

A Review of the Gaps in the Academic Literature on Child Labor, Forced Labor, and Human Trafficking

Eric V. Edmonds, Dartmouth College

December 17, 2019

Executive Summary

A query of child labor and forced labor in the major research bibliographic databases yielded 614 original research papers published since 2010. Relative to prevalence, the literature overemphasizes research on forced labor. Impact evaluations, especially randomized control trials are rare. Hence, the literature is far from providing a complete toolkit of what works for child labor and forced labor. It is especially unfortunate in the forced labor space where a large body of research documents circumstances of forced labor with very little research guiding what to do to reduce forced labor. Even in the child labor space where there are more impact evaluations, there is a mismatch between the target of those evaluations and how child labor projects actually operate. For child labor research, there appears to be no relationship between the number of studies and the prevalence rate of child labor.

1 Motivation

The Sustainable Development Goals include targets of eliminating child labor by 2025 and forced labor by 2030. Policy progress towards meeting these goals depends on reliable evidence on the extent of child labor, forced labor, and human trafficking as well as causal research on what can influence the prevalence of these circumstances.

The goal of this document is to review the available evidence that can be used as the basis for policy choices to eliminate child labor, forced labor, and human trafficking. A total of 27,587 research papers were evaluated for inclusion. 614 of these research papers were principally based on original data collection and analysis with direct policy relevance and hence for the basis for this review. A companion document contains a complete database of all 614 studies used in this review.

2 Methodology

2.1 Definitions

This study does not impose definitions on the terms child labor, forced labor, and human trafficking in its review of the literature, allowing researchers to self-select into what label is appropriate for their work. However, it is useful to fix ideas in the discussion. As such, this overview defines these three concepts as follows.

2.1.1 Child Labor

Child labor is often defined as work that deprives children of their childhood, their potential and their dignity, and that is harmful to physical and mental development. It refers to work that:

- is mentally, physically, socially or morally dangerous and harmful to children; and
- interferes with their schooling by:
 - depriving them of the opportunity to attend school;
 - obliging them to leave school prematurely; or
 - requiring them to attempt to combine school attendance with excessively long and heavy work

Not all work is child labor. Where possible, we have tried to respect local laws in applying the phrase only to work that is illegal in the country context.

2.1.2 Forced Labor

Forced labor is work for which a person has not offered him or herself voluntarily (criterion of “involuntariness”) and which is performed under coercion (criterion of “menace of penalty”) applied by an employer or a third party. The coercion may take place during the worker’s recruitment process to force him or her to accept the job or, once the person is working, to force him or her to do tasks that were not part of what was agreed to at the time of recruitment or to prevent him or her from leaving the job.

2.1.3 Human Trafficking

In the labor sphere, human trafficking differs from forced labor and child labor in that it describes a process through which someone enters into child labor or forced labor. Trafficking is the process of coercing people into situations in which they cannot exit. Some human trafficking is outside of the labor sphere, including trafficking for organ removal, forced military service, forced marriage or adoption.

2.2 Databases Examined

Google Scholar, ProQuest, ERIC, PubMed, and Web of Science are the 5 main citation ecosystems in academic research, and they were the focus of the review conducted herein. Scopus, an Elsevier owned product, was examined but found to include entirely resources found in the 5 databases listed above.

2.3 Keyword Searches

These databases were searched for keywords that are associated with research related to this project. The following keywords were used:

- Child labor
- Child Work
- Working Children

- Child Employment
- Economically Active Children
- Forced labor
- Debt-Bondage
- Forced Sexual Exploitation
- Human Trafficking

2.4 Ancillary Website Searches

Beyond the databases examined above, several websites were examined for relevant research papers. These websites were selected, because they were known to be sources of research on child labor, forced labor, and human trafficking. Websites were searched using Google's tools for searching within a website for the keywords listed above. The following websites were examined:

