Quantifying Hidden Exploitation: Dual-Method Prevalence Estimates of Modern Slavery Risk Among UK Domestic Workers*

Caroline Emberson   Scott Moser

08, October 2025

Purpose

The purpose of this article is to demonstrate a quantitative approach to the construction of a risk index of labour exploitation and alternative estimators of the prevalence of exploitation.

Design/ Methodology/ Approach

Using data from a survey of domestic workers based in the United Kingdom (UK), we use statistical techniques, including Respondent Driven Sampling (RDS) methods RDS-I and RDS-II and Network Scale Up (NSUM) methods, to produce an index of labour exploitation risk and estimators of the prevalence of labour exploitation.

Findings

The labour exploitation risk index shows a reverse correlation between the increasing seriousness of exploitation and the number of exploitation cases reported. The various prevalence estimators examined show significant differences in population level exploitation.

Research implications/ limitations

Further research into the application of different quantitative statistical estimators of the prevalence of labour exploitation is urgently required.

Practical implications

Robust estimators are necessary if policy makers are to make informed choices about the appropriate allocation of scarce resources to help to eradicate severe forms of labour exploitation and labour abuse.

Social implications

---

*Authors' names are listed in alphabetical order.

Even by more conservative estimates, thousands of domestic workers in the UK are subject to labour exploitation. Urgent policy attention is needed if structural vulnerabilities are to be removed.

Originality

We believe this paper is the first to compare the use of RDS and NSUM methods in the quantitative estimation of the prevalence of labour exploitation and to construct a quantitative, composite index of labour exploitation risk.

Table of contents

1 Introduction

Labour exploitation has been defined as 'work situations that deviate significantly from standard working conditions as defined by legislation or other binding legal regulations, concerning in particular remuneration, working hours, leave entitlements, health and safety standards and decent treatment'(European Union for Fundamental Rights, 2015, p. 10). In the operations and supply chain management literature, interest in businesses' respect for these kinds of employee labour rights began with studies focused upon labour rights transgressions related to risk reduction and risk communication and how to improve employees' health and safety (Chinander, 2001; Wolf, 2001). More recently, serious labour rights abuses have come to the fore with studies examining the challenges of severe labour exploitation under the umbrella term 'modern slavery' (Benstead et al., 2018; Gold et al., 2015; New, 2015; Stevenson & Cole, 2018). While this literature offers important insights into severe forms of labour exploitation, particularly in global supply chains, this and the wider social sustainability literature has been criticised for its de-humanised approach to the understanding of workers and their working conditions (Soundararajan et al., 2021). While at least one current project seeks to examine the phenomenon of worker voice in factory settings (Leverhulme Trust, 2022), there appears to be little attention paid to severe forms of labour exploitation from the workers' perspective in the private sphere. Nowhere are the realities of individual workers' experiences of employer exploitation brought into sharper relief than in the setting of domestic work in private households.

The authors of the Global Slavery Index estimate that there are seventy-six million people employed in domestic work worldwide (International Labour Organization et al., 2022)). According to Bonnet et al. (2022), eighty percent of this domestic work is unregulated and informal. Labour exploitation has been identified as an extensive global problem within the sector, with domestic work identified as one of five private sector groupings which contribute the most to forced labour. Defined in the ILO Forced Labour Convention, 1930 No.29, forced or compulsory labour is 'all work or service which is exacted from a person under the threat of a penalty and for which the person has not offered himself or herself voluntarily' (I.L.O, 2024). Seventy-six percent of domestic workers are women, and these workers represent four percent of the total female workforce (International Labour Organization et al., 2022). Indeed, women in forced labour are much more likely to be in domestic work than in any other occupation (International Labour Organization et al., 2022). The ILO suggest that female domestic workers may be coerced through non-payment of wages; abuse of vulnerability; subjected to physical and sexual violence or experience threats against their family members. Such severe forms of labour exploitation may be present alongside other, perhaps less severe but equally illegal, practices which constitute various forms of labour abuse. The criminalisation of both labour exploitation and abuse in a domestic setting has developed in recent times, with legislation enacted in the United Kingdom (UK), Europe, Australia and Norway to criminalise such severe exploitation under the term 'modern slavery'. However, even where modern slavery laws are in place, reliance on traditional, inspection-led, approaches to detection designed primarily to ensure labour rights compliance within communal workplaces such as factories mean that the number of reported cases of labour exploitation in private dwellings may well severely underestimate actual exploitation levels. Though an exploration

of labour exploitation within private residences, our research seeks to redress the paucity of rigorous quantitative research in the modern slavery field. Specifically, this article aims to contribute to a more nuanced understanding of how quantitative methodologies may be deployed to improve understanding of the realities of workers' conditions by demonstrating the use of a statistically robust estimation of the nature and proportion of labour exploitation and abuse among domestic workers in the UK. This setting was chosen due to long-standing national legislation criminalising modern slavery introduced to the UK in 2015. Despite, or perhaps because of this legislation, in recent years the number of potential victims entering the UK's National Referral Mechanism (NRM), a scheme which provides government support for those suspected to be modern slavery survivors, has continued to increase. Nineteen thousand, one hundred and twenty-five potential victims were recorded in 2024: the highest annual figure since the NRM began (Home Office, 2025). In 2024, for the first time the number of cases of potential modern slavery among females handled by the charity Unseen, who run the UK's modern slavery helpline, were more prevalent than those among men (Carter, 2025). Despite these worrying headline statistics, and the persistence of specific concerns about high levels of exploitation among domestic workers in the grey literature (Kalayaan, 2008; Latin American Women's Rights Service, 2023; Mantouvalou, 2016), to our knowledge no-one has yet estimated the nature and extent of labour exploitation and abuse that may exist among domestic workers in the UK.

