A Multi-Modal Approach to Estimating Prevalence: The Case of Labor Trafficking

Laura V. Zimmermann, David Okech, Hui Yi, Lydia Aletradis

University of Georgia*

Abstract

Policymakers, researchers, and other stakeholders often care about the prevalence of complex, sensitive outcomes or behaviors, such as human trafficking and forced labor, gender-based violence, child labor, or corruption. But the measurement of such issues is challenging: Comprehensive administrative data is usually unavailable, survey respondents may define these concepts very differently than existing legislation, and social stigma may lead to substantial underreporting. In this paper, we set up a conceptual framework that identifies four theoretically distinct types of prevalence rates. We then propose a survey design approach that is practically feasible in many contexts and that, with only one additional assumption beyond those commonly made in the literature, can recover all the conceptual information. We apply that approach to estimate the prevalence of labor trafficking in two countries, Malawi and Zambia.

Keywords: measurement, prevalence, labor trafficking, forced labor, human trafficking, list experiment, survey experiment, social stigma, working conditions, exploitation, safe work, multi-modal, Malawi, Zambia

JEL: Codes: I32, J47, J61, J81, J82

*Financial support from the Department of State, Office to Monitor and Combat Trafficking in Persons Grant SSJTIP22CA0029 is gratefully acknowledged. The views and opinions expressed in this paper are those of the authors and do not necessarily represent the views or positions of the Department of State.

1 Introduction

Policymakers, researchers, non-government organizations and other stakeholders are often interested in measuring the prevalence of outcomes or behaviors. Prevalence estimates allow tracking progress, designing and testing better policies, and spotting implementation issues, for example. But especially for complex, sensitive issues, such as human trafficking and forced labor, gender-based violence, child labor, or corruption, measuring how widespread they are is challenging: Comprehensive and accurate administrative data representative of the underlying population is usually unavailable, making surveys an important data collection tool. Survey respondents may define these concepts very differently than the official national or international legal definition of the term, however. And social stigma, fear of persecution, or distrust in authorities may lead to substantial misreporting.

So what needs to be considered in the collection of prevalence data on complex, sensitive variables? In this paper, we first set up a conceptual framework that identifies four theoretically distinct types of prevalence rates that differ on two dimensions. The first dimension is whether the used definition is the formal one, for example based on a national law or an international treaty, or the subjective one used by the survey respondent. The second dimension is whether the prevalence rate is the one a respondent is willing to share directly during a survey or is their internal truth. Each possible combination yields a prevalence rate that could be of interest. Our conceptual framework makes the required decisions for a given application explicit, which should guide the appropriate data collection efforts and the correct interpretation of the information for policy purposes.

We argue that with a careful survey design approach and subject to assumptions commonly made in the existing literature, at least three of the four prevalence rates can be directly estimated in one survey. The fourth one, the internal prevalence rate that uses the formal definition of the term of interest, may be too complex in many contexts to yield high-quality responses. In these situations, we show that this prevalence rate can be recovered from the other prevalence rates by making one additional assumption. This approach

is practically feasible in many situations and does not require adding many extra questions. Beyond the prevalence rates themselves, it also provides information on the importance of the definition dimension relative to the social stigma dimension as an automatic by-product, which may in itself yield highly policy-relevant insights that so far have rarely been exploited in their comprehensiveness: In many contexts, proposed policy solutions to address a complex issue will be very different if the main concern is low awareness of the official legal definition of a concept than when the dominant hurdle is social stigma. So estimating the relative importance of the two dimensions may help prioritize policy interventions.

We apply this conceptual framework to estimate the prevalence of labor trafficking, a complex and sensitive issue of high policy relevance. An estimated 27.6 million people worldwide are subjected to forced labor every single day, generating annual profits of 236 billion dollars for perpetrators.[1] While forced labor is a concern in all countries including high-income countries, the incidence is believed to be highest in low-income countries (ILO, Walk Free and IOM, 2022; ILO, 2024).[2] To combat this labor trafficking, UN Sustainable Development Goal 8.7 seeks to end modern slavery by 2030. But accurately estimating the prevalence of labor trafficking is difficult due to low availability of administrative data and concerns about the representativeness of known cases relative to the total population of trafficking victims and survivors.

Surveys present a promising alternative but come with their own challenges. Individuals need to understand concepts like forced labor and how they are defined. If awareness is low, or if respondents' own definitions of the key terms are wider or more narrow than those used by researchers and policymakers, the gap between formal and subjective prevalence rates could be high. Connections to the perpetrator or the perceived normalcy of similar experiences in the community could lead individuals to categorize incidents as less severe than existing international legislation, for example. If being a survivor of labor trafficking

[1] This translates into 3.5 out of every 1000 people worldwide in 2021 and an increase by 2.7 million cases relative to the most recent pre-pandemic estimates from 2016 (ILO, Walk Free and IOM, 2022).

[2] 6.3 per thousand people in low-income countries as compared to 4.4 per thousand people in high-income countries.

is associated with personal shame, social stigma or fear of persecution, individuals may also under-report their experiences.

We collect information on labor trafficking through household surveys representative of the youth and young adult (YYA) population aged 18-37 years in two developing countries. We fous on two districts in Malawi and four districts in Zambia identified as high-migration districts based on Census data. Respondents are randomized into either the treatment or the control group, which determines which question wording they receive in two survey experiments.

After some general questions on socio-demographic characteristics, current work experience and wages, participants are first asked to respond to a list experiment: Both treatment and control groups are read a set of statements and are asked to report the total number of statements that is true for them, but not which specific statements those are. While the control group receives four statements on other topics, the treatment group receives the same four statements plus an additional statement on having experienced forced labor in the past two years. This setup provides more privacy to respondents in the treatment group to truthfully report a sensitive experience, and the randomization allows us to back out the estimated internal prevalence rate using individuals' own definitions of forced labor.

After the list experiment, the control group is directly asked whether they have experienced forced labor in the last two years, equivalent to the expressed prevalence rate using individuals' own definition of the term. Comparing the direct question prevalence rate to that of the list experiment allows us to test whether the increased privacy of the survey experiment makes respondents more likely to report.

To explore how different prevalence rates are if the used definition is the existing formal one, both treatment and control group are then asked direct questions about a range of experiences that map to the UN Palermo Protocol definition of labor trafficking. This provides us with a third measure of forced labor prevalence from aggregating responses to the relevant 'objective' UN thresholds of labor trafficking.

In addition, this module may serve to raise the awareness among survey participants about what kinds of experiences count towards forced labor, improve recollection and affect the way in which connections are made to personal experiences. After this module, we therefore implement a second list experiment to test whether responses are updated, followed again by the direct question on forced labor.