- BRAC
- BREAD
- CARE
- Catholic Relief Services
- Child Fund International
- Concern Worldwide
- Creative Associates International
- Desarrollo y Autogestion
- Devtech
- DFID
- Fair Labor Association
- FAO
- Free the Slaves
- Goodweave
- Grameen Foundation
- Heartland Alliance International
- ICF
- ILO
- Impaq International
- International Cocoa Initiative
- IPA
- IRC
- JBS International
- JPAL
- Management Systems International
- Mathematica
- MDRC
- NBER
- NORC
- Oxfam
- Partners for the Americas

- Private Agencies Collaborating Together (PACT) Incorporated
- Save the Children
- Solidarity Center
- Trickle Up
- UCW
- UN
- UNHRC
- UNICEF
- USAID
- USDOL
- Verite
- Winrock
- World Bank
- World Education
- World Vision

Many websites describe programmatic activities related to the areas of this bibliography. Only formal research available in research paper format and related to the areas of this study were evaluated for inclusion based on the criteria of the next section.

2.5 Criteria for Inclusion

This annotated bibliography will include *original* research papers, studies, and reports in the child labor and forced labor fields published in 2010 or later. Documents that are not original research are excluded. This will mean that literature reviews, bibliographies, and reports based on other studies are not included in this deliverable. Documents that are primarily historical in nature, referencing child labor or forced labor in contexts that predate 2000 are excluded.

Documents that are theoretical in nature or whose primary contribution is related to theoretical modeling concerns are excluded.

Included documents will be categorized as either assessments or impact evaluations. An assessment is defined as document that describes a setting where child labor or forced labor is present. For inclusion, the primary focus of the assessment must be on child labor, a specific form of child labor, forced labor, or a specific form of forced labor. Assessments in which topics related to child labor and forced labor are just incidental are not included. Assessments that rely on secondary reports and not original primary data collection are excluded. An impact evaluation is defined as a document that attempts to assess the impact of project or policy that is relevant to child labor or forced labor. "Relevance" is defined based on whether a component of child or forced labor is an outcome in the empirical analysis. Relevant impact evaluations do not need to have child labor or forced labor as a primary focus. This different treatments of assessments and impact evaluations is necessary, because most impact evaluations will focus on the intervention itself, not the topic of child labor or forced labor as the intervention is what is being evaluated.

2.6 Results of Database Searches

Table 1: Source of the 614 Articles Evaluated in this Review

Source	# Articles	Screened out by title	Excluded based on earlier search	Assessed for Eligibility	Excluded by abstract	Full text assessed	Excluded based on full text	Included
Google Scholar	8,550*	7,470	419	661	156	505	290	215
ProQuest	8,326	5,300	308	2,718	2,504	214	192	22
ERIC	846	126	120	600	496	104	101	3
PubMed	422	17	80	325	100	225	221	4
Web of Science Website Searches	1,676	617	0	1,059	541	518	224	294
	7,767	7,217	47	503	263	240	164	76

*The original Google Scholar search returned more than 26,000 articles. We limited the articles to all those published in 2019 and all those published between 2010-2018 with at least 5 citations.

Each index was searched in August 2019. Web of Science was the first index searched. Hence, none of the 1,676 articles found in that search were excluded based on an earlier search. A total of 27,587 articles were evaluated for inclusion in this literature review and a total of 614 articles met the criteria for inclusion.

There is year to year variation in number of articles included from the database search. Figure 1 plots the total number of included articles, across all sources listed above