In contrast to overseas factory workers in globally dispersed, product, supply chains, many service workers engaged in domestic work have migrated to work in the UK. These transnational workers enter on restricted visas where their employment—and their right to remain in the country—is tied to their continuing employment. It is now ten years since the UK's Modern Slavery Act was enacted. During its passage through parliament, those advocating for the rights of domestic workers were successful in expanding the final category boundaries of the legislation to include, in Section 53, the specific definition of (overseas) domestic workers as modern slavery victims (Caruana et al., 2025). These transnational migrants are at particular risk of exploitation due to regulatory visa restrictions and intersecting structural issues related to their gender, the relative isolation of domestic work and a lack of supportive social networks. This can mean that they fall out of legal migratory status. Due to the social stigma attached to such illegal working, transnational workers remaining in the UK without the right to work may be considered a hidden, hard-to-reach, population. Extracting a sample of domestic workers which includes this group raises difficulties when trying to employ the normal statistical sampling methods considered necessary for robust prevalence estimation. Perhaps due to these sampling difficulties, we know relatively little about the nature of labour exploitation among this particularly at-risk group of workers. Fortunately, there has been significant interest in the development of alternative methods for prevalence estimation which include such hard-to-reach groups, with many scholars advocating and developing the use of respondent-driven sampling (RDS) techniques to support statistically robust estimators.

In this paper, we make two specific contributions to the operations and supply chain management literature. First, we demonstrate the use of RDS coupled with Network Scale-up Methods (N-SUM) to reach and sample respondents' views of their working conditions among these, predominantly female, transnational migrant domestic workers. We use the data we

obtain from these respondents to show how quantitative survey data can be used to estimate the proportion of workers experiencing labour exploitation. Second, we begin to capture the nature and extent of modern slavery as voiced by domestic workers, thus, we believe, expanding the nascent literature on worker voice which has, in the main, focused primarily on factory workers (Stephens et al., 2024). These contributions not only extend our understanding of the risks of labour exploitation and abuse among service workers engaged in domestic settings but also show how it is possible to shed light on the severity of the individuals' experience of exploitation through the construction of a novel risk index. The remainder of this paper is structured as follows. our study in more detail, highlighting what is already known about the current population of domestic workers in the UK and the conditions in which they work. Next, we describe our research methods. We review the development of the respondent-driven sampling (RDS) techniques we used and explain why this sampling method is suitable for our study. We then describe our survey methods, including how we designed our survey instrument, contacted our sample seeds and analysed our data. We then present and discuss our findings, detailing the proportional estimate that we calculated and the risk index we constructed. In our discussion, we expand upon the implications of our findings for government policy, enforcement practices and further research, including how these methods may be used in future studies of labour exploitation in other sectoral and geographic contexts. The limitations of our study are outlined, before, finally, we conclude our article.

1.1 Conceptualising Labour Exploitation and the Degree of Risk

- Binary vs. continuous definitions.

- Risk index construction and theoretical justification.

Modern slavery has been criticised by some for its overly extensive scope: encapsulating a broad range of divergent sub-categories of exploitation (O'Connell Davidson (2015); Gutierrez-Huerter O et al. (2023)). For this reason, we used the International Labour Organization's (I.L.O. (2012)) 'Indicators of Forced Labour' to identify the potential for severe labour exploitation and as a basis for the quantification of our labour exploitation and abuse risk index. The ILO identify eleven indicators designed to help understand how forced labour arises and how it affects victims. These indicators include: abuse of vulnerability; deception; restriction of movement; isolation; physical and sexual violence; intimidation and threats; retention of identity documents; withholding of wages; debt bondage; abusive working and living conditions and excessive overtime. According to the ILO, the presence of a single indicator in any given situation may in some cases imply the existence of forced labour. However, it also suggests that in other cases it may be necessary to look for several indications which, taken together, may point to a case of forced labour. We seek to refine this statement through the construction of a composite index by which means a degree of risk related to the likelihood of a domestic worker experiencing this most severe form of exploitation may be distinguished from the likely occurrence of less severe, though similarly illegal, forms of labour abuse.

2 Evaluating the Degree of Risk

The study of risk management has a long tradition in operations and supply chain management. Initially, the risks under consideration were primarily related to ensuring continuity of the supply of goods and services (see for example, Jüttner et al. (2003)). Beginning with Anderson (2006) and Anderson and Anderson (2009), however, a literature stream of sustainability-related supply chain risk management developed related specifically to the risks associated with the environment and social justice. A normative consensus related to the main stages of supply chain risk management has developed in the literature, with a five-stage sequential model typically presented. There have also been empirical studies of risk management within various industrial supply chains in the United States and India (Tarei et al. (2018); Dellana et al. (2021)), including the quantification of a risk index for the petroleum supply chain (Tarei et al. (2018)). Yet, while these authors recognize the need for responsible management and its effect on societal values, in line with other literature in the field they view risk from the perspective of the corporate supply chain rather than examining the risk of harm to the worker.

In our study, we conceptualise the risk of labour exploitation from the workers' perspective. We conceive severe forms of labour exploitation such as forced labour as one end of a spectrum ranging from illegal employment practices that constitute labour abuse, such as wage payments below legal minimum wage levels and health and safety violations, through to the likelihood of criminal exploitation recognized in the UK as modern slavery. Our assessment of this personal risk permits a degree of risk to be assigned to various clusters of forced labour indicators with the more indicators present, the stronger the likelihood that the working conditions may be considered exploitative. Our approach, therefore, includes, but goes beyond, assessing the likelihood of forced labour by simply quantifying the proportion of survivors entering the UK's National Referral Mechanism (NRM): a government system for survivor support set up to identify whether there are positive grounds for the identification of Modern Slavery. In our method, an NRM referral is used as the strongest indicator of modern slavery risk, with lesser risks assessed according to the degree to which cumulative indicators of forced labour are reported.

2.1 Case Setting: Labour Exploitation Risk Among Transnational Migrant Domestic Workers In The UK

Domestic work forms part of a broader industrial category of Personal and Household Service work (PHS). Work in this category includes those employed in 'social work activities without accommodation' and 'activities of households as employers of domestic personnel' (European Commission (2012)). In 2017, an estimated 980,000 people were engaged in PHS work in the UK (Manoudi et al. (2018)). Manoudi et al. (2018) highlight that the PHS sector is dominated by women and migrants, with many undeclared foreign workers. Detailed statistics related to the country of origin of domestic workers migrating to work in PHS in the UK are difficult to isolate before 2019. Since that time, annual migration has fluctuated – falling sharply in 2021 due in part to the COVID-19 pandemic, before later rising again

above pre-pandemic levels. In the year to December 2022, the UK Home Office reported that it had issued 18,533 Overseas Domestic Worker visas (Home Office (2023)). These domestic workers came from various countries in South America and Asia, including many from the Philippines.

