**Methodological Problems in the
Study of Child Soldiers**

Barry Ames
University of Pittsburgh
2007

About the Ford Institute for Human Security

The mission of the Ford Institute for Human Security is to conduct research that focuses on a series of transnational threats to the human rights of civilian populations. The Institute's purpose is to generate independent research, disseminate policy papers, and advocate nonpartisan policy proposals available to both domestic and international policymakers.

In today's increasingly globalized world, there has been a shift from traditional threats focused on the territorial integrity of nations to global threats that center on the safety of individuals. The Ford Institute for Human Security at the University of Pittsburgh recognizes the critical importance of these emergent problems in human security.

This working paper is a product of the Ford Institute's working group, "Child Soldiers Initative: Building Knowledge about Children and Armed Conflict". The Child Soldiers Initiative is an ongoing network of scholars, policymakers and representatives of civil society engaged in promoting and developing policy proposals addressing the recruitment and reintegration of child soldiers.

Methodological Problems in the Study of Child Soldiers

Barry Ames

A powerful moral dimension characterizes research on the problem of child soldiers. Given the importance of the moral issues, how critical is methodology and scientific neutrality? Child soldiers may be the most depressing topic in the field of international conflict, and the elimination of child soldiering is a completely laudable goal. But the methodological principles promoting good scholarship are still crucial. A more precise understanding of why children join violent conflicts, why irregular armies recruit and abduct children, why children leave irregular armies, and what problems they face after leaving will help us craft policies to reduce child soldiering.

Much of the literature on child soldiers comes from advocacy groups. Not only would it be unrealistic to expect such advocates to jump into complex statistical analyses; it would also be counterproductive. Advocacy groups are valuable precisely because they awaken readers to the horrors of child soldiering. The research of advocacy groups, nevertheless, can be informed and guided by the theoretical models and rigorous statistical techniques of the scientific community.

This essay adopts a somewhat "extremist" methodological perspective. Scientific inference, in this view, depends not only on what one knows but also on recognizing what one does not know, that is, on recognizing the gap between the data you have and the data you want. This does not mean that the real limitations of data availability should be ignored; rather, we

should be explicit about the consequences of the inevitable compromises imposed by limitations of time and money.

Research on child soldiers falls into four, somewhat arbitrarily defined categories. Perhaps the best-known type is work based on compiling reports on ex-combatants in multiple countries. A good example is Ilene Cohn and Guy Goodwin-Gill's monograph *Child Soldiers: The Role of Children in Armed Conflict* (1994). Cohn and Goodwin-Gill use data from Sri Lanka, Guatemala, El Salvador, Liberia, and the Israeli-occupied territories. A second type, qualitative single-country (quasi-ethnographic) studies, includes Sebastian Brett's *You'll Learn Not to Cry: Child Combatants in Colombia,* published in 2003 for Human Rights Watch.[1] Quantitative studies focusing purely on child soldiers in individual countries seem nonexistent, but quantitative studies of ex-combatants in general (both children and adults) include Humphreys and Weinstein's important papers on Sierra Leone (2006a and 2006b).[2] Multiple-country quantitative projects are of two subtypes. *Child Soldiers Global Report, 2004*, a worldwide compilation of facts, laws and protocols sponsored by the Coalition to Stop the Use of Child Soldiers (SUCH), is essentially a data fact book. The only quantitative large-N (multicountry) project of which I am aware is Achvarina and Reich's "Why Do Children 'Fight'? Explaining the Phenomenon of Child Soldiers in African Intra-State Conflicts" (2006).

How we do research usually depends on the substantive questions we ask. Here I focus on the most commonly asked question: Why do children (typically defined as persons younger than 18 years of age) participate in armed conflicts? Sometimes guerrilla and paramilitary groups abduct children; sometimes children join voluntarily; sometimes they are sold into combat groups by relatives. Note the vagueness of these categories; there may be little distinction in practice between coercion and seemingly voluntary enlistment. And the data used

to measure children's participation may take the form of an aggregate rate (how many children in a given population participate) or an individual decision to join or not to join.

A cursory reading of the qualitative literature on participation produces a common set of explanations. "Grievance" factors include poverty, loss of parents, lack of economic opportunity, abuse at home, ethnicity, political belief, and so on (Brett 2003). "Inducement" factors include pay, glory, and promises of future material gains. "Solidarity" factors include group cohesion, village networks, and friends (Petersen 1993). "Accessibility" factors include the presence and vulnerability of refugee camps (Achvarina & Reich 2006).