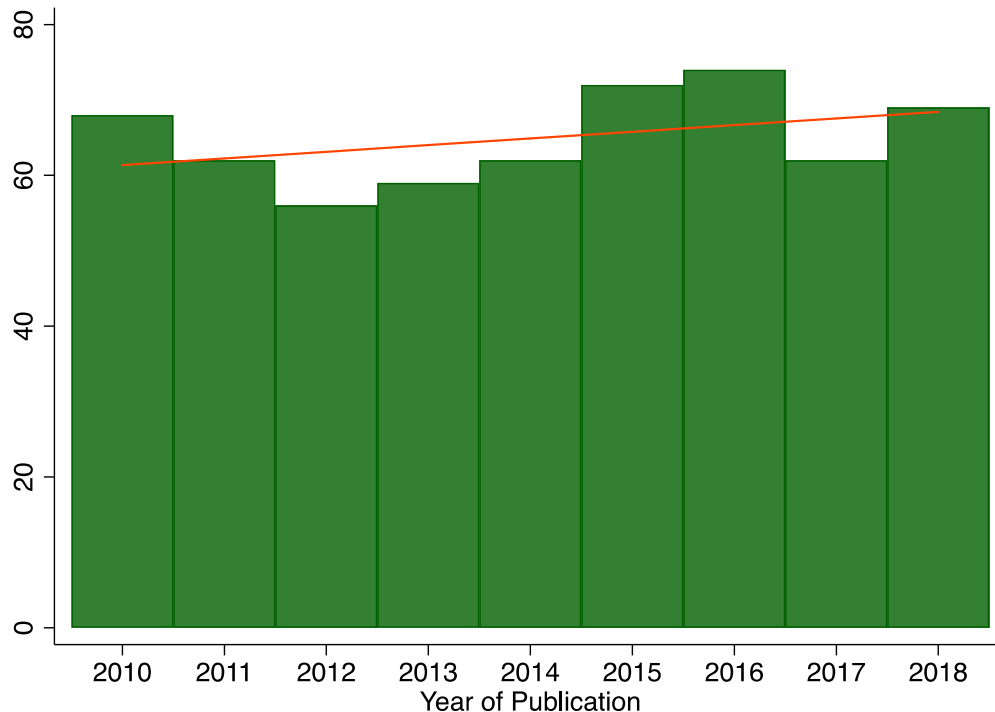


Figure 1: Number of Included Articles by Year 2010-2018

2019 is excluded from Figure 1, because we do not have a complete year and a different selection criterion was applied to 2019 because of the lack of time to accumulate citations in Google Scholar. Overall, there is a slight upward trend (pictured), but the variation between 2010 and 2018 from the trend line is smaller than some of the year to year variation in number of articles. Hence, Figure 1 cannot reject the hypothesis that there is no change in the number of articles that meet the search criteria over time. 35 percent of included articles come from Google Scholar, and it is worth emphasizing that from 2010-2018 articles found in Google Scholar (and not other databases) with fewer than 5 citations were not included. Citations accumulate over time. Thus, the selection criteria used to limit Google Scholar results would push us against finding an upward trend. A downward trend in Figure 1 would not have been surprising given the selection criteria applied to the Google Scholar results.

3 Findings

3.1 The forced labor literature is relatively large and growing

Most of the research found is related to child labor. Of the 614 identified studies for this review, 17 percent have forced labor as a primary outcome. While this might appear to be a literature dominated by child labor research, global estimates of child labor and forced labor put the number of people involved at 152 and 25 million respectively. This implies that the literature review produced found 0.4 studies per 100,000 forced laborers and 0.3 studies per 100,000 child laborers.

The forced labor literature is also growing more rapidly. Figure 2 plots the number of studies found in each year (as in Figure 1) but separately by whether the primary subject is child labor or forced labor.