In 2023, Strauss and Cocco (2023) reported a big shift in the source countries of migrants arriving in the UK on the Overseas Domestic Worker and other types of worker visas. Transnational domestic workers from the Philippines and India accounted for the single largest number of applications granted (10,186 and 3,858 visas respectively), followed by smaller, but still significant, numbers of workers arriving from Bangladesh (465), Nigeria (446), Sri Lanka (444), Egypt (422), and Ethiopia (285). In the same period, smaller numbers of visa applications to work as a domestic worker in the UK were also accepted from workers from other source countries including, but not limited to, the Sudan, Nepal, Ghana, Kenya, Lebanon, Eritrea, Iran, Turkey, Yemen, Malaysia, Thailand, and Morocco. This post-Brexit increase in the diversity of source countries from which transnational workers are drawn makes a more detailed analysis of the risk of labour exploitation in the sector both timely and more urgent.

There is a long history of reports of exploitation in the domestic work sector in the UK. In 2008, the civil society organisation Kalayaan, which was formed to campaign for the formal recognition of migrant domestic workers' rights in the UK, reported on the impact of proposed changes to the UK immigration system on migrant domestic workers (Kalayaan (2008)). Their report highlights government recognition of documented and unacceptable levels of abuse and exploitation among domestic workers in the UK as early as 1996. At this stage, new policies, including the development of a specialised visa allowing domestic workers to change employer during their stay were introduced. However, in 2012, these visa conditions were modified, tying domestic workers to a single employer and restricting the length of time that they are permitted to remain in the country to a period of six months (Gower (2016)). Overseas domestic worker visa holders are now, again, permitted to change employers, but not to apply to renew their six-month long visa unless they receive a positive 'Conclusive Grounds' decision related to exploitation considered to be modern slavery through the UK's National Referral Mechanism (NRM) (Romero and Fisher (2025)).

These reports highlight the underlying reasons for migrant domestic workers' vulnerability, including workers' relative desperation for work; their lack of social ties; unfamiliarity with English language and culture; long working hours; lack of knowledge of their legal rights; a lack of oversight of the private home as a workplace; their work forming part of the informal economy; their reliance on their employer for permission to work in the UK; and their lack of recourse to public funds. As a result, migrant domestic workers are vulnerable to abuse ranging from minor breaches of employment and health and safety law, to physical and sexual violence, slavery, forced labour and trafficking.

That these conditions may persist is evidenced by a report from another civil society organisation, the Latin American Women's Rights Service, which describes the results from twelve in-depth interviews with Latin American domestic workers in the UK. This report depicts high levels of isolation, exploitation and abuse including a failure by employers to

provide written contracts or payslips; breaches of verbal agreements; a requirement to perform different tasks from those indicated during recruitment; increasing working hours with little or no time off; excessive work days; a lack of paid holiday; many domestic workers not registered with a GP; sexual harassment in the workplace; verbal or physical abuse; employer surveillance; a lack of opportunity to change working conditions; isolation and fear of seeking help; and high reported levels of trafficking for labour exploitation (Latin American Women's Rights Service (2023)).

Against this backdrop, we used respondent driven sampling (RDS) as a sampling technique to recruit and survey domestic workers in the UK about the working conditions they were experiencing to estimate the nature and scale of abuse and exploitation based upon reports of their conditions by domestic workers themselves.

3 Research Methods

3.1 Respondent-Driven Sampling (RDS) And Survey Method

Comprehensive descriptions and literature reviews of the development and use of RDS to estimate the population size of a hidden population are available elsewhere (Heckathorn (2011); Gile et al. (2018)). Suffice it to say, the possibilities of the use of a one-wave snowball sampling to allow researchers to obtain a sample of personal networks was posited by Frank and Snijders (1994). Following the identification of a set of original sample members known as seeds, Heckathorn (1997); Heckathorn (2002) advocate the use of a double incentive to recompense participants not only for their involvement, but also for their recruitment of further participants in subsequent 'waves' of participation by drawing upon the social ties through which members of the hidden population are connected to each other.

The typical number of original sample seeds is between two and ten: chosen as heterogeneously as possible (Gile et al. (2018)). Though they may be subject to both systematic and non-systematic errors, the use of snowballing methods for the study of hidden populations, with the support of monetary or symbolic rewards, has been advocated as a way of creating robust recruitment embodying diversity in characteristics such as ethnicity, gender and geographical location (Heckathorn (1997); Heckathorn (2002)). In these papers, Heckathorn advances the development of RDS to include self-reported network size as a population estimator and bootstrapping techniques to support the development of an estimator's confidence intervals, an approach that has since been refined by others (Gile and Handcock (2015)). Such developments derive a new class of indicators for the population mean and define a corresponding bootstrap method to estimate the errors in RDS. The resulting 'network working model' permits the individual's connectedness in the network to be tested, while reducing bias with respect to the composition of the seeds. Snowball sampling is based upon the initial recruitment of the original sample selection by means of convenience. RDS also takes a non-random approach to seed selection, but relies upon the social network structure that exists between participants to produce a non-probabilistic sample (Goodman (2011)). Incentive structure is important—though this weakness is not a feature

of our target hidden population, some researchers have identified that younger men with higher socio-economic status are less likely to participate (McCreesh et al. (2013)). Perhaps of more concern, RDS has been described as a risky strategy since researchers cannot be sure whether enough respondents have been recruited through subsequent waves to eliminate bias within the original sample members (Vincent and Thompson (2017)).

RDS has been widely used to sample a variety of hidden populations, including HIV prevalence, rape and client-initiated gender-based violence among sex workers (McCreesh et al. (2012); Schwitters et al. (2012)). While the RDS method has proved limited when seeking to provide population heterogeneity by geographical location (McCreesh et al. (2011)), where these population features are of lesser importance, such methods have been used successfully. RDS methods have been used to survey other migrant populations (Tyldum (2021)), while such network-based referrals have been described as the only viable method to reach many types of labour trafficking victims (Zhang (2012)) and have been used to research exploitation among low-wage workers in three American cities (Bernhardt et al. (2009)); a study of labour trafficking in migrant communities in the city of San Diego (Vincent and Thompson (2017)); examination of the worst forms of child labour in the Indian state of Bihar (Zhang et al. (2019)); and the commercial sexual exploitation of children in Nepal (Jordan et al. (2020)).