These factors may be thought of as competing, but all might be true -- indeed, all probably are true -- at least under specified and varying circumstances. Our task is to assess the causes of children's participation and to determine the relative importance of various causal factors. Because ultimately only statistical modeling can make such assessments, I focus on large-N comparison.[3] My suggestions, specific as well to the particular problem of child soldiers in civil wars, come in only a small measure from the single large-N study so far undertaken (Achvarina & Reich 2006). To a larger degree they come from the kinds of explanations, explicit and implicit, found in the ethnographic literature surveying ex-combatants. Though I focus on large-N quantitative comparisons, the suggestions developed here also apply to the qualitative, quasi-ethnographic undertakings of advocacy groups in single or multiple-country settings.

What is the Unit of Analysis?

In political science, most large-N statistical comparisons of aggregate data are based on nation-states. Sometimes nations are grouped into dyads, as in the study of interstate conflict;

sometimes individual countries are maintained as units, as in the study of conditions favoring democratization. In either case, we take for granted that the nation-state is the unit of analysis.[4]

In studies of child soldiers, I suspect that taking the nation as the unit of analysis is almost always a mistake. If the object of explanation (the dependent variable) is an attribute or behavior of a fixed political or geographical entity (as in studies of levels of democratization), then that entity, typically a nation, is the appropriate unit of analysis. And if all we care about is predicting some outcome, such as an election winner, a nationwide sample ignoring spatial variation may work quite well. But in studies of child soldiers we seek to understand whether or how many children become part of irregular forces, and we hope we can intervene in some of the variables identified as central to recruitment.[5] The nation-state is an inappropriate unit if such predictive factors vary across the national space. Of course in analyses of child soldiers the predictive factors nearly always do vary spatially. Conflicts are localized within particular communities, ethnic groups, or regions. Even in cases without ethnic conflict, such as Colombia, rural areas are much more likely than urban areas to be involved.

Suppose, as one indicator of "grievances," we want to assess the effects of poverty on the willingness of children to join irregular forces. Poverty rates are likely to be quite a bit higher in areas contested by guerrillas and government forces, because conflict destroys assets and disrupts agricultural activities. If we take the national poverty rate as our indicator of poverty, its observed effects will inevitably appear weaker than its true effects. Similarly, we expect that children are more likely to become child soldiers if one or both parents are dead. But the percentage of children missing one or both parents is much higher in disputed areas (areas contested by rebels, paramilitaries, and/or government forces) than in the nation as a whole.

Whether irregular forces are abducting kids or they join more or less voluntarily, local context matters.   Children join because their friends join (Brett 2003); guerrillas abduct because they have already penetrated a village; the presence of other parentless children affects individual parentless children. Contextual effects need to be assessed at the appropriate level. Sometimes the appropriate level is the faction, as in Pugel's discussion of four factional groups in Liberia (this volume), and sometimes it is the village or group of closely related villages. Researchers must, in addition, take particular care with post-hoc explanations or rationalizations provided by ex-combatants who have left irregular forces.

People in Context

In the early years of public opinion surveys, two perspectives, the Michigan school and the Columbia school, competed. The Michigan model, represented by the Institute for Social Research and the Interuniversity Consortium for Political and Social Research (ICPSR) at the University of Michigan, focused on national surveys. Because such surveys are nationally representative, they are very good at predicting the results of national elections. The Columbia (University) model, represented by the work of Lazarsfeld, et al. (1968), was more sociological. Its surveys frequently limited themselves to communities and neighborhoods. Such surveys could not be generalized to the national level, but they were better at illuminating the processes leading people to change their opinions and behaviors. In the end, the Michigan model dominated, in part because its surveys were better able to predict broad electoral outcomes and in part because a model assuming an atomized citizenry fit 20th century American society.

The places where children fight are worlds away from late industrial society. No television, no air conditioning, no bedroom suburbs; these are places where face-to-face contact

is the primary mode of persuasion, where ideas move through dense social networks.[6] And it is exactly here that surveys of smaller ecological units, surveys that include respondents' neighborhood and workplace contracts, may be better able to reveal the deeper motivations of adults and children as they confront irregular forces.