We find that the estimated prevalence of labor trafficking in Malawi in the first list experiment is 17 percent, more than three times as high as the 5 percent reported if the forced labor question is asked directly. This suggests that privacy concerns matter to respondents. In the Zambian sample, on the other hand, we cannot reject the hypothesis that the prevalence rates between direct question and list experiment are the same. Among the sub-sample of respondents who reported having worked at least for some time in the last two years, the list experiment prevalence rate is about 20 percent in both countries.

When comparing the first and the second set of list experiments, we find very little change in the overall estimates. Neither responses to the direct forced labor question nor to the list experiment appear to get updated substantially after responding to an extensive set of detailed behaviors and experiences that would count towards forced labor. However, in the Zambian sample this masks heterogeneity between rural and urban areas where rural respondents revise their reporting upwards by about the same magnitude as urban respondents revise their reports downwards. There is little evidence of similar updating in Malawi. In Malawi, response patterns are consistent with participants believing that they have a clear idea of what forced labor entails. The gap between reporting in the list experiment and the direct questioning remains, further supporting the importance of added privacy for accurate responses. In Zambia, ongoing analyses explore the potential mechanisms behind the updating process, which appears to suggest that while there is low social stigma, the detailed survey questions may have provided additional information that led to the refinement of personal definitions of forced labor.

We can also test how the self-reported prevalence rates compare to the UN labor traffick-

ing thresholds that can be constructed from the detailed survey questions on different types of experienced exploitation. This allows us to check for the likely definition and understanding respondents have of the term forced labor relative to its UN definition. In first results, we find that the prevalence rates from the individual behaviors lie between those reported in direct questions and list experiments, but align more closely with the forced labor direct reports. For Malawi, this suggests that the definition dimension is less important than the social stigma dimension.

Overall, our surveys provide labor trafficking estimates in a context where little data is available and show that while added layers of privacy are important in Malawi, the aggregated estimates in both Malawi and Zambia are robust to multiple other measurement concerns. This has potentially important policy implications in contexts where prevalence estimates are heavily reliant on survey data.

Our paper adds to the understanding of labor market outcomes for workers from developing countries. An existing literature analyzes the domestic and international challenges of working-age adults to find jobs, the sensitivity of employment decisions to wage changes and shocks, and the migration and human capital investment decisions individuals make (Beam, McKenzie and Yang, 2016; Goldberg, 2016; Shrestha and Yang, 2019; Theoharides, 2020).[3] This paper adds more detailed and explicit measurement and definitions of labor trafficking that operationalize concepts used in the international policymaking community. We show that about 20 percent of young adults in Malawi and Zambia are survivors of forced labor. To the best of our knowledge, this is the first paper across the social sciences that explores whether list experiments improve on the prevalence rate estimates relative to direct questioning and whether respondents update their responses during the survey.[4]

We also contribute to the literature on list experiments. List experiments are still a fairly

[3]See e.g. Bandiera et al. (2022) for a recent overview. They document that young adults in many African countries are less likely to get paid for their work, are less likely to have a salaried job, and are more likely to be in self-employment than young adults elsewhere.

[4]For a survey experiment on perceived relative economic deprivation and potential vulnerability through increased risk-taking see Mo (2018).

new type of survey experiment in economics.[5] Important technical advances have been made to test underlying assumptions and suggest estimation techniques as well as robustness, much of it in political science. We add to this by clarifying the additional measurement and reporting concerns that are not typically considered in other list experiment applications, and test the robustness of list experiment estimates to these issues. Our work also contributes to the literature on measurement error in household surveys, where having the ability to see whether respondents update their answers during the survey is rare.

More broadly, our multi-modal approach provides added clarity and nuance for researchers and practitioners interested in the estimation of prevalence rates for complex, sensitive issues. To the best of our knowledge, this is the first paper to explicitly distinguish the four concepts of prevalence rates in a conceptual framework, match them to the best way of estimating them in the existing literature, and propose a systematic survey design approach that is practically feasible in many situations while also allowing information on all four rates and the importance of definition and stigma dimensions to be recovered simultaneously.

2 Conceptual Framework

2.1 Four Types of Prevalence Rates

With complex and sensitive issues, there are four distinct prevalence rates differing on two dimensions. The first dimension is whether the used definition is the formal one, for example based on a national law or an international treaty, or the subjective one used by the survey respondent. For complex concepts, the average individual's definition could differ quite substantially from the official one, for example due to lack of awareness, cultural or local factors.

The second dimension is whether the prevalence rate is the one a respondent is willing

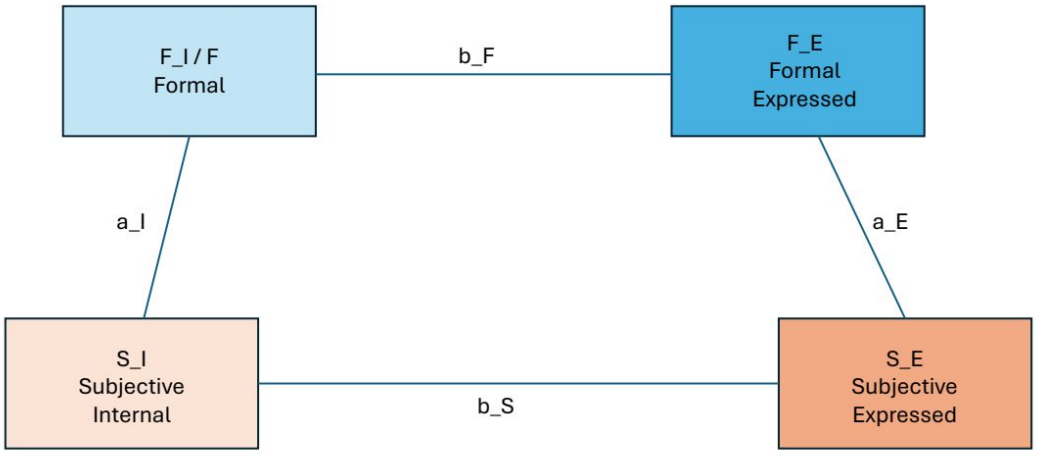
[5]For examples, please see Chuang and Seban (2021) and Cullen (2023).

to share during a survey or is their internal rate. This difference between internal and expressed prevalence could be substantial for sensitive issues where respondents fear social stigma, persecution or retaliation, or where they distrust authorities or the interviewer.

The possible combinations of the concepts lead to four different prevalence rates. We denote the prevalence rates by upper-case letters: formal expressed, which we refer to as F_E; subjective internal, which we call S_I; and subjective expressed, which we call S_E. The fourth combination, formal internal, could be referred to as F_I, but more cleanly this also denotes the formal prevalence F, i.e. the prevalence rate an independent objective observer with complete information about the sample of individuals would create when using the formal definition of a concept.