The survey instrument included modules on demographic and employment characteristics, social network size and composition, and indicators of labour exploitation. The exploitation indicators were based on the International Labour Organization's framework of forced labour, adapted for the UK domestic work context. These indicators allowed us to operationalise exploitation in two ways. First, we constructed binary indicators classifying respondents as exploited or not exploited, based on threshold criteria. Second, we developed a continuous risk index, designed to capture gradations of vulnerability across the full sample.

In the following section, we describe our methods, including how we designed our survey, contacted our sample seeds, and analysed our data. Our approach can best be described as Web-based RDS (Wejnert and Heckathorn (2008)). We designed a web survey using the JISC online survey interface, suitable for our respondents to complete via a mobile phone. Composite measures to quantify the extent to which respondents were at risk of labour exploitation, including severe forms of exploitation such as forced labour, were constructed from existing exploitation typologies, notably the ILO's Indicators (I.L.O. (2012)). The survey consisted of these 11 composite indicators and included questions related to domestic workers' level of job satisfaction, employment conditions, and demographic data such as nationality, age, and gender. The main survey was conducted in the five months between February and July 2023.

3.2 Initial Sample Selection

We selected our first wave of participants nonrandomly by convenience sampling. Mobile phone numbers were used both to identify seed participants and to act as a unique identifier

for those whom they referred. To avoid sample homophily, original sample members were selected from three distinct domestic worker communities. This was facilitated by civil society organisations who represented distinct domestic worker communities. One was an exclusively online community of transnational domestic workers working in the UK, the second represented UK domestic workers of Filipino origin, and the third drew its membership from the Latin American community of domestic workers, also in the UK. Along with other academics with expertise in exploitation within domestic work, representatives from these three organisations also contributed to survey question design and facilitated the piloting of an initial version of the survey (which was translated and made available in four languages: English, Spanish, Tagalog, and Portuguese) to selected domestic workers within each community.

3.3 Survey Incentives: Incentive Design and Participation Verification

A double incentive scheme rewarded respondents both for completing the questionnaire and for each referral who went on to engage with the survey. The challenge of incentive design is to set the incentive at a level that adequately rewards respondents' time and participation, but that also avoids the risk of fraudulent participation due to too high a monetary gain (Jordan et al. (2020)). A sum of £10 was provided for survey completion with a further £5 for each successful nomination. While respondents were asked to nominate up to 10 domestic workers within their existing social network, it was the first three of these from whom participation was requested in subsequent waves. This approach is akin to the use of vouchers in face-to-face studies as advocated by Thompson (2020).

The ethical and practical issues related to the design and effective use of incentives for RDS among vulnerable populations has been much discussed in the literature; see, for example, Wang et al. (2005); Abdul-Quader et al. (2006); Singer and Bossarte (2006); DeJong et al. (2009); Semaan et al. (2009); Brunovskis and Surtees (2010); Semaan (2010); Platt et al. (2015), including the specificities of incentive use within web-based surveys (Cobanoglu and Cobanoglu (2003)). Following the principles of lottery use established by Brown et al. (2006) and Laguilles et al. (2011), we also designed our survey to encourage the maximum extent of participation by entering all respondents completing the questionnaire into a free prize draw for £150. Research suggests that a high lottery provides the most cost-effective incentive for obtaining complete responses (Gajic et al. (2012)). While using incentives to encourage participation would seem to be desirable, it is worth noting the potential downside of respondents fabricating responses to increase their remuneration (Robinson (2014)). To minimise this risk, mobile phone numbers for each respondent and those whom they referred were collated, and each of these numbers was called by one of the authors of the paper to ascertain the veracity of the respondent as a migrant domestic worker.

3.4 Descriptive Statistics

In total, we received completed online surveys from 97 respondents. Of these respondents, 90 identified themselves as transnational migrants. Forty-five percent regarded themselves

as self-employed, 39% identified themselves as employees, and 16% categorised their employment status as that of a worker.

Of the 97 respondents, 64 (66% of the total), and the largest single nationality group, reported that they had a Filipina background. Other nationalities represented included Dominican, Brazilian, Spanish, Colombian, Bolivian, Venezuelan, Cuban, and Panamanian. Female domestic workers made up 97% of the sample, with 3% of the sample comprised of male domestic workers. The age structure of the domestic workers was skewed towards those over 45 years old, with such workers representing over half of the sample (see Table 1).