Selection Bias

Surveys of children as ex-combatants typically (perhaps always) suffer from a simple form of selection bias, because they fail to interview non-combatants.[7]   Some children join irregular forces; most children do not.   Guerrillas and paramilitaries offer selective incentives to children, only some accept. What is it about these children that makes them more likely to respond to incentives? Only by comparing combatants to non-combatants can we assess these causal factors.[8]

Another form of selection bias comes from focusing solely on children. Recruitment of children may be no different from recruitment of adults. Moreover, the way irregular forces treat civilians affects the recruitment of children.

If the relevant population is not merely children and not merely ex-combatants, a given population then consists of non-combatants, volunteers, and victims of abduction.[9]   Both volunteers and victims of abduction may be further subdivided into the various factions and subfactions with which they may be involved. At the level of the individual, through survey responses, the appropriate methodological technique is then some sort of multinomial logit, that is, a statistical technique allowing the simultaneous comparison of distinct, non-ordinally related categories.[10]

Finally, large-N studies need to include cases in which intrastate conflict did not result in the utilization of child soldiers. In terms of the factors that induce recruitment of children, whether by guerrilla, paramilitary, or governmental forces, exclusion of no-recruitment cases leads to an under appreciation of the forces leading different kinds of organizations to decide to include or exclude children (See Gutierrez-Sanin, this volume).

The Centrality of Time

Cross-sectional surveys cannot provide a basis for inference unless we assume that the variables included in the model are stable over time. In most political science research this assumption is questionable, but the temporal variation is often random and simply produces noise.[11] In research on child soldiers, however, time variation is an inherent and systematic problem.

Suppose each conflict contributes one data point, one observation, in a large-N study, where the observations are regions, villages, or other relevant units of analysis. This single data point, however, may represent very different stages in a civil war. Brett notes that the use of children by paramilitaries and guerrillas in Colombia only began in the 1990s, while fighting – and the emergence of paramilitaries – began long before (2003, 20). Recruiting changed, including the initiation of forcible abductions, because conditions in the countryside changed. Poverty for children increased, and the guerrillas needed manpower from both boys and girls. In Moçambique, Renamo began recruiting children after 1979, when Rhodesia ceased to be its main source of funds and support (Weinstein 2005, 612). And in Sierra Leone, the RUF rebel movement, once dominated by university students, began recruiting unemployed youths and urban marginals after the collapse of its student base. Not only did the nature of the recruits

change, but the kinds of incentives provided to them changed. Looting, for example, became more prominent.

Collecting data from ex-combatants is obviously very difficult, and it would be foolish to mandate the use of panel surveys. But coding schemes can be "time-aware." At the individual level, this means including variables measuring the stages at which respondents joined and left conflicts. For aggregate data (where the object of explanation might be the percentage of some community's child soldiers who joined or left irregular forces), researchers might be able to project or impute data at parallel stages of a conflict. Humphreys and Weinstein (2006a, 15), in their surveys of ex-combatants in Sierra Leone, asked respondents to give their location and faction membership for seven designated time periods "marked by major events in the history of the conflict." Interviewers then asked about one of these periods of activity, chosen at random.[12] Other, perhaps simpler, schemes to build time-awareness into surveys are possible; what is not acceptable is to ignore the problem.

Model Interactions

Why do children join irregular forces? In terms of grievances, broadly defined, most ethnographic studies stress the same factors. In Colombia, for example, Brett (2003) notes that most recruits came from poor homes, had dropped out of school, and had suffered abuse at home. But in the areas where irregular forces recruit, almost everyone is poor, and most children refuse to enlist. Does this mean that poverty really has no impact? In a linear additive model, poverty might drop out (as it does in Achvarina and Reich). But our stories about poverty, lack of opportunity and abuse are really stories about *interactions*, that is, about conditions working jointly. We hypothesize that abuse, for really poor people, may trigger the kind of hopelessness

that leads to abandoning the community and joining irregular forces. For middle-income people, abuse may not trigger such enlistment, because more social support is available from relatives and friends. Similarly, poverty may have different consequences for different ethnic groups, because some groups engage in more communal sharing of resources, thus cushioning individual-level poverty.[13]

Scholars should consider every independent variable in conjunction with every other independent variable, asking whether the effects of one variable might depend on the level of another. If so, interactive tests are appropriate. Modern statistical software makes this simple; all the major statistical packages (Stata, SAS, etc.) allow easy calculation and graphing of interaction terms (Gary King's 'Clarify' program is very user friendly).

