column computes probabilities for household without landholdings.

(1) - Change in probability that child is a wage worker if father moves from not disabled and mean work to disabled and no work.

(2) - same as (1) for mother

(3) - change in probability that child is a wage worker if dad is literate with average schooling and mom moves drom illiterate to literate (with no schooling)

(4) - change in probability that child is a wage worker if illiterate mom and dad shifts to a illiterate mom with literate dad (no schooling)

(5) - change in probability that a child is a wage worker if illiterate mom and dad shifts to literate mom and dad (no schooling)

(6) - change in probability that child is wage worker if household moves from no self employment to mom self employment

(7) - same as (6) only for father

(8) - change in probability that child is a wage worker if household moves from no self-employment to both mom and dad in self-employment

Table 8: Attibutable Risk Estimates for Background Characteristics in Ragpickers Survey

	Unconditional Attributable Risk	95% Confidence Interval		Conditional Attributable Risk	95% Confidence Interval	
	Estimate	Lower	Upper	Estimate	Lower	Upper
Household Background						
Owns Farmland	-0.00017	-0.00031	-0.00008	0.00000	-0.00001	0.00000
Father Characteristics						
Reports Characteristics	-0.00002	-0.00006	0.00000			
Can Read and Write	-0.00004	-0.00008	-0.00002	-0.00001	-0.00002	0.00000
Disabled	0.00154	0.00052	0.00376	0.00024	0.00006	0.00064
Not Working	0.00002	0.00000	0.00006	0.00000	-0.00002	0.00000
Owns Small Business	0.00000	-0.00004	0.00009	0.00000	-0.00001	0.00001
Employed in Agriculture	-0.00016	-0.00030	-0.00007	-0.00003	-0.00009	-0.00001
Mother Characteristics						
Reports Characteristics	-0.00012	-0.00023	-0.00005			
Can Read and Write	-0.00001	-0.00003	0.00002	0.00000	-0.00001	0.00000
Disabled	0.00016	0.00001	0.00057	0.00002	0.00000	0.00015
Not Working	0.00001	-0.00003	0.00008	0.00000	0.00000	0.00001
Owns Small Business	-0.00004	-0.00007	-0.00001	-0.00001	-0.00002	0.00000
Employed in Agriculture	-0.00011	-0.00021	-0.00005	-0.00002	-0.00005	0.00000

All regressions include controls for child age, gender, ethnicity, language, belt, and development region. All standard errors corrected for clustering at the block level (primary sampling unit). Estimates computed using King and Zeng's relogit code with prior correction: http://gking.harvard.edu/stats.shtml#relogit. Each estimate of attributable risk in the "unconditional" column is from a separate regression. Each estimate in the "conditional" column is from one regresion, including all of the listed covariates. All estimates assume an incidence rate of ragpicking of 0.03 percent. Attributable risks are computed for a change in the row variable from 0 to 1 at the mean of all other covariates except all "conditional" estimates are computed at father and mother reports characteristics =1.

Table 9: Attibutable Risk Estimates for various scenarios in Ragpickers Survey

	At Mean Landholding Rate			Landless		
	Attributable Risk	95% Confidence Interval		Attributable Risk	95% Confidence Interval	
	Estimate	Lower	Upper	Estimate	Lower	Upper
Disability						
Dad is disabled & cannot work (1)	0.00046	0.00011	0.00121	0.00075	0.00017	0.00208
Mom is disabled & cannot work (2)	0.00009	0.00001	0.00061	0.00014	0.00001	0.00064
Literacy						
Literate dad (avg sch) & illiterate mom (no sch.) to literate mom (3)	0.00000	-0.00001	0.00000	0.00000	-0.00001	0.00000
Illiterate mom & dad to literate dad (no schooling) (4)	-0.00001	-0.00003	0.00000	-0.00001	-0.00004	0.00000
Illiterate mom & dad to literate mom & dad (no schooling) (5)	-0.00001	-0.00004	0.00000	-0.00001	-0.00005	0.00000
Home Enterprises						
Household without any self employment to Mom self employment (6)	0.00000	-0.00002	0.00000	-0.00001	-0.00002	0.00000
Household w/o self emp. to dad self emp. (7)	0.00000	-0.00001	0.00001	0.00000	-0.00001	0.00001
Household w/o self emp. to mom & dad self emp (8)	0.00000	-0.00002	0.00000	-0.00001	-0.00002	0.00000
Wage Labor						
Household w/ no wage work to dad (9)	0.00000	0.00000	0.00002	0.00001	0.00000	0.00002
Household w/ no wage work to mom & dad (10)	0.00005	0.00001	0.00017	0.00008	0.00002	0.00024