2.2 The Prevalence Rate Quadrilateral

Figure 1: The Prevalence Rate Quadrilateral



Legend: F_I or F: formal internal prevlaence rate; F_E: formal expressed prevalence rate; S_I: subjective internal prevalence rate; S_E: subjective expressed prevalence rate.a_I and a_E: importance of definition dimension. b_f and b_s: importance of stigma dimension.

The relationships between the four types of prevalence rates can be visualized in Figure 1 above, where a_I and a_E show how far apart prevalence rates are due to differences in definition

among the internal and external truths, respectively. b_f and b_s show the magnitude of the social stigma and of other factors that lead to differences between expressed and internal prevalence among the formal and subjective definitions, respectively.

A plausible hypothesis is that a_I and a_E are of similar magnitude: How the survey respondents subjectively define the outcome of interest should not be affected by whether or not they share their responses with the interviewer.

That social stigma is similar for the formal definition and the subjective definition may also be a plausible simplifying assumption but is less likely to be true *a priori*. In both cases respondents consider the social stigma entailed in admitting, say, to being survivors of labor trafficking. But depending on how systematically formal and subjective definitions differ, the social stigma of one may be higher than that of the other. If respondents have, say, a narrower definition of labor trafficking than the formal definition that focuses only on the most severe forms of forced labor, then the social stigma from revealing experiences with them may be higher than those from a wider definition that can also include some less severe instances that are perceived as fairly common.

2.3 Comprehensive versus Disaggregated Questions

While researchers and stakeholders may be more interested in some of these prevalence rates than others, in practice a third decision is made in survey design that impacts the conceptual measurement. When studying a complex concept with many different nuances or types of behavior that fall under a given definition, surveyors can either opt to ask about the concept comprehensively or they can pick a sub-set of experiences to ask about specifically. For a concept like labor trafficking, for example, one could ask about experiences with forced labor directly or provide specific examples that would fall under forced labor but that may not capture the full nuance of the legal definition.

These distinctions mainly matter when the formal definition of an outcome of interest is so complex that it cannot be easily shared with a respondent during the survey. As we

will discuss in more detail in the Background section below, to qualify formally as labor trafficking, for example, requires an act (recruitment, transportation, transfer, harboring, receipt, or soliciting), a means (force or other forms of coercion, abduction, fraud, deception, the abuse of power or a position of vulnerability) and a purpose (forced labor or services, slavery or practices similar to slavery, servitude) (Okech et al., 2021). Since this information is difficult to share fully with respondents in one go, surveyors may instead opt to ask a number of questions about specific aspects of the formal definition, such as recruitment or experiences of coercion, that are easier for a respondent to operationalize. Similarly, studies of intimate partner violence (IPV) often ask questions about specific different examples of physical or emotional violence rather than expecting that a survey respondent will be perfectly aware of the entire breadth of experiences that would fall under IPV.

If questions are asked in a disaggregated manner, there will be a prevalence rate quadrilateral for each question. For very specific descriptions of behaviors or experiences, however, there is little room for a subjective interpretation that is different from the formal one, so a_I and a_E will be about 0. The quadrilateral will then collapse to just one dimension, the difference between the internal and expressed prevalence rate.

Individual questions can be aggregated to generate prevalence rates at the higher level. If the questions do not cover all potential sub-cases or nuances of the overall concept, these will be partial prevalence rates. On the other hand, if questions can be aggregated to match the thresholds in the overall definition, the aggregated prevalence rate corresponds to the formal comprehensive definition of the outcome of interest.

3 Survey Design Approach: Empirically Identifying Prevalence Rates

In this section, we focus on how each of the four prevalence rates can be empirically estimated.

3.1 Estimating S_E

The expressed subjective prevalence rate is the easiest to collect since it comes from directly asking survey questions. To get the expressed subjective prevalence rate of forced labor, for example, we can directly ask respondents whether they experienced forced labor in some defined time interval. Since the survey does not define forced labor, respondents will draw on their own personal definition of forced labor when answering. Social stigma and other factors like fear of persecution or retaliation will also determine what an individual is willing to express.

For researchers and practitioners, the concern with this prevalence estimate is that it is unclear how big social stigma is and how much the subjective definition differs from the formal definition. At the same time, understanding how individuals see themselves and their experiences can be very valuable when designing policy interventions to support survivors or prevent abuse. Policy tools for individuals that view themselves as survivors may be different than the optimal intervention for individuals that do not view themselves as survivors despite meeting the formal definition thresholds.

3.2 Estimating S_I

More recently, various survey design techniques have been developed and improved upon that seek to provide more privacy to respondents, thereby ideally eliminating social stigma and other reasons for misreporting. The most well-established method to get at the internal subjective truth is a list experiment. Assuming the assumptions are met, the fact that respondents only share a number of statements that are true for them rather than their responses to individual statements eliminates b_s. A list experiment where the treatment statement is, say, that the respondent has experienced forced labor in a specified time period would yield an estimate of S_I. It can be compared to the prevalence estimate S_E from above to calculate the magnitude of the social stigma b_s.

Identifying S_I is conditional on the assumptions of a list experiment being met, so list

experiments need to be designed carefully and internal consistency and identification checks proposed in the recent literature need to be carried out.

3.3 Estimating F_E

For complex issues where the formal definition is long, complicated and with many nuances or sub-cases, the best way of collecting data on the expressed formal prevalence rate is to break the overall concept down into a series of more specific questions that can be aggregated to an overall prevalence rate in accordance with thresholds identified in the formal definition.

In the labor trafficking application, this is achieved by asking the PRIF module and using appropriately aggregated responses from this module to generate prevalence rates of labor trafficking that are consistent with the UN Palermo Protocol (Okech et al., 2021).

Comparing this estimate of F_E to the estimate of S_E provides us with a_E, which tells us how different the subjective expressed prevalence is from the formal expressed prevalence.

3.4 Estimating F_I / F

In most applications of complex sensitive issues, F_I / F cannot be feasibly directly estimated. Doing so would require creating a list experiment for each of the PRIF module questions, for example. As recent papers on list experiments have shown, it is difficult to design a large number of list experiments while ensuring that the assumptions are met. List experiment questions also take longer to answer than direct questions, so a large number of list experiments would add substantial interview time, increasing the risk of survey fatigue, especially when these questions are integrated into a larger survey. Alternatively, recovering F would require having an independent objective observer with perfect information who can correctly assess whether each individual in the sample meets the formal threshold for the outcome of interest. This is typically not possible. Administrative data is also usually either unavailable or not representative of the underlying population, so cannot be used as the gold standard.

At the same time, F is usually the prevalence rate that policymakers implicitly or explic-

itly have in mind when collecting data to track progress towards goals such as the Sustainable Development Goals or to test whether a specific intervention is effective. Fortunately, F_I / F can be estimated or at least bounded by making assumptions about a_I and/or b_F, so by imposing one additional assumption above and beyond the assumptions typically made in the literature.