Recruitment diagnostics indicate that equilibrium was reached across key demographic variables by wave X. Reported personal network sizes ranged from X to X, with a mean of X and a standard deviation of X. Figure 1 presents the recruitment tree, showing that Latinx respondents generated longer referral chains, while British respondents tended to form shorter, more fragmented networks.
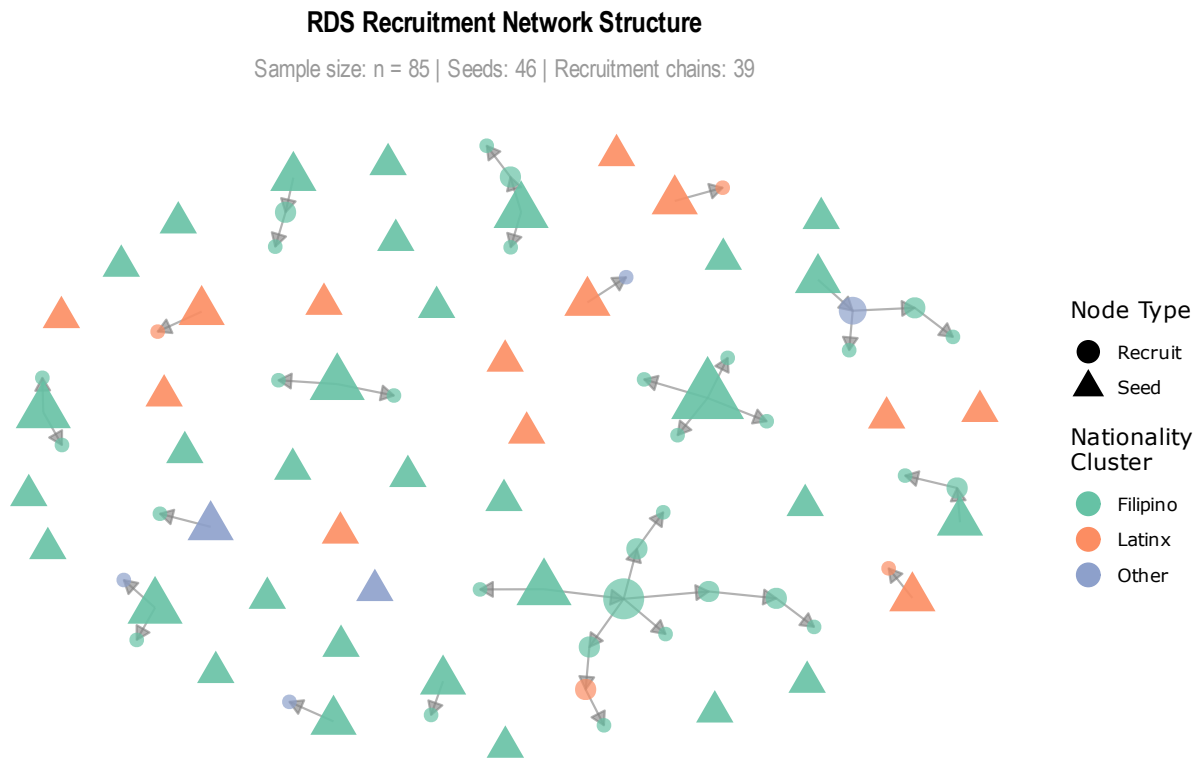

Figure 1: RDS Network Structure: Recruitment chains showing seed nodes (triangles) and recruit nodes (circles). Node colors represent nationality clusters, with node size proportional to out-degree (number of recruits). Lines show recruiter-recruit relationships. Source: Authors' Own Work.

Table 1: RDS Sample Characteristics by Nationality Cluster and Recruitment Wave. Source: Authors' Own Work.

Table 2: Sample Characteristics by Nationality Cluster

Nationality	N	%	Seeds	Recruits	Mean Degree	Median Degree
Filipino	62	72.9	31	31	11.3	5
Latinx	17	20.0	13	4	6.9	5
Other	6	7.1	2	4	7.5	4

Note: Degree refers to reported network size (Q13: number of domestic workers known).

3.5 Indicators

creating comparable variables to bridge the Respondent-Driven Sampling (RDS) and Network Scale-Up Method (NSUM) estimations is a key methodological challenge the research team is actively working to solve. The core issue is that the two methods rely on differently framed questions: • RDS estimation uses "egocentric" questions, which focus on the respondent's personal experiences (e.g., "Have you been forced to work?"). • NSUM estimation uses questions about the respondent's knowledge of others in their network (e.g., "How many domestic workers do you know who have experienced…"). The survey questions designed for these two purposes do not always match up directly, making a fair comparison between the estimates difficult. The Strategy: Creating "Linking" Variables To overcome this, the team's strategy is to identify and group questions that cover the same underlying themes, even if they are worded differently for each method. The goal is to aggregate some of the specific "egocentric" questions to make them comparable to the broader "how many do you know" NSUM questions. Based on your sources, several sets of "linking" questions have been identified with varying degrees of confidence in their comparability: • Document Withholding (High Confidence): ◦ RDS Question: Q70 asks if the employer has withheld the respondent's travel and identity documents. ◦ NSUM Question: Q71 asks how many other domestic workers the respondent knows who do not have access to their own documents. • Debt and Pay Issues (High Confidence): ◦ RDS Questions: This involves combining Q39 (having to pay off debt to someone who helped find work) and Q42 (has your pay ever been withheld?). ◦ NSUM Question: Q43 asks how many others the respondent knows who have experienced problems with debt or pay. • Abuse and Threats (Medium Confidence): ◦ RDS Questions: This requires grouping Q45 (forced, deceived, or threatened into poor working conditions), Q47 (employer intimidation or threats), and Q48 (verbal abuse). ◦ NSUM Question: Q49 asks how many others the respondent knows who have experienced the use of threat or force. • Excessive Hours (Lower Confidence): ◦ RDS Questions: Q61 (weekly rest does not last 24 hours) and Q62 (working excessive overtime) are combined. ◦ NSUM Question: Q64 asks about knowing others who work excessive overtime, lack breaks, or lack annual leave. The comparison here is considered "flakier" because the RDS questions do not ask about annual leave, creating a partial mismatch. • Access to Help (Lower Confidence): ◦ RDS Question: Q78 asks if the respondent knows who might help them (coded for "No" answers). ◦ NSUM Question: Q79 asks how many other domestic workers the respondent knows who do NOT know where to go for help. By creating these comparable variables, even

if just for one or two highly confident themes like debt bondage or document withholding, the team hopes to perform a fair comparison between the estimates derived from the two different methodologies. This process is considered a crucial step to integrate the RDS and NSUM approaches and to produce more robust findings.

3.6 Ethical Approval

The data collection that underpins the analysis presented in this paper was given favourable ethical approval by the lead author's School Research Ethics Committee in January 2023. All participants were informed about the aims of the study, provided informed consent, and were assured that participation was voluntary and confidential.

3.7 Use of Artificial Intelligence in Research

Large Language Models (LLMs) were used for brainstorming the organisation of the paper and editing of text. Code co-pilot (Claude Code) was used to test and debug R scripts employed in the statistical analysis. No generative models were used to generate or simulate data.

4 Estimation Methods

We analysed the survey sample using multiple estimation models in order to assess robustness and conduct sensitivity analyses (see Appendix XX). The survey instrument contained both ego questions (which capture information about respondents and their personal network ties) and alter questions (which capture information about the people respondents know). These two types of network data enable two fundamentally different estimation strategies: respondent-driven sampling (RDS) estimators and network scale-up methods (NSUM). In addition, we developed and implemented a novel three-step bootstrap procedure to address uncertainty in NSUM estimates, which we describe in more detail below.