Impute Missing Data

In large-N analyses of child soldiering, missing data are almost a given. Earlier generations of scholars were trained simply to delete cases 'listwise,' that is, to drop entire cases if a measurement on any variable was missing. Such deletion is no longer acceptable, because it is now understood that the biases introduced by deletion can seriously distort inference (King et al. 2001). Because data are rarely missing purely randomly, deletion biases results.

Listwise deletion is not always the wrong strategy. When the cases dropped are only a small portion of the dataset, and when they are representative of the whole dataset, i.e., a truly random sample, then deletion can work. But "representative" means that the chance an observation is missing does not depend on any data values, whether observed or missing.

Researchers should first ask why data are missing. Suppose we are modeling weight (Y) as a function of gender (X).[14] If some respondents refuse to give their weight, Y will have missing values.

1. There may be no particular reason why some respondents give their weights and others do not. Thus the probability that Y is missing may have no relationship to X or Y. Such data are "Missing Completely at Random (MCAR)."

2. One gender may be less likely to report its weight. Thus the probability that Y is missing depends only on the value of X. Such data are "Missing at Random (MAR)."

3. Heavy (or light) people may be less likely to disclose their weight. In this case, the probability that Y is missing depends on the unobserved value of Y itself. Such data are not missing at random (NMAR). This is the problem of "Non-Ignorable Missing Data."

In the study of child soldiering, missing data are especially unlikely to be missing completely at random (MCAR). Conflicts attracting less international attention are likely to be those with more missing data. International advocacy groups with limited resources naturally concentrate on the biggest conflicts. Smaller conflicts, conflicts farther back in time, conflicts initiated closer to the data collection point, conflicts involving political rather than ethnic grievances; all these will attract less attention.[15] But in all these cases, if we are trying to explain child soldier recruitment rates, the missing data can be modeled as a function of various independent variables (such as political rather than ethnic grievances). Hence these are all examples of "Missing at Random" (MAR) data.

Recall our earlier discussion of units of analysis or observation. It is easy to see that the unit of analysis problem impacts data availability. Though imprecise, estimates for variables predicting grievances or material inducements may exist at the national level. If such indicators

have to be specified for multiple, smaller, units of observation, the likelihood of variables with missing observations is quite high.

Awareness of the problem is a first step toward its amelioration. If NGOs working on child soldiers understand that deleting cases because of missing data hinders inference, they will pay more attention to gathering complete data at appropriate observational levels in the first instance. It is always more difficult to go back and fill in data gaps than to structure a data-gathering design correctly in the first place.

What about scholars working in large-N context with previously gathered data? If the data are MCAR or MAR or, the best solution is some kind of imputation. Analyzing data one does not have may seem bizarre, but imputation has become standard practice in statistics (Little & Rubin 1987; Allison 2001).

All forms of imputation create values for missing data by modeling the relationships between the variables of interest (including those with no missing data at all). But we cannot simply replace a missing value with a value generated by an equation involving the other variables in the model, because we have less confidence in the imputed values than in the non-missing values. In other words, inserting single values for missing data would result in estimates of variances that are too small. Multiple imputation techniques address this by modeling our uncertainty, both in the parameters generating missing values (since these are generated in part by the missing values) and in the values themselves.[16] With modern software packages, this is much less difficult than it sounds.[17]

What if the data are characterized by non-ignorable missingness? In studies of child soldiers at the individual level, non-ignorable missingness might result when we model the post-combat propensity of child soldiers to commit violent acts, because the most violent ex-

combatants might be reluctant to admit to violent acts. At the aggregate level non-ignorable missingness would occur if cases with the highest rates of child recruitment are reluctant to allow researchers to enter their territories.