If it is plausible that a_E and a_I are of similar magnitude, as argued above, then the best estimate of F is the estimate of S_I plus the estimate of a_E. Similarly, if one is willing to make the assumption that b_F is of about the same magnitude as b_S, then an estimate of F can be derived from taking the estimate of F_E and adding b_S. Alternatively, one can provide bounds similar in spirit to Lee bounds by making assumptions about the minimum and maximum magnitudes of a_I and/or b_F. A good internal consistency check is whether the estimate of F derived from S_I plus a_I is similar to the one derived from F_E plus b_F.

3.5 Summary of Survey Design Intervention

Taken together, the suggested order in a survey is therefore:

1. Randomly allocate respondents to either the treatment group or the control group
2. Ask the list experiment where the treatment statement is the concept of interest (for example forced labor), given to the treatment group
3. Ask the control group the equivalent question of the treatment statement from the list experiment
4. Administer the questions (potentially disaggregated) to measure F_E

As far as we are aware, this approach extends what is usually done in the literature, thereby generating new insights: List experiment papers usually estimate S_I and S_E, but typically cannot say anything about the difference between formal and subjective prevalence rates. This information may often be valuable when designing incentive-compatible policies,

or when deciding on the right kind of policy tool. Awareness or information campaigns may be an effective tool when the difference between formal and respondent definitions is large, for example, but be unlikely to be a first-order concern if the difference is small.

Examples from the literature that carefully use sub-questions and then aggregate them to thresholds that meet the formal definitions and provide an estimate of F_E are rare in much of the social science literature. It is more common to employ a small set of vignettes or questions on specific behaviors, for example in the intimate-partner violence literature or the corruption literature. Even then, without the addition of S_E and S_I the estimate of F_E only allows the recovering of F if social stigma is close to zero. Adding the estimates of S_E and S_I is feasible in many studies since the added number of questions and extra response time is small.

Overall, our proposed approach is often practically feasible and allows a number of additional conceptual insights that are not possible by focusing just on one side of the rectangle as both of the established approaches in the literature implicitly do.

We now turn to applying this framework to estimate the prevlaence of labor trafficking in Malawi and Zambia.

4 Background on Labor Trafficking

4.1 Trafficking Definitions

Article 3 of the UN Protocol to Prevent, Suppress and Punish Trafficking in Persons, also referred to as the Palermo Protocol, defines human trafficking as follows (United Nations, 2000):

> *(a) 'Trafficking in persons' shall mean the recruitment, transportation, transfer, harbouring or receipt of persons, by means of the threat or use of force or other forms of coercion, of abduction, of fraud, of deception, of the abuse of power or of a position of vulnerability or of the giving or receiving of payments or*

benefits to achieve the consent of a person having control over another person, for the purpose of exploitation. Exploitation shall include, at a minimum, the exploitation of the prostitution of others or other forms of sexual exploitation, forced labour or services, slavery or practices similar to slavery, servitude or the removal of organs;

(b) The consent of a victim of trafficking in persons to the intended exploitation set forth in subparagraph (a) of this article shall be irrelevant where any of the means set forth in subparagraph (a) have been used;

(c) The recruitment, transportation, transfer, harbouring or receipt of a child for the purpose of exploitation shall be considered 'trafficking in persons' even if this does not involve any of the means set forth in subparagraph (a) of this article;

(d) 'Child' shall mean any person under eighteen years of age.

To qualify as trafficking therefore requires three elements for individuals aged 18 and above: An act (recruitment, transportation, transfer, harboring, receipt, or soliciting), a means (force or other forms of coercion, abduction, fraud, deception, the abuse of power or a position of vulnerability), and a purpose. The purpose breaks instances of human trafficking down into its two main sub-categories: sex trafficking (exploitation of the prostitution of others or other forms of sexual exploitation) and labor trafficking (forced labor or services, slavery or practices similar to slavery, servitude) (Okech et al., 2021).

Consistent with the Palermo Protocol's description of labor trafficking, the U.S. Department of State defines forced labor as (U.S. Department of State, 2016)

labor obtained by any of the following methods: the recruitment, harboring, transportation, provision, or obtaining of a person for labor or services, through the use of force, fraud, or coercion for the purpose of subjection to involuntary servitude, peonage, debt bondage, or slavery.

Crucially, both definitions clarify that while abduction, force or coercion are sufficient to satisfy the means portion of an instance of forced labor or labor trafficking, they are not necessary. The labor trafficking threshold may also be met through fraud or deception as long as the work experience satisfies the criteria for act and purpose. This implies, for example, that individuals may voluntarily agree to a job under false pretenses. Similarly, the movement of individuals is not necessary to meet the labor trafficking threshold, which means that workers may experience trafficking without migrating.

4.2 Labor Trafficking in Malawi and Zambia

Anecdotal concerns about labor trafficking in the neighboring countries of Malawi and Zambia are high and show a large variety of channels through which youths and young adults end up as victims and survivors of forced labor. The U.S. Department of State lists both countries as so-called Tier 2 countries, which do 'not fully meet the minimum standards for the elimination of trafficking but [are] making significant efforts to do so'.[6] The plight of around 400 Malawian women who were able to escape modern-slavery-like positions as domestic helpers in Oman has received international attention.[7] Malawian men and women have been forced to work in agriculture, been offered seemingly lucrative contracts on tobacco estates in Zambia, or promised employment in South Africa or the Middle East.[8]

In Mangochi, one of our survey districts, the prevalence of human trafficking is believed to have increased in recent years. According to the Mangochi police station coordinator Inspector, contributing factors are the low engagement and lack of knowledge on human trafficking in the population, poor training of police officers, lack of coordination between government and non-government organizations, and the issue that many cases of trafficking may begin by voluntary agreements of workers to seemingly generous employment or scholarship offers

[6]See https://www.state.gov/reports/2023-trafficking-in-persons-report/zambia/ and https://www.state.gov/reports/2023-trafficking-in-persons-report/malawi/

[7]See e.g. Phiri and Ford (March 18, 2024), 'How a Malawi WhatsApp Group Helped Save Women Trafficked to Oman', BBC. See https://www.bbc.com/news/world-africa-68565425

[8]'Forced Labour, Prostitution and Child Marriages in Malawi', June 16, 2023. https://africa.cgtn.com/forced-labour-prostitution-and-child-marriages-in-malawi/

abroad.[9]

Similarly, organizations like UNICEF have reported on cases of trafficking from Zambia's Sesheke district, one of our survey districts. As in Malawi, trafficking in Zambia can take many different forms and involve a wide range of perpetrators, including promises being made of employment and medical treatment.[10]

5 Data and Empirical Specification

5.1 Data

We draw on representative household survey data on young adults aged 18-37 years in 3 rural and 3 urban districts across Malawi and Zambia identified as high-migration areas according to Census data.[11] In Malawi, the two districts are Blantyre (urban) and Mangochi (rural). In Zambia, the districts include Lusaka and Livingstone (urban) as well as Chipata and Sesheke (rural).