4.1 RDS-Based Estimation

RDS estimators use ego-based information. Each participant reported the number of other domestic workers they knew, and this degree information was used to adjust for the over-representation of highly connected individuals in the sample. We implemented RDS-II and Gile's successive sampling (SS) estimator, the latter of which accounts for finite population effects and improves performance when the sample fraction is relatively large (Gile, 2011; Gile and Handcock, 2010). These estimators were used to generate prevalence estimates for binary indicators of exploitation.

For continuous traits, such as the exploitation risk index, we applied model-assisted inference approaches (**gile15-network<empty citation>**, Gile, 2011; Gile, Beaudry, and Handcock, 2018). These approaches combine design-based adjustments with regression models that incorporate auxiliary covariates, producing valid estimates of sample means and distributions of continuous outcomes under the RDS design.

4.2 Estimation Strategy

Given that our data were collected using a respondent-driven sampling (RDS) design, the choice of estimator is critical. We considered several well-established RDS estimators, each with distinct assumptions and applicability.

First, the **RDS-I estimator salg04-samplin<empty citation>** provides an early design-based approach. However, it performs poorly in small samples, is highly sensitive to seed dependence, and requires strong assumptions about equilibrium. Since our sample comprises fewer than 100 respondents recruited from multiple seeds, we regard RDS-I primarily as a historical benchmark rather than a viable option for inference.

Second, the **Volz–Heckathorn (VH) estimator volz08-probabi<empty citation>** offers a probability-based refinement of RDS-I. It is more robust but assumes a large population relative to the sample size and does not adjust for finite-population effects. Although our diagnostic checks indicated approximate equilibrium across key demographics, the VH estimator is best used here as a robustness check rather than the primary estimator.

Third, the **Successive Sampling (RDS-SS) estimator** (also called RDS-II) **gile11-improve<empty citation>** accounts for finite population effects, adjusting for the non-negligible sampling fraction that arises when the target population is not extremely large relative to the sample. This property makes RDS-SS particularly appropriate for our study of migrant domestic workers, and we therefore use it as our primary estimator for binary outcomes (e.g., presence or absence of exploitation).

Fourth, to estimate **continuous outcomes** such as our exploitation risk index, we employ the **Model-Assisted (MA-RDS) estimator gile15-network<empty citation>**. This approach integrates regression modelling with RDS weights, thereby extending inference beyond binary outcomes and improving efficiency by leveraging auxiliary covariates.

Finally, although recent work on **Clustered Successive Sampling Population Size Estimation (Clustered SS-PSE) gamb23-clustered<empty citation>** extends RDS-based methods to estimating the size of clustered hidden populations, our study does not attempt to estimate the absolute number of domestic workers. Instead, we focus on the prevalence and severity of exploitation. As such, Clustered SS-PSE is not central to our analysis, though it remains a promising avenue for future research.

In summary, we rely primarily on **RDS-SS** for binary outcomes and MA-RDS for continuous outcomes, with **VH** used for robustness checks. This combined strategy balances methodological rigor with the substantive goals of our study.

We use model-assisted RDS estimators (**gile15-network<empty citation>**) because they are design-based yet leverage a working ERGM (with degree and homophily terms) to approximate inclusion probabilities conditional on our observed seeds. This approach directly addresses seed bias and homophily that conventional RDS estimators cannot correct, accommodates finite-population effects through successive-sampling logic, and supports valid estimation of both binary and continuous outcomes. By conditioning on the actual seed composition and subgroup structure in our data, the method reduces bias, improves efficiency, and provides design-compatible bootstrap uncertainty, making it particularly well-suited to our small, heterogeneous sample of domestic workers."

4.3 Network Scale-Up Methods (NSUM)

NSUM relies on alter-based information. Rather than depending on respondents' own position in the referral network and their reported degree, NSUM uses information about alters—other people in respondents' networks. Participants reported on the number and characteristics of people they knew who met specific exploitation criteria. These reports were aggregated to estimate prevalence in the wider population of domestic workers.

The key methodological distinction between RDS and NSUM lies in how the social network is used. RDS leverages ego-level network size and recruitment paths to adjust for biases in the referral process. NSUM treats respondents as informants about a larger social universe, using alter data to infer prevalence. RDS depends on accurate self-reporting of degree and on the properties of recruitment chains, while NSUM depends on the accuracy of respondents' knowledge about others and the representativeness of their social networks

4.3.1 How RDS and NSUM Use Social Network Information

Respondent-driven sampling (RDS) and network scale-up methods (NSUM) both rely on social network structures, but they exploit different aspects of those structures for inference.

RDS uses **ego-based information**. Each participant reports the size of their personal network of eligible individuals, and these degree reports are combined with the wave at which respondents were recruited to adjust for unequal inclusion probabilities. The underlying logic is that individuals with larger networks are more likely to be recruited earlier and more often, creating a bias toward highly connected respondents. RDS estimators, including Gile's successive sampling estimator, explicitly correct for this bias by weighting observations according to network degree and recruitment path. For continuous traits, such as our risk index, RDS model-assisted estimators incorporate auxiliary covariates into this weighting process to further reduce bias.

In contrast, NSUM uses **alter-based information**. Rather than focusing on the ego's probability of inclusion, NSUM treats respondents as informants about the wider hidden population. Respondents are asked how many people they know with a given trait (for example, "How many domestic workers do you know who have experienced exploitation?").

These responses are then scaled up, using assumptions about network size and visibility, to infer prevalence in the broader population. This method does not depend on recruitment chains but instead on the accuracy of respondents' knowledge about others in their social circles.

The practical difference is therefore twofold. RDS estimates are anchored in "who recruited whom" and "how many do you know," while NSUM estimates are anchored in "how many of your alters fit this category." RDS leverages inclusion probabilities tied to ego network size; NSUM leverages alter reports to extend beyond the sample. Applying both methods to the same dataset allows for triangulation across two fundamentally different inferential logics.