One approach to non-ignorable missingness is to impute values based on data otherwise external to the research design. Race, for instance, can be estimated with Census block data associated with the address of the respondent. Another approach is the use of expert information, that is, the judgments of country and regional experts. This approach should be particularly useful in time series cross-sections, that is, analyses of a large number of cases over time. In these TSCS analyses, we might be missing an estimate of the number of child soldiers for a case in a single year, but we have data for the years before and after. If experts on the conflict know that a spike up or down was likely or unlikely, we can use that information to constrain (and thus improve) our imputation model. Likewise, if the poverty level is missing for a year characterized by a famine, we can use that information to help estimate the missing poverty rate. Honaker and King (2006) have recently discussed, with some interesting African examples, new techniques for analyzing situations of non-ignorability.

. *Judging Explanations*

In the words of King, Keohane and Verba, a social science theory is a "reasoned and precise speculation about the answer to a research question, including a statement about why the proposed answer is correct" (1994, 19). Theories generate *testable hypotheses*. From Achvarina and Reich's argument about refugee camps (this volume), one can derive the following theoretical statement: If refugee camps are unprotected, guerrilla groups will use them as a base for the recruitment of child soldiers. This hypothesis is inductively derived in the work of

Achvarina and Reich, but it now takes on the status of a research finding, and can thus be evaluated in new settings as part of a broader explanation of child recruitment. Their hypothesis has some fundamental characteristics: it may be falsified, it has observable implications, it is concrete.

Some theories begin with findings from prior work on the same broad problem but then emphasize new explanatory variables not previously examined. Jo Becker's paper on child soldier recruitment in Sri Lanka, Nepal and Burma (this volume) stresses factors unique to these nations, including political indoctrination and past government abuses. The next logical research step is the evaluation of Becker's factors in the setting of the original research, i.e., Africa.

In this volume, three papers stand out as self-consciously theory-driven, i.e., these three papers present comprehensive causal arguments derived from prior research. Jens Christopher Andvig and Scott Gates "Recruiting Children for Armed Conflict" emphasizes the economics of child labor as well as the psychology of children. For example, research on child behavior finds that children underestimate low probability risks when associated with losses, even if the risk of loss is fairly high. Children underestimate risks more than adolescents. Of course the world of child soldiers is unlike a Western laboratory setting, but this result provides some support to the idea that it may be easier for guerrilla organizations to employ children than adults, given the inherently high risk-reward ratios of these conflicts.

Francisco Gutiérrez Sanín's "Organizing Minors" stresses organizational factors. Within a given country, why do some armed forces recruit children while others do not? Noting that rational choice arguments are inherently limited – no rational calculation of costs vs. benefits would lead anyone to volunteer for a guerrilla force – Gutiérrez Sanín points to the differences between rent-seeking bands (as in Sierra Leone) and armies with a consistent political ideology

and minimal rent-seeking, such as the FARC in Colombia. The FARC was too big and too serious to rely on selective incentives, hence its leaders were forced to develop socialization mechanisms promoting gregariousness and a sense of the collective in its recruits. The organization learned this through trial and error, i.e., after several military disasters. Gutiérrez-Sanín's analysis suggests that political armies, those whose objectives transcend rent seeking, either begin with mechanisms for socializing young recruits, develop such mechanisms in response to military debacles, or eventually cease to recruit young soldiers altogether.

James Pugel's "Disaggregating the Causal Factors Unique to Child Soldiering" demonstrates that eliminating one form of selection bias – by comparing child combatants to adult combatants – helps refine arguments about recruitment, because it turns out that certain factors thought to be unique to children are also present for adults. When Pugel disaggregated the warring groups in Liberia by faction and age of soldier, it turns out that the reasons cited by children for their participation in armed groups – money, family protection, and abduction – were the same as those cited by adults.

Conclusion

This essay began by defending a "principled" methodological perspective, laying out the data requirements and analytical techniques necessary for rigorous scientific testing of theories predicting whether and under what conditions children will participate in irregular forces. Questions of data requirements and statistical methodologies are interwoven, because the statistical techniques suggested here usually require more and better quality data. Given the limited budgets of child soldier researchers, it would be naive to think that this "principled" perspective will be adopted wholesale. Where should researchers compromise?

Individual research groups (mainly sponsored by NGOs) have skills and resources linked to particular conflicts. Instead of seeking to mount multi-conflict research designs (especially designs in which a handful of ex-child soldiers are interviewed from each conflict), it should be cheaper and easier to focus on carefully selected units of observation within particular conflicts. This means, of course, a loss in generality, but the gain in the accuracy of the theoretical argument should easily compensate. At the same time, the various NGOs interested in child soldiering could coordinate their research (through conferences like the PRIO-Ford Institute workshop), so that their research questions are the same and their methodologies mutually consistent.