In each district, a target number of 500 households were chosen using a two-stage stratified sampling design. 63 enumeration areas were first chosen with probability proportional to size based on the most recent Census in both countries. In each enumeration area, 8 households were then selected by systematic random sampling. In each household, the most knowledgeable adult household member was administered a questionnaire on household-level information and completed a household roster. Each youth and young adult (YYA) aged 18-37 years who resides in the household was given a separate young adults survey, covering socio-economic background questions, information about recent employment and financial capability in addition to detailed survey modules on labor trafficking.

[9]'Mangochi Targeted by Human Traffickers', Malawi24, October 2, 2021. https://malawi24.com/2021/10/02/mangochi-targeted-by-human-traffickers/

[10]Shahryar (October 20, 2023), 'Children Should Not Be Trafficked, They Belong at Home with their Families', Unicef. https://www.unicef.org/zambia/stories/children-should-not-be-trafficked-they-belong-home-their-families

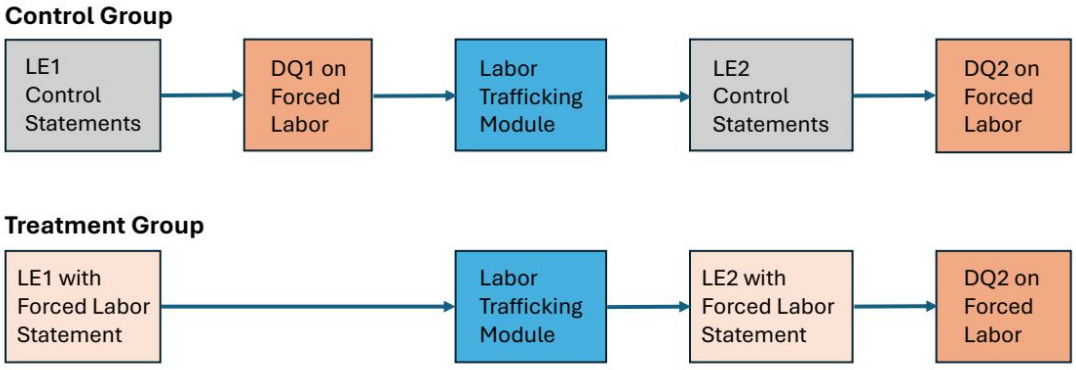
[11]The study was pre-registered in the American Economic Association trial registry.

Data collection took place in November and December of 2023. In total, the final Malawian sample includes 562 households (983 YYAs) from Blantyre and 568 households (880 YYAs) from Mangochi. In Zambia, the data contains 507 households (795 YYAs) from Lusaka, 505 households (880 YYAs) from Livingstone, 504 households (847 YYAs) from Chipata, and 508 households (709 YYAs) from Sesheke.

5.2 Intervention Overview

After answering some questions on socio-economic characteristics and their job market experience, respondents in the treatment and control groups are exposed to different versions of survey questions. Figure 2 shows the intervention setup stages graphically.

Figure 2: Overview of Survey Structure



Legend: This figure shows the structure of the survey. LE1 and LE2 refer to list experiments 1 and 2, respectively. DQ1 and DQ2 refer to a direct question on forced labor. In Zambia, DQ2 was only asked to the control group.

The control group first answers the list experiment question, but is just exposed to the control group version of the statements, which include no questions on labor trafficking or forced labor.[12] We refer to this list experiment as LE1. Afterwards, a respondent is

[12]Some of the control group statements for Malawi and Zambia are different, so the results are comparable across countries if the assumption holds that the specific control group statements do not affect the treatment group's behavior.

asked directly whether they have experienced forced labor in the last two years, which we refer to as DQ1. After some questions on other labor topics, a respondent then answers questions from an entire survey module devoted to questions on labor trafficking. While the words labor trafficking or forced labor are not explicitly mentioned anywhere, the questions focus on specific experiences that would fall under the UN Definition of labor trafficking, either by themselves or in combination with other indicators. These include questions about experienced physical and psychological violence, fear, losing freedom of movement, having identification documents taken away, and various forms of deception about aspects of the job (Okech et al., 2021).

We extend the survey design setup from above by exploring whether the Labor Trafficking Module leads respondents to update their personal definition of the term forced labor. While the questions in this module do not directly provide the full formal definition to individuals and do not share information about the threshold that needs to be crossed, the more specific examples or sub-cases could allow respondents to refine their own definition to more closely match the formal definition. To the extent that the F_E sub-questions are priming respondents that this topic is important or are lowering recollection error about specific events, this might also lead to updating. Whether updating happens is a very understudied topic in general and especially so for prevalence rates.

After the module, the control group participants are therefore exposed to the control statements of a second list experiment (LE2), which exactly match the statements from the first list experiment, after which they are directly asked again whether they have experienced forced labor in the last two years (DQ2).

For respondents in the treatment group, the setup is instead the following: They are first exposed to the treatment group version of the list experiment (LE1), which includes the statement of having experienced forced labor in the last two years in addition to the same four statements that the control group receives. They skip the direct question on forced labor (DQ1). Instead, they receive the same module on labor trafficking as the control group, with

direct questions on a range of potential indicators of labor trafficking. Participants are then given the second list experiment (LE2), with the exact same statements as the first list experiment. In Malawi, respondents are then asked directly whether they have experienced forced labor in the last two years (DQ2), whereas they do not receive this questions in the Zambian sample.

5.3 Empirical Specification

Given the randomization of respondents into treatment or control group versions of the list experiments, the prevalence rate of forced labor can be estimated using the following simple regression

$$y_i = \beta_0 + \beta_1 T_i + u_i$$

where y_i is the number of true statements respondent i reports, and T_i is an indicator equal to 1 if the individual is part of the treatment group (and 0 otherwise). β_1 captures the difference in the mean number of statements between the two groups and provides an estimate of the prevalence of forced labor in the sample. Heteroskedasticity-robust standard errors are estimated.

We estimate prevalence rates both in the overall sample of all respondents as well as for the sub-sample of young adults who worked in the last two years in both Malawi and Zambia. We also explore heterogeneity between rural and urban areas.