4.4 Bootstrap Procedure for NSUM

To appropriately characterise uncertainty in NSUM estimates, we developed a novel three-step bootstrap procedure. This approach resamples respondents, their reported alters, and the exploitation classifications simultaneously, thereby capturing uncertainty at each stage of the inference process. This procedure provides more realistic confidence intervals than those generated by conventional variance estimators, particularly for small samples such as ours. Details of the bootstrap implementation and diagnostic checks are provided in Appendix **app-3step<empty citation>**.

4.5 Comparative Rationale

Respondent-driven sampling (RDS) and network scale-up methods (NSUM) both rely on social network structures, but they exploit different aspects of those structures for inference.

RDS uses ego-based information. Each participant reports the size of their personal network of eligible individuals, and these degree reports are combined with the wave at which respondents were recruited to adjust for unequal inclusion probabilities. The underlying logic is that individuals with larger networks are more likely to be recruited earlier and more often, creating a bias toward highly connected respondents. RDS estimators, including Gile's successive sampling estimator, explicitly correct for this bias by weighting observations according to network degree and recruitment path. For continuous traits, such as our risk index, RDS model-assisted estimators incorporate auxiliary covariates into this weighting process to further reduce bias.

In contrast, NSUM uses alter-based information. Rather than focusing on the ego's probability of inclusion, NSUM treats respondents as informants about the wider hidden population. Respondents are asked how many people they know with a given trait (for example, "How many domestic workers do you know who have experienced exploitation?"). These responses are then scaled up, using assumptions about network size and visibility, to infer prevalence in the broader population. This method does not depend on recruitment chains but instead on the accuracy of respondents' knowledge about others in their social networks.

The practical difference is therefore twofold. RDS estimates are anchored in "who recruited whom" and "how many do you know," while NSUM estimates are anchored in "how many of your alters fit this category." RDS leverages inclusion probabilities tied to ego network size; NSUM leverages alter reports to extend beyond the sample. Applying both methods to the same dataset allows for triangulation across two fundamentally different inferential logics.

Applying both RDS and NSUM to the same survey allows triangulation across two fundamentally different inferential paradigms. For continuous outcomes, such as the exploitation risk index, only model-assisted RDS estimators are appropriate. For binary outcomes, both RDS and NSUM can be applied, enabling direct comparison of results. This dual approach strengthens the empirical credibility of our findings, highlights the conceptual value of considering exploitation both as a binary threshold and as a continuum, and demonstrates the methodological trade-offs involved in studying hidden populations.

5 Results

5.1 Model-Assisted Estimates of the Risk Index (Continuous Conceptualisation)

One of the central contributions of this study is the introduction of a continuous risk index to measure degrees of exposure to labour exploitation. The index was constructed from multiple indicators aligned with the International Labour Organization's forced labour framework, weighted inductively and refined through expert consultation. Rather than treating exploitation as a dichotomy, the risk index conceptualises all domestic workers as facing some degree of potential exploitation, albeit with significant variation in intensity.

Because continuous traits cannot be estimated directly with conventional RDS estimators or NSUM, we employed model-assisted inference methods. These methods adjust for the non-random structure of RDS recruitment while permitting reliable estimation of means and distributions of continuous outcomes (Gile, 2011; Gile and Handcock, 2010; Gile, Beaudry, and Handcock, 2018).

Composite Risk Index: Sample vs. Population-Adjusted Estimates
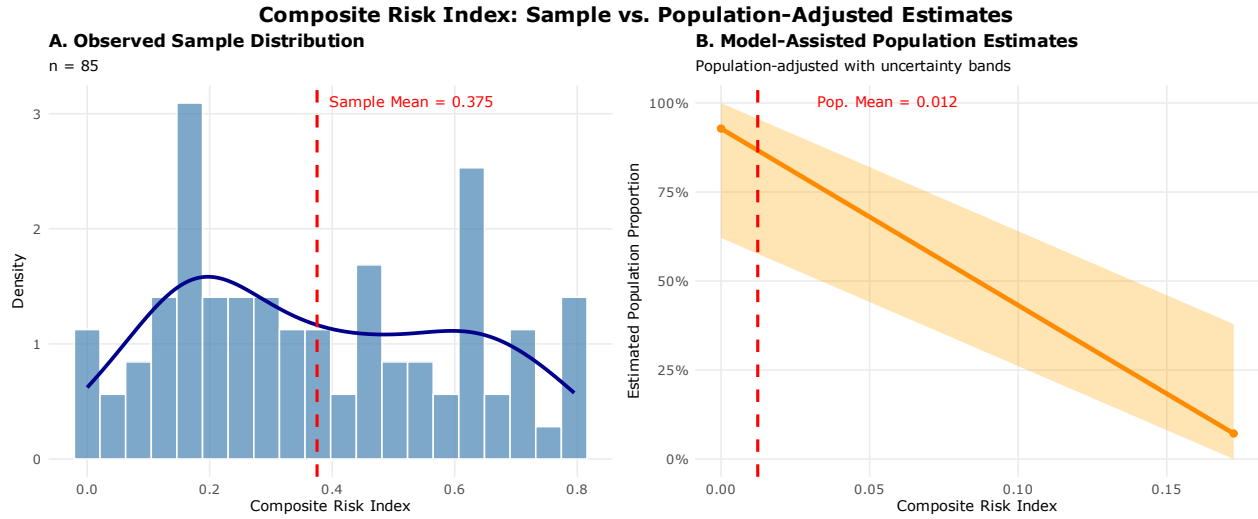
Figure 2: Model-Assisted Estimates of Composite Risk Index: Distribution comparison between observed sample data and population-adjusted estimates. The left panel shows the empirical distribution of risk scores in the RDS sample, while the right panel presents the model-assisted population proportion estimates for specific risk categories with uncertainty bounds (the light band shows confidence intervals around the proportion estimates). Source: Authors' Own Work.