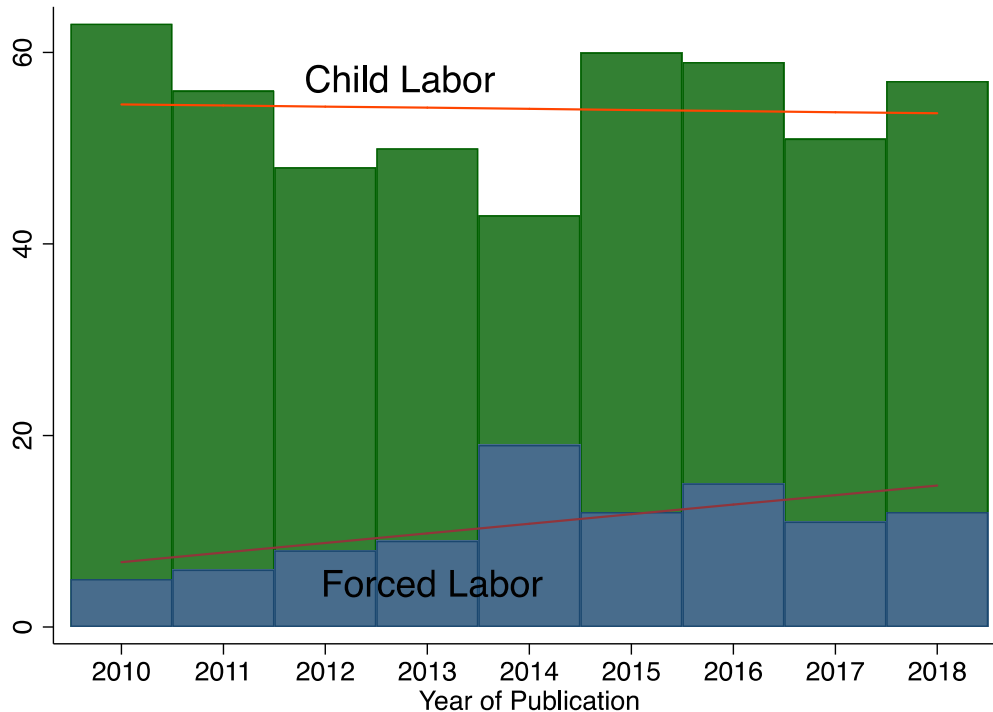


Figure 2: Number of Included Articles by Primary Topic and Year, 2010-2018

Also pictured are the trend lines. We see that while identified research in child labor is flat over this time period (actually slightly declining), forced labor research is on the rise. Hence, the imbalance in favor of forced labor (relative to prevalence) seems to be growing.

3.2 Child labor research is increasing in population, not economic activity rates

At the time of writing, country level estimates of the prevalence of forced labor do not appear to be available. However, economic activity rates of children 7-14 are available in the World Development Indicators. In this section, we document that child labor research is increasing in the number of economically active children.

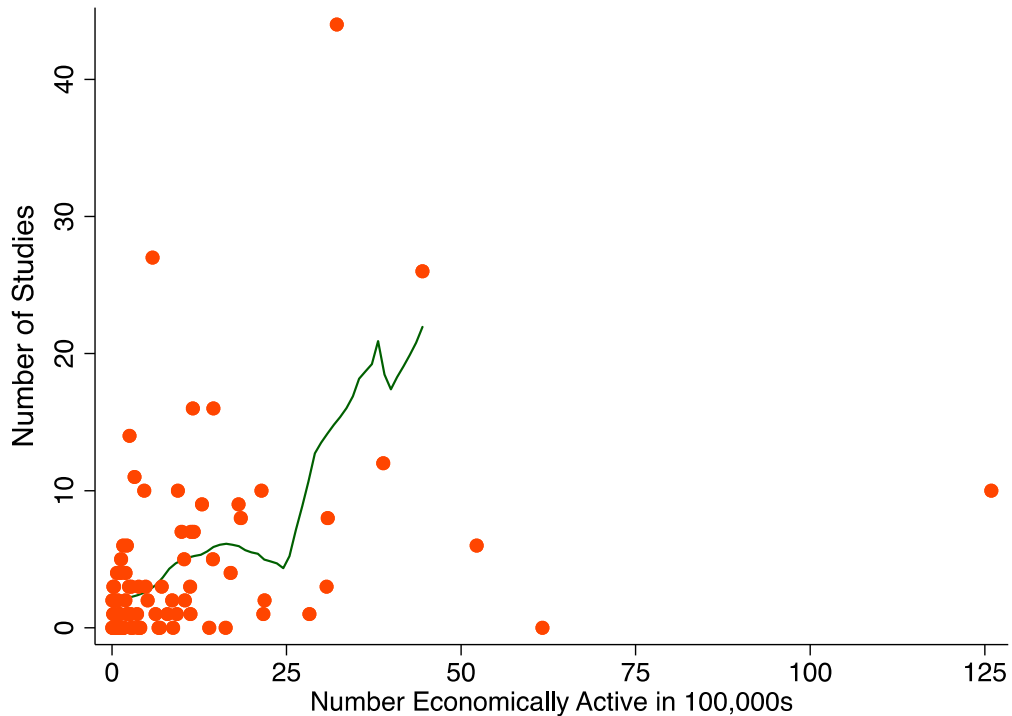


Figure 3: Number of Child Labor Studies and the Number of Working Children, 7-14

Source: World Development Indicators for economic activity rates and population

China is a problem in this examination. China might have the most child laborers in the world, but we do not have estimates of economic activity rates. With only 5 studies found, it is probably safe to assume that if China were able to be included in the analysis, it would attenuate the pattern observed above.