6 Results

6.1 First List Experiment and Direct Question: LE1 and DQ1

Tables 1 and 2 show the results for the first list experiment and the direct question on forced labor for the Malawian and Zambian samples, respectively. Panel A in both Tables shows the estimates for the total sample of all young adults, whereas Panel B restricts the sample

Table 1: Comparison of LE1 and DQ1 (Malawi)

	Panel A: All YYAs		
	(1) total	(2) rural	(3) urban
List Experiment (LE1)	0.1749***	0.1897***	0.1598***
	(0.0426)	(0.0624)	(0.0581)
N	1,787	862	925
Direct Question (DQ1)	0.0536	0.0254	0.0566
N	914	437	477
p-value LE1 - DQ1	0.0049	0.0287	0.0702
	Panel B: All YYAs who worked in last 2 years		
	(4) total	(5) rural	(6) urban
List Experiment (LE1)	0.2071***	0.2412***	0.1669**
	(0.0571)	(0.0814)	(0.0795)
N	1,024	514	510
Direct Question (DQ1)	0.0837	0.0779	0.0891
N	502	244	258
p-value LE1 - DQ1	0.0430	0.0509	0.3610

Notes: Rural refers to the district Mangochi, urban to Blantyre. LE1 is the first list experiment, in which respondents report the total number of statements that are true for them, which for the treatment group includes the statement 'I have experienced forced labor in the last two years.' DQ1 shows the proportion of respondents in the control group who said yes to the question 'Have you experienced forced labor in the last two years?' Heteroskedasticity-robust standard errors in parentheses. Reported p-value for LE1 - DQ1 is from bootstrapping with 1000 repetitions. ***99%, **95%, *90%.

to respondents who earlier in the survey, i.e. before any questions on forced labor were asked, reported that they had worked at least at some point in the last two years. Technically, only individuals who worked in the last two years could have experienced labor trafficking during this time period, but in practice respondents may not count forced labor as work experience.

Column 1 of Panel A in Table 1 shows that among all Malawian young adults in the sample, the prevalence estimate of forced labor from the list experiment is 17.49 percent, an

Table 2: Comparison of LE1 and DQ1 (Zambia)

	Panel A: All YYAs		
	(1) total	(2) rural	(3) urban
List Experiment (LE1)	0.0882**	0.0254	0.1538***
	(0.0352)	(0.0520)	(0.0472)
N	2,971	1,497	1,480
Direct Question (DQ1)	0.1019	0.0717	0.1322
N	1,502	753	749
p-value LE1 - DQ1	0.7032	0.3734	0.6473
	Panel B: All YYAs who worked in last 2 years		
	(4) total	(5) rural	(6) urban
List Experiment (LE1)	0.2055***	0.1338*	0.2726***
	(0.0500)	(0.0766)	(0.0648)
N	1,494	727	767
Direct Question (DQ1)	0.1669	0.1317	0.2016
N	749	372	377
p-value LE1 - DQ1	0.4406	0.9783	0.2738

Notes: Rural refers to the districts of Chipata and Sesheke, urban to Lusaka and Livingstone. LE1 is the first list experiment, in which respondents report the total number of statements that are true for them, which for the treatment group includes the statement 'I have experienced forced labor in the last two years.' DQ1 shows the proportion of respondents in the control group who said yes to the question 'Have you experienced forced labor in the last two years?' Heteroskedasticity-robust standard errors in parentheses. Reported p-value for LE1 - DQ1 is from bootstrapping with 1000 repetitions. ***99%, **95%, *90%.

effect that is precisely estimated and more than three times larger than the prevalence of 5.36 percent reported by the control group in response to the direct question. As columns 2 and 3 show, these patterns are very similar in the rural district of Mangochi and the urban district of Blantyre, and we cannot reject the hypothesis that the estimated list experiment coefficients are the same. This suggests that the increased privacy of responses in the list experiment increases respondents willingness of reporting sensitive experiences of forced labor.

Panel B shows that the results are qualitatively similar in the sample of young adults who worked in the last two years. The overall list experiment estimate is 21 percent relative to a direct question prevalence of 8 percent. Again, we cannot reject the hypothesis that there is no heterogeneity between rural and urban areas. Overall, these estimates suggest that roughly every fifth 18-37 year old in Malawi is a survivor of forced labor when that information is self-reported in a privacy-improved setting.

Table 2 reports the results for the Zambian sample. The total Zambian sample is roughly twice as large as the Malawian sample since data was collected in four rather than two districts. But Zambian young adults are much less likely to report having worked in the last two years than young Malawians, leading to a bigger difference in the prevalence estimates between Panels A and B. For the total sample in Panel A, the list experiment estimate of 8.82 percent and the direct question prevalence of 10.19 percent are very similar to one another. Most of that effect comes from the urban areas, whereas responses in the rural area list experiment yield a labor trafficking estimate of 2.5 percent that is not statistically significantly different from zero.

In Panel B, which includes roughly half of the sample from Panel A, the results show an estimated list experiment prevalence of 21 percent, which is very similar to the corresponding rate in Malawi. As in Panel A, the reported likelihood of a forced labor experience is higher in the urban than in the rural areas of Zambia, but we cannot reject the hypothesis that the coefficients are the same, and list experiment and direct question estimates are similar to one another.

Relative to Table 1, the results in Table 2 therefore suggest that while the vulnerability of young adults to forced labor among the group with recent work experience is very similar, in contrast to Malawi social stigma, shame or fear of persecution do not appear to be first-order concerns in Zambia. Differences in the prevalence of labor trafficking in the overall samples across both countries arise from the lower labor-force participation rate in Zambia.

6.2 Second List Experiment and Direct Question: LE2 and DQ2

Table 3: Comparison of LE2 and DQ2 (Malawi)

	Panel A: All YYAs		
	(1) total	(2) rural	(3) urban
List Experiment (LE2)	0.1583***	0.1786***	0.1381**
	(0.0407)	(0.0598)	(0.0555)
N	1,783	862	921
Direct Question (DQ2)	0.0580	0.0482	0.0669
N	914	436	478
p-value LE2 - DQ2	0.0160	0.0322	0.1968
	Panel B: All YYAs who worked in last 2 years		
	(4) total	(5) rural	(6) urban
List Experiment (LE2)	0.2030***	0.2305***	0.1699**
	(0.0540)	(0.0787)	(0.0736)
N	1,021	514	507
Direct Question (DQ2)	0.0962	0.0658	0.125
N	499	243	256
p-value LE2 - DQ2	0.0609	0.0457	0.5565

Notes: Rural refers to the district Mangochi, urban to Blantyre. LE2 is the second list experiment, in which respondents report the total number of statements that are true for them, which for the treatment group includes the statement 'I have experienced forced labor in the last two years.' DQ2 shows the proportion of respondents in the control and treatment group who said yes to the question 'Have you experienced forced labor in the last two years?' Heteroskedasticity-robust standard errors in parentheses. Reported p-value for LE1 - DQ1 is from bootstrapping with 1000 repetitions. ***99%, **95%, *90%.