Figure 2 presents the comparison between the observed sample distribution and the model-assisted population estimates, demonstrating how the bias correction affects our understanding of exploitation risk in the broader domestic worker population. The population-weighted average risk is 0.0124 This represents the Model-Assisted estimate of mean exploitation risk in the population, adjusted for RDS sampling bias. It is a design-based estimate that corrects for the non-random recruitment process in RDS. The population-adjusted distribution shows less concentration in the lower risk values compared to the raw sample. The MA estimate suggests that 7.2% of DWs in the UK have moderate risk (0.1725) This pattern reinforces the conceptual claim that exploitation is best understood as a continuum rather than a simple binary condition.

5.2 Binary Exploitation Indicators (Exploited or Not Exploited)

To complement the continuous measure, we also operationalised exploitation as a binary outcome. Respondents were classified as exploited if they met threshold indicators consistent with ILO definitions. This allows estimation using both respondent-driven sampling estimators, which rely on ego-based network data, and network scale-up methods, which rely on alter-based information.

Table 3: Model-Assisted Estimates of Binary Exploitation Indicators: Population prevalence estimates adjusted for RDS sampling bias. Values represent the estimated proportion of domestic workers experiencing each form of exploitation. Source: Authors' Own Work.

Table 4: Model-Assisted Estimates of Binary Exploitation Indicators

Exploitation Indicator	Population Prevalence (95% CrI)
Any Exploitation	95.0% (62.2-100.0%)
Excessive Hours	83.7% (71.9-95.6%)
Limited Access to Help	58.3% (46.1-70.6%)
Threats/Abuse	57.3% (35.4-79.3%)
Pay-Related Issues	49.5% (35.6-63.5%)
Document Withholding	34.3% (16.4-52.2%)

Note: Estimates adjusted for RDS sampling bias using Bayesian Model Assissted Inference of Gile & Handcock (2015) (N=980,000, sample seed selection). Credible intervals from posterior distribution.

5.2.1 Model-Assisted (MA) Estimates
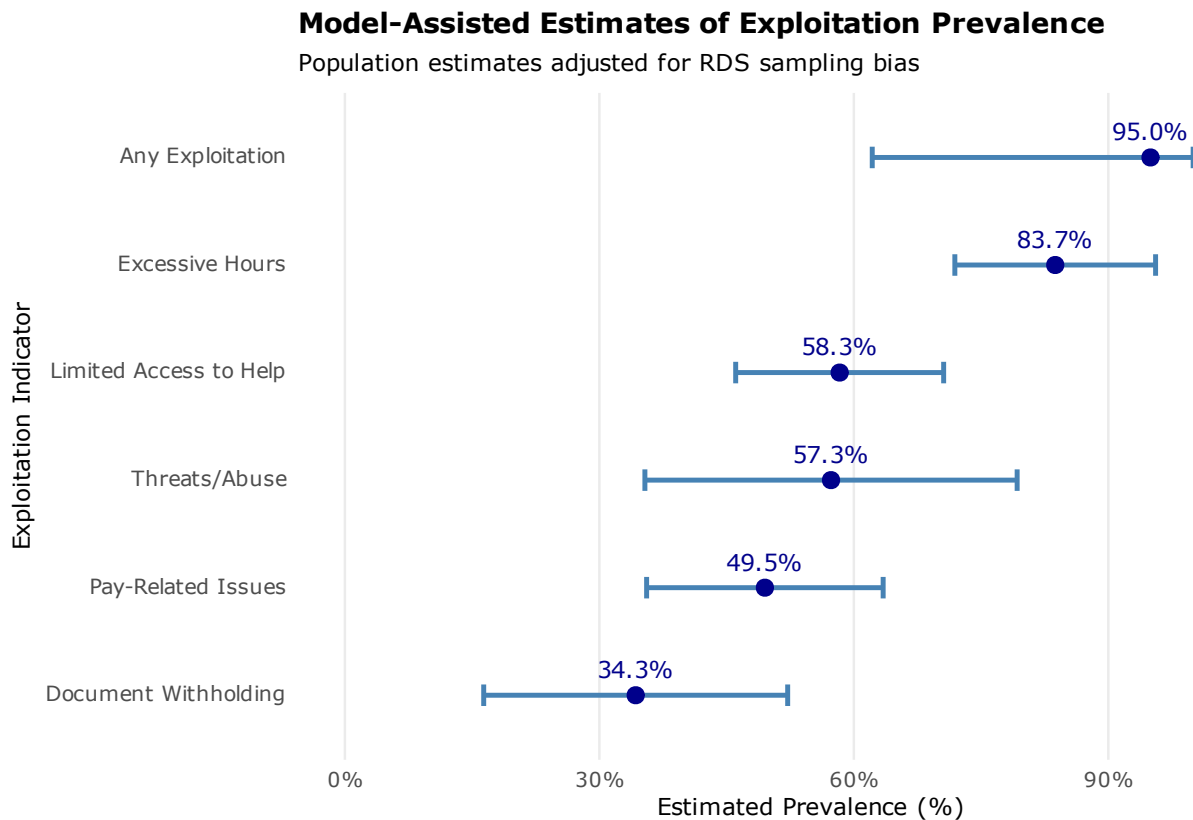
Figure 3: Forest Plot of Model-Assisted Binary Exploitation Estimates: Population prevalence estimates with 95% credible intervals for each exploitation indicator, ordered by prevalence magnitude. Error bars represent posterior uncertainty from Bayesian SS-PSE (N=980,000, sample seed selection). Source: Authors' Own Work.

Table 5: RDS Estimates of Binary Exploitation Indicators: Comparison of RDS-I and RDS-SS estimators with neighborhood bootstrap confidence intervals (N=980,000). Source: Authors' Own Work.

Exploitation Indicator	RDS-I	RDS-SS
Document withholding	49.4% (25.3–56.3)	34.2% (25.2–56.0)
Excessive working hours	92.8% (71.4–92.9)	83.7% (71.5–93.0)
Limited access to help	57.7% (55.1–80.3)	59.4% (54.7–80.2)
Pay-related issues	59.3% (34.2–63.5)	49.7% (34.8–63.4)
Threats and abuse	65.2% (40.5–72.1)	57.2% (40.1–72.3)

Note: RDS-I: Salganik-Heckathorn estimator. RDS-SS: Gile's Successive Sampling estimator. Confidence intervals from neighborhood bootstrap (n=1000). Population size: 980,000 domestic workers.

5.2.2 RDS Estimates