The pattern observed in Figure 3 in part reflects that there are more child labor studies in countries with more children (again the omission of China is important). Figure 4 depicts the number of child labor studies plotted against the natural log of the number of children 7-14. With China again omitted, the number of studies is increasing in population. India is the clear outlier with a large number of studies and children.

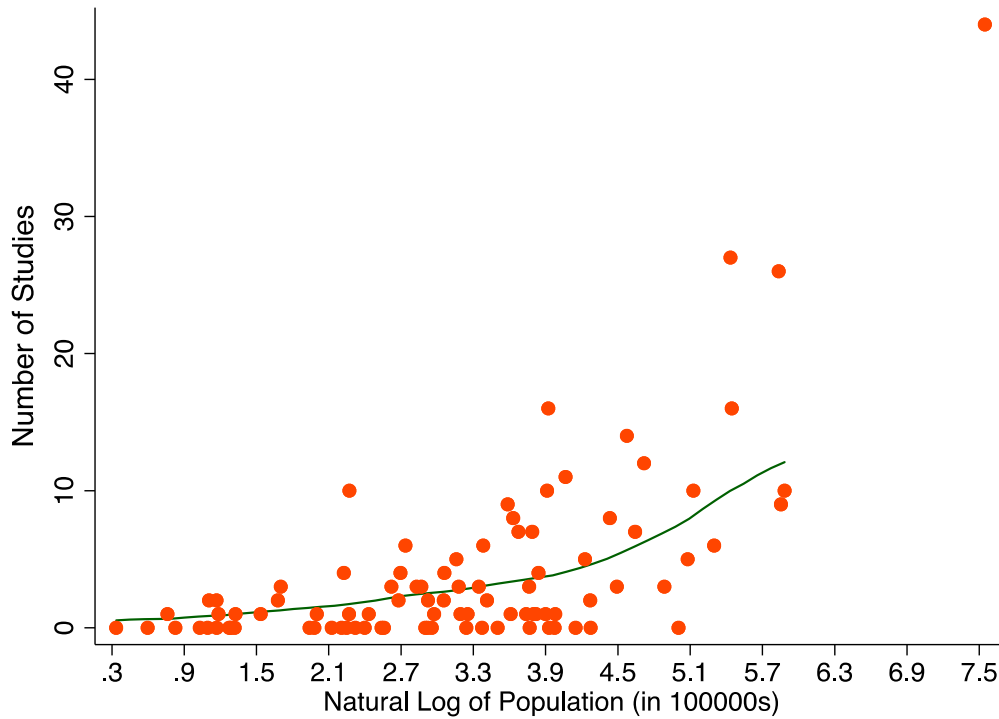


Figure 4: Number of Child Labor Studies and Population 7-14

Source: World Development Indicators for population

Abstracting from population, we do not see a clear relationship between economic activity rates and the number of child labor studies. This is evident in Figure 5. Economic activity rates are not strongly correlated with population. Hence, when we look at economic activity rates alone, they do not predict the number of child labor related studies.