The results from both the first list experiment and the direct question rely on respondents' understanding of the term 'forced labor'. To test whether survey participants update their responses after they answer an intensive module on specific experiences that fall under labor trafficking we repeat the list experiment and direct questions after the module. The setup in the second list experiment is exactly the same as in the first. In Malawi, the direct question

Table 4: Comparison of LE2 and DQ2 (Zambia)

	Panel A: All YYAs		
	(1) total	(2) rural	(3) urban
List Experiment (LE2)	0.0699**	0.026	0.1137**
	(0.0354)	(0.0518)	(0.0484)
N	2,961	1,486	1,475
Direct Question (DQ2)	0.1076	0.0764	0.1392
N	1,506	759	747
p-value LE2 - DQ2	0.2869	0.3304	0.599
	Panel B: All YYAs who worked in last 2 years		
	(4) total	(5) rural	(6) urban
List Experiment (LE2)	0.1982***	0.2342***	0.1646**
	(0.0509)	(0.0762)	(0.0679)
N	1,486	724	762
Direct Question (DQ2)	0.1748	0.136	0.2132
N	755	375	380
p-value LE2 - DQ2	0.6449	0.198	0.4741

Notes: Rural refers to the districts of Chipata and Sesheke, urban to Lusaka and Livingstone. LE2 is the second list experiment, in which respondents report the total number of statements that are true for them, which for the treatment group includes the statement 'I have experienced forced labor in the last two years.' DQ2 shows the proportion of respondents in the control group who said yes to the question 'Have you experienced forced labor in the last two years?' Heteroskedasticity-robust standard errors in parentheses. Reported p-value for LE1 - DQ1 is from bootstrapping with 1000 repetitions. ***99%, **95%, *90%.

(DQ2) is asked to both the treatment and control group after the list experiment, whereas in Zambia only the control group was given the question. If the survey module raises awareness or improves recollection of relevant experiences, we would expect this to increase the forced labor prevalence estimates relative to the earlier estimates.

Tables 3 and 4 show the results of this second round of forced labor questions. As Table 3 reveals, the responses for the Malawian sample are quite similar to those from Table 1. In

particular, the list experiment estimate for the sample of individuals who have worked in the last two years is 20.30 percent in list experiment 2, compared to a list experiment 1 estimate of 20.71 percent. Rural workers slightly revise their responses downward by one percentage point, whereas the urban prevalence is almost unchanged.

In the direct responses, all workers are about 1 percentage point more likely to reveal that they are survivors, which is similar across rural and urban areas. Since the direct question was asked to participants in both the treatment and the control group, we can test whether both groups respond differently, but find no evidence that they do. Taken together, this suggests some mild updating but no large-scale changes due to raised awareness. The gap between list experiment and direct question prevalence rates also persists, supporting the idea that even after a long module on specific forms of labor exploitation the added privacy of the list experiment remains important.

Table 4 reports the analogous results for the Zambian sample. For individuals who have worked in the last two years, the direct responses are again slightly revised upwards by about a percentage point or less, similar to the response in the Malawian sample. The list experiment estimate is about the same (20.55 percent in LE1 compared to 19.82 percent in LE2), but this masks a flip in the response patterns in rural and urban areas. While in LE1 the prevalence was 13 percent in rural and 27 percent in urban areas, in LE2 they are 23 percent and 16 percent, respectively. In other words, rural participants substantially revise their responses upward while those in urban areas revise them downward by about the same magnitude, leaving the overall prevalence rate almost unchanged.[13]

In further analysis, we will explore potential explanations for this behavior.

6.3 Comparison to UN Labor Trafficking Thresholds

In this sub-section, we will compare the two rounds of prevalence estimates from list experiments and direct questions to the information from the detailed labor trafficking module.

[13]These effects are not driven by different response rates for the two list experiments. The patterns persist when we restrict the sample to individuals who answered both list experiments.

Using Okech et al. (2021), information on a range of specific forms of exploitation can be combined to see whether a given work experience crosses the threshold of labor trafficking as defined by the UN Palermo Protocol. The comparison of the implied prevalence rates to those of the other two approaches provides a test of whether the definition of forced labor that respondents have is likely similar or very different from the international definition.

The disaggregated module responses also have the advantage that we can construct indicators of different severity of labor trafficking, which provide more information on the types of acts and means of the labor trafficking definition that most young adults have experienced. Following the existing literature, we calculate the prevalence rates of labor trafficking based on two thresholds: PRIF1 and PRIF2. PRIF1 only requires one or more severe forms of trafficking experiences to cross the threshold. The PRIF2 threshold is satisfied with at least moderate experiences across at least two of the seven possible categories of experiences. We also create a joint PRIF variable that captures respondents that meet either the PRIF1 or the PRIF2 threshold, or both. This is the closest equivalent to the formal expressed truth from the conceptual framework.

Table 5 presents the prevalence rates in Malawi for the overall sample and for the list experiment treatment and control groups separately. Like previous tables, it estimates labor trafficking also separately by rural and urban districts. As the table shows, the prevalence of labor trafficking with the formal UN definition in PRIF Joint is higher than the rate from DQ1 and DQ2 from above for both the overall sample and those who have worked in the last two years, but is substantially smaller than the list experiment estimates. This suggests that social stigma plays a larger role for prevalence than the definition dimension.

6.4 List Experiment Assumptions

List experiments rely on two assumptions: First, that the way respondents answer the control statements is not influenced by the treatment statement (no design effects). And second, that respondents truthfully answer the treatment statement. We carry out Blair and Imai

Table 5: Labor Trafficking Module (Malawi)

	Panel A: All YYAs		
	(1) total	(2) rural	(3) urban
PRIF Joint	0.0750	0.0672	0.0822
PRIF1 (Severe)	0.0196	0.0209	0.0184
PRIF2 (Moderate)	0.0722	0.0638	0.0800
N	1,787	862	925
PRIF Joint (T only)	0.0816	0.0824	0.0809
PRIF1 (Severe) (T only)	0.0195	0.0188	0.0202
PRIF2 (Moderate) (T only)	0.0782	0.0776	0.0787
N	870	425	445
PRIF Joint (C only)	0.0687	0.0526	0.0833
PRIF1 (Severe) (C only)	0.0196	0.0229	0.0167
PRIF2 (Moderate) (C only)	0.0665	0.0503	0.0813
N	917	437	480
	Panel B: All YYAs who worked in last 2 years		
	(4) total	(5) rural	(6) urban
PRIF Joint	0.1240	0.1012	0.1471
PRIF1 (Severe)	0.0313	0.0311	0.0314
PRIF2 (Moderate)	0.1201	0.0973	0.1431
N	1,024	514	510
PRIF Joint (T only)	0.1308	0.1185	0.1440
PRIF1 (Severe) (T only)	0.0308	0.0259	0.0360
PRIF2 (Moderate) (T only)	0.1269	0.1148	0.1400
N	520	270	250
PRIF Joint (C only)	0.1171	0.0820	0.1500
PRIF1 (Severe) (C only)	0.0317	0.0369	0.0269
PRIF2 (Moderate) (C only)	0.1131	0.0779	0.1462
N	504	244	260

Notes: Rural refers to the district Mangochi, urban to Blantyre. PRIF1 captures severe, PRIF2 captures moderate experiences with labor trafficking. PRIF Joint includes everyone who meets the PRIF1 or PRIF2 thresholds.