If we accept the importance of the social contexts in which potential child combatants live, we ought to be able to enrich our analyses of these contexts. One possibility is to interview enough people at the level of the community or neighborhood so that the aggregate opinion – political and social – of the social unit can be measured. Another possibility is to develop "snowball" samples, samples in which respondents provide information identifying their immediate social contacts. These contacts are in turn interviewed, and the resulting network constitutes the real social group encouraging or discouraging children from abandoning their communities to join irregular forces. Because interviewers are likely to be on the ground in the community already, such techniques can deepen our understanding of the forces impinging upon children without dramatically increasing the cost of research.

Notes

[1] Notable among single-country studies of ex-combatants in general in Colombia is Arjona (2006).

[2] The group known as SWAY (Survey of War Affected Youth) is undertaking surveys in displacement camps in Northern Uganda, but their focus is on boys and girls abducted by the Lord's Resistance Army and living in these camps. See http://www.sway-uganda.org/

[3] This does not mean that any large-N statistical model is better than any small-N study; indeed, that is certainly untrue. But ultimately only statistical testing has the ability to disconfirm hypotheses by "holding constant" other predictors. In effect, statistical modeling is analogous to the "most-similar systems" design of Przeworski and Teune (1970). See also King et al. (1994).

[4] This is the case for Achvarina and Reich (2006).

[5] Achvarina and Reich (2006) note that the causal status of refugee camps is particularly important, because policy interventions can reduce their vulnerability to incursions by irregular forces.

[6] Roger Petersen's (1993) work on Lithuania in the1930s is particularly useful here. Another example is David Laitin's analysis of rebellion in Catalonia and the Basque country (Laitin 1995).

[7] Humphreys and Weinstein (2006a) do not focus on children, but their survey includes ex-combatants and non-combatants.

[8] A good example of such comparisons is the work of SWAY (Survey of War-Affected Youth) in Uganda. See SWAY-Uganda (2006).

[9] One significant group is perforce left out, i.e., the children who do not return from irregular armies. If the children killed in the bush are just a random sample of all combatant children, their exclusion does not matter. But if they are smaller, sicker, or malnourished, or if their attitudes affected their fighting behavior or their relationships with older combatants, then a problem of selection bias results.

[10] Logit (or logistic) analysis is appropriate when the dependent variable is a dichotomous choice, such as voting Republican or Democratic, and can only take the values '0' or '1'. If the dependent variable has multiple categories (stayed home, volunteered for rebels, volunteered for paramilitaries, abducted), then the multinomial logit is used. One of the categories becomes a reference or baseline, and the other categories are measured with respect to that baseline outcome.

[11] If error is random, then the coefficient will be accurate but the standard errors will be too large. We are thus likely to reject a hypothesis that should not be rejected. In most social science

research this is a lesser evil than accepting a false hypothesis. But in the case of child soldiers, where we hope to intervene in certain variables to change outcomes, the rejection of a true predictor is more serious.

[12] Cf. Humphreys and Weinstein's description of the conflict in Sierra Leone. "The war ... was long and complex. It lasted for over a decade and involved five primary factions, numerous sub-factions and various external actors. Over the course of the conflict, the government changed hands four times and two peace accords were negotiated and failed. Individual experiences of the fighting were also complex. Some ex-combatants were involved in the conflict for short periods of time, while others entered early in the conflict and stayed to the end. Some changed sub-faction or primary faction during the conflict and almost all moved locations" (2006a, 14).

[13]

[14] This example comes from http://www.sociology.osu.edu/people/ptv/faq/missing/missing.htm.

[15] An interesting case in point is the Naxalite rebellion in India. Though this conflict has been ongoing for many years, and though it is included in data fact books, the international press had essentially forgotten about it until NY Times articles in the spring of 2006.

[16] MI programs produce multiple data sets, each with different imputed values. After a multiple regression (or any other statistical technique) is implemented on each data set, the coefficients are averaged and the standard errors calculated by formulas known as Rubin's Rules.

[17] Stata users can download a particularly good MI utility called 'ICE.' Gary King has developed 'Amelia,' and there are many other packages as well.