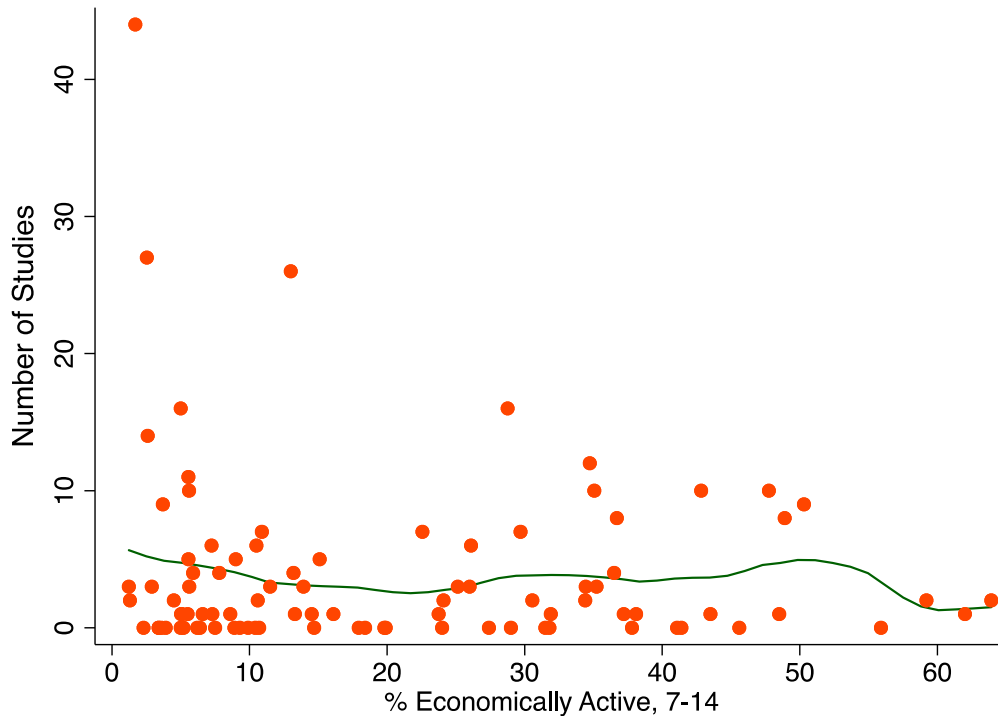


Figure 5: Number of Child Labor Studies and Economic Activity Rates, 7-14

Source: World Development Indicators for economic activity rates

The implication of Figure 5 is that there is not more child labor research in places where the fraction of children working is higher. From an anti-child labor standpoint, it is not clear whether policy should prioritize research in areas where there are more working children or where a larger share of children are working.

3.3 There are Few Randomized Control Trials

23 percent of identified studies are classified as impact evaluations. The definition of impact evaluations includes studies that examine the impact of factors that are not direct policy levers. Hence, fewer than a third of studies examine the impact of factors that are direct policy levers to combat child or forced labor.

Impact evaluations are especially rare among forced labor studies. 10 percent of forced labor studies are impact evaluations whereas 25 percent of child labor studies are.

A total of 6.5 percent of studies are randomized control trials of policy related interventions. This includes 1 forced labor related study and 39 child labor related studies.

The 1 forced labor study is an evaluation of different approaches to raising awareness about human trafficking (438), but it does not examine the impact of awareness raising on human trafficking itself. The study was not designed to do so and would be underpowered if that

analysis had been attempted. Hence, there are no impact evaluations aimed at examining interventions that are hypothesized to directly impact forced labor.

3.4 A majority of randomized control trials involve cash transfers

Within the 39 child labor related randomized control trials, 56 percent examine cash transfers. Most studies find that either conditional or unconditional transfers increase schooling and reduce child employment. A dominant theme in the literature is that the increases in schooling and declines in employment can be achieved without full replacement of child wages.

Interestingly, two studies find cash transfers increasing child employment. In one study from the Philippines (487), the authors find increases in child employment in reaction to a partial education subsidy that did not full cover the costs of schooling. Their interpretation is that the partial subsidy made education affordable when combined with additional child labor earnings. Hence, children worked more. In a study from Malawi, authors document increases in child employment in reaction to an unconditional cash transfer (134). In this case, the highest return investment for the beneficiary household was in household based productive assets such as livestock that children in turn would work.

These two examples illustrate the value of conducting a relatively large number of evaluations. Repeated replications of the same intervention in different contexts can help us understand when the intervention is an appropriate treatment for a problem and when it is not. In both cases where child employment increased, households may be better off, but if the focus is reducing child labor, the two examples highlight that cash transfers are not always the best solution for child labor.