(2012)'s test of the no design assumption: It estimates the proportion of the sample that is of each possible response type in the treatment and the control group, respectively, using generalized moment selection and the flexibility of a Monte Carlo simulation since responses to control statements might not be independent of each other. If all the estimated response type proportions are positive, or if it cannot be rejected that any negative estimates could have been randomly generated, the overall null hypothesis of no design effects cannot be rejected.

The p-values in Table 6 show that the null hypothesis of no design effect cannot be rejected in any of the four main samples: all YYAs and those that worked in the last two years in Malawi and Zambia, respectively. This implies that adding the forced labor statement does not significantly alter responses to the control items. With the exception of one estimated response type in the all YYA Zambia sample, all estimated proportions are positive, and the method cannot reject that the negative proportion occurred by chance. In results not shown here, we similarly find that all rural and urban sub-samples for both countries pass this test.

This analysis also suggests that ceiling and floor effects in which respondents alter their responses because they are aware that reporting the maximum or minimum possible numbers makes their responses to all of the individual statements known to the interviewer are not large concerns. We carefully chose our control statements to avoid floor and ceiling effects by choosing one control statement that was likely true for most respondents, one statement that was likely false for most respondents, and two other control statements that were likely true for some but not for others.

7 Conclusion

We explore the robustness of labor trafficking prevalence estimates across different measurement approaches and two countries, Malawi and Zambia. Our preliminary results suggest that providing additional privacy for reporting potentially sensitive information on labor

Table 6: Test of No Design Effects

Panel A: Malawi

	(1)	(2)	(3)	(4)
	All YYAs		Worked in last 2 years	
Control statements said yes to	π Control	π Treatment	π Control	π Treatment
0	0.1368	0.0093	0.125	0.0198
1	0.3343	0.091	0.3129	0.0979
2	0.2998	0.0546	0.2906	0.0685
3	0.05	0.0154	0.0584	0.017
4	0.0041	0.0046	0.0061	0.0038
Total	0.825	0.1749	0.793	0.207
p value no design effect	1		1	

Panel B: Zambia

	(1)	(2)	(3)	(4)
	All YYAs		Worked in last 2 years	
Control statements said yes to	π Control	π Treatment	π Control	π Treatment
0	0.1346	-0.0103	0.0993	0.0179
1	0.3245	0.0311	0.2583	0.0817
2	0.3446	0.042	0.3333	0.0699
3	0.1035	0.0175	0.0988	0.0224
4	0.0044	0.0082	0.0048	0.0136
Total	0.9116	0.0885	0.7945	0.2055
p value no design effect	0.4026		1	

Notes: Table reports the estimated proportions of all possible response types using Blair and Imai (2012). Columns (1) and (3) indicate the estimated response types for individuals saying no to the forced labor statement. Columns (2) and (4) indicate the estimated response types of individuals saying yes to the forced labor statements. Each column sums up to the estimated proportion of individuals saying no and yes to the forced labor statement in the first list experiment, respectively. The p-value for the null hypothesis of no design effect is Bonferroni-corrected for multiple hypothesis testing.

trafficking consistently matters in Malawi but is of much lower importance in Zambia. On the other hand, respondents in rural and urban areas of our Zambian sample revise their responses in opposite directions based on additional information provided through a detailed survey module on different specific forms of labor exploitation.

More broadly, we set up a conceptual framework and propose a survey design approach that operationalizes this framework. It clarifies the differences between the four types of

prevalence rates one may be interested in, and offers a practically feasible way for researchers, policymakers and stakeholders to generate richer data on the different concepts and their relative importance.

References

Bandiera, Oriana, Ahmed Elsayed, Andrea Smurra, and Céline Zipfel. 2022. "Young Adults and Labor Markets in Africa." *Journal of Economic Perspectives*, 36(1): 81–100.

Beam, Emily, David McKenzie, and Dean Yang. 2016. "Unilateral Facilitation Does Not Raise International Migration from the Philippines." *Economic Development and Cultural Change*, 64(2): 323–368.

Blair, Graeme, and Kosuke Imai. 2012. "Statistical Analysis of List Experiments." *Political Analysis*, 20: 47–77.

Chuang, Erica, Dupas Pascaline Huillery Elise, and Juliette Seban. 2021. "Sex, Lies, and Measurement: Consistency Tests for Indirect Response Survey Methods." *Journal of Development Economics*, 148: 102582.

Cullen, Claire. 2023. "Method Matters: The Underreporting of Initmate Partner Violence." *World Bank Economic Review*, 37(1): 49–73.

Goldberg, Jessica. 2016. "Kwacha Gonna Do? Experimental Evidence about Labor Supply in Rural Malawi." *American Economic Journal: Applied Economics*, 8(1): 129–149.

ILO. 2024. "Profits and Poverty: The Economics of Forced Labour. Second Edition." International Labour Organization, Geneva: International Labour Office.

ILO, Walk Free, and IOM. 2022. "Global Estimates of Modern Slavery: Forced Labour and Forced Marriage." International Labour Organization, Walk Free, and International Organization for Migration, Geneva.

Mo, Cecilia Hyunjung. 2018. "Perceived Relative Deprivation and Risk: An Aspiration-Based Model of Human Trafficking Vulnerability." *Political Behavior*, 40: 247–277.

Okech, D., L. Aletraris, N. Kagotho, and C. Risko. 2021. "Human Trafficking Core Indicators: Proposed Field Questions for Researchers." Center on Human Trafficking Outreach, University of Georgia.

Shrestha, Slesh, and Dean Yang. 2019. "Facilitating Worker Mobility: A Randomized Information Intervention among Migrant Workers in Singapore." *Economic Development and Cultural Change*, 68(1): 63–91.

Theoharides, Caroline. 2020. "The Unintended Consequences of Migration Policy on Origin-Country Labor Market Decisions." *Journal of Development Economics*, 142: 102271.

United Nations. 2000. "Protocol to Prevent, Suppress and Punish Trafficking in Persons Especially Women and Children, supplementing the United Nations Convention against Transnational Organized Crime."

U.S. Department of State. 2016. "Standard Terms and Conditions."