Attributable risks computed using results from the "conditional regression" results in table 8. The first columns compute probabilities for households with mean probability of holding land. The second column computes probabilities for household without landholdings.

(1) - Change in probability that child is a ragpicker if father moves from not disabled and mean work to disabled and no work (any catagory).

(2) - same as (1) for mother

(3) - change in probability that child is a ragpicker if dad is literate with average schooling and mom moves drom illiterate to literate (with no schooling)

(4) - change in probability that child is a ragpicker if illiterate mom and dad shifts to a illiterate mom with literate dad (no schooling)

(5) - change in probability that a child is a ragpicker if illiterate mom and dad shifts to literate mom and dad (no schooling)

(6) - change in probability that child is ragpicker if household moves from no self employment to mom self employment

(7) - same as (6) only for father

(8) - change in probability that child is a ragpicker if household moves from no self-employment to both mom and dad in self-employment

(9) - change in probability that child is a ragpicker if household moves from no wage work to father wage work

(10) - same as (9) except mom & dad in wage work

Table 10: Bounds on Attibutable Risk Estimates for various scenarios in Ragpickers Survey, landless households

Incidence rates bounded between 0.3 and 0.03 percent

	Estimated Bounds		95 % Confidence Intervals for Bounds	
	Upper	Lower	Lower Value	Upper Value
Disability				
Dad is disabled & cannot work (1)	0.00742	0.00075	0.00017	0.01995
Mom is disabled & cannot work (2)	0.00136	0.00014	0.00001	0.00621
Literacy				
Literate dad (avg sch) & illiterate mom (no sch.) to literate mom (3)	0.00000	-0.00003	-0.00009	0.00000
Illiterate mom & dad to literate dad (no schooling) (4)	-0.00001	-0.00010	-0.00039	0.00000
Illiterate mom & dad to literate mom & dad (no schooling) (5)	-0.00001	-0.00013	-0.00047	0.00000
Home Enterprises				
Household without any self employment to Mom self employment (6)	-0.00001	-0.00007	-0.00022	0.00000
Household w/o self emp. to dad self emp. (7)	0.00000	-0.00004	-0.00011	0.00001
Household w/o self emp. to mom & dad self emp (8)	-0.00001	-0.00008	-0.00020	0.00000
Wage Labor				
Household w/ no wage work to dad (9)	0.00007	0.00001	0.00000	0.00022
Household w/ no wage work to mom & dad (10)	0.00080	0.00008	0.00002	0.00238

Attributable risks computed using results from the "conditional regression" results in table 8 assuming an incidence of 0.03 percent and unreported regressions assuming an incidence of 0.3 percent.

(1) - Change in probability that child is a ragpicker if father moves from not disabled and mean work to disabled and no work (any catagory).

(2) - same as (1) for mother

(3) - change in probability that child is a ragpicker if dad is literate with average schooling and mom moves drom illiterate to literate (with no schooling)

(4) - change in probability that child is a ragpicker if illiterate mom and dad shifts to a illiterate mom with literate dad (no schooling)

(5) - change in probability that a child is a ragpicker if illiterate mom and dad shifts to literate mom and dad (no schooling)

(6) - change in probability that child is ragpicker if household moves from no self employment to mom self employment

(7) - same as (6) only for father

(8) - change in probability that child is a ragpicker if household moves from no self-employment to both mom and dad in self-employment

(9) - change in probability that child is a ragpicker if household moves from no wage work to father wage work

(10) - same as (9) except mom & dad in wage work