4 Recommendations

4.1 Impact evaluations must have clear counterfactuals

There is no hierarchy of evidence, and randomized control trials should not be treated as the gold standard for impact evaluations in the child labor and forced labor space. However, an impact evaluation is informative only if the reader can understand how the researcher solved the problem of the counterfactual.

The problem of the counterfactual is that we observe outcomes of interest for the treated population (the factual), but we do not observe those same outcomes for the treated population absent treatment (the counterfactual). This is a problem, because the impact of a treatment is the difference between the outcomes observed on the treated and what would have happened to the treated group absent the treatment.

Every impact evaluation is comparing what is observed in the treated population to some guess of what would have happened in the treatment population absent treatment. The most common approach in impact evaluations in the child labor and forced labor space is to ask stakeholders in the treatment what how they think outcomes have changed with treatment. In these studies, the problem of the counterfactual is placed on the respondent, and there is no way to know how this counterfactual is formed.

The second most common approach is to track treated subjects over time. Often there is a baseline survey (for quantitative studies) or set of interviews, followed by an endline. In this approach, a counterfactual is formed by assuming that nothing changes over time for the treated other than the treatment. Sadly, one of the universal truths we know for the poorest of poor where child labor and forced labor are most prevalent is that their lives face a lot of volatility. A lot of things change over time, and it is unrealistic to ignore the influence of these confounding factors on measured outcomes or respondent's perceptions of how their lives have changed with the program.

Sometimes studies have control groups to control for changes that would happen over time absent the program. How are these control interviews chosen? Are they comparable to the treatment group in that they provide an approximation as to what would happen in the treated group absent treatment? Ex-post evaluations often attempt to identify this control group ex-post, but it is natural to worry about how treatment influence the comparability of this control group.

The appeal of randomization is that it simplifies the reader's understanding how the problem of the counterfactual was solved. Start with a group of potential subjects and use a lottery to determine their treatment status. Because treatment assignment is random, the control group presents a valid guess as to what would have happened to outcomes in the treatment group absent treatment.

A lot can go wrong in a randomization, especially when the number of things being randomized is small. By chance, researchers might end up with too many people with a particular outcome in the control or treatment population. That is why it is standard practice in any randomized control trial to begin with a table of pre-treatment characteristics to assess the comparability of the treatment and control groups. It is easy to identify problems in comparing the treatment and control groups with that sort of format. Such transparency is often missing from other approaches to impact evaluation.

4.2 Assessments and descriptions of forced labor and child labor situations need clear policy recommendations

77 percent of identified studies are assessments that describe a condition of child or forced labor. Such assessments help raise awareness about the circumstances of child and forced labor. 12 percent of assessments focus specifically on traded goods which may be especially relevant to the goal of raising global awareness around child and forced labor.

However, the volume of descriptions of situations of forced and child labor is large and growing, and the value added of further descriptive work is limited if there are not clear policy actions that follow out of that descriptive work. For example, documenting yet again that workers are more vulnerable to exploitation when they live within their worksite is not especially useful.

There appear to be two barriers to making these assessments more useful for policy. First, with relatively little impact evaluation research, there are no robust toolkits available for policymakers to apply when a situation of child or forced labor is documented and detailed. Second, the assessments themselves are often conducted in isolation without the deliberate goal of informing the different policy options that might be available in a given context. Even if the state of knowledge is not ready to tell us what works when, assessments could have more impact if they are conducted with attention to what policy levers are realistically available in the setting being studied.

4.3 Community monitoring and local enforcement models need more evaluation

Perhaps the most common approach to combatting child and forced labor around the world is to create community monitoring system that enforce local labor laws. Sometimes, these monitoring systems are in place in conjunction with local labor inspectors or law enforcement, but more often, they exist separately, outside of a formal legal framework.

We identified no randomized control trials of community monitoring programs. Four impact evaluations were identified using observational methods (363 364 440 441). Of those four, three only evaluated whether the program conveyed information about the relevant laws to participants. The one remaining had no clear research design. Hence, we effectively have no clear evidence on community monitoring and local enforcement programs. Given the difficulty of this work, it is unreasonable to just assume that this class of programs is effective at reducing child or forced labor.

4.4 Information interventions need more evaluation

Awareness raising activities are often inexpensive ways to combat child labor, so much so that they are not often evaluated. This is unfortunate, because although they are inexpensive, they may have substantive potential to influence child time allocation. Several education studies that did not arise from the above search criteria document large responses in schooling to information about returns to education. Could information provided to families with child laborers have an analogous effect?

Two studies provide some basis for optimism. 556 shows that coaching families about the dangers of child labor improved the impact of an economic strengthening program in Burkina Faso. 438 examines how to communicate messages related to forced labor and human trafficking in Nepal. While it is underpowered to find an effect on trafficking, it raises the possibility of using a positive, effective message to communicate with people about how to reduce their risks.

4.5 Research must be attentive to policy needs

So much of the research in child labor and forced labor area is funded by donor governments that we cannot assert that research is inattentive to the needs of policymakers. However, in reviewing the 614 studies included in this review, it was not obvious that research was considering the informational needs of the countries that were the subjects of the research. The fact that there appears to be no correlation between the prevalence rate of child labor and the number of studies done in a country should raise some questions about the targeting of research.

One vivid example that creates the impression of a disconnect between research and policy is that there appear to be no RCTs related to how child labor projects are actually implemented. In a typical child labor or forced labor project, an implementing agent engages in a wide array of activities in order to meet targets around the number of beneficiaries. One way to understand the goal of existing impact evaluation research is that it is trying to develop a toolkit that will allow implementing agents to have information on what should work in which circumstances. The

infancy of the literature means that it is not a very useful toolkit at present. However, building the toolkit abstracts from all the issues around what the implementing agent actually does.

Consider a world in which there was a fully developed toolkit of what works when. It is naïve to suppose that implementing agents will simply follow the toolkit or that the toolkit will contain a clear way to prioritize resources including time and money. More work needs to be done to understand the behavior of implementing agents and to study how explicit contracting incentives influence the activities of the implementing agents. There is no work in this area identified by this study.

4.6 Add-on questions to existing interventions are unlikely to inform marginalized populations

One strategy to cheaply improve learning on what works for child and forced labor is to add on relevant questions to studies that are not explicitly targeting children in child labor or forced labor or settings where children are not especially vulnerable. This has been tried for child labor related outcomes, and these studies with add on questions generally find no impact on child labor.

For example, several studies of microcredit related interventions (565 171 211 229 314 315 437 337 112 107) and education (587 506) ask questions related to child time allocation. In most of these cases, the study found no effect on child labor related outcomes. There are a few exceptions. One health and accident insurance product added on top of a microcredit loan appears to have reduced child labor in Pakistan (337). Child economic activity also appears to have increased in two studies (171 211).

While the vast majority of these microcredit and education related interventions find no effect on child labor related outcomes, it is impossible to tell whether the non-result is because of how the intervention was implemented or its target population. Statistical power can be an issue in general population studies where the presence of actual child or forced laborers can be low. We should not conclude based on these populations that generally microcredit has no meaningful effect on child labor if implemented in a population where child labor is prevalent any more than we should generalize from the two studies that find an increase in child work.

It is very difficult to interpret “no result” evidence from studies that were not designed to reach child and forced labor parties.

4.7 Progress on outcome measurement is critical to make impact evaluations more valuable

Another issue that arises in these studies that aren't targeted at child labor is measurement error in the child labor related outcomes. Generally, measurement error will bias research away from finding an effect of a treatment. When some questions are implemented as an afterthought or without proper piloting and localization, measurement error is apt to be a problem. This is likely an important issue with add-on questions or with researchers unfamiliar with measuring child labor and forced labor.

There has been an emphasis in developing a standardized set of questions related to child and forced labor. Without appropriate piloting and localization, these generic questions may generate measurement error that undermines the ability of impact evaluations to measure changes in child labor and forced labor. If measurement error from standardized questions leads to false conclusions of no effect of a treatment, it may actually diminish rather than create knowledge of what works for working children and forced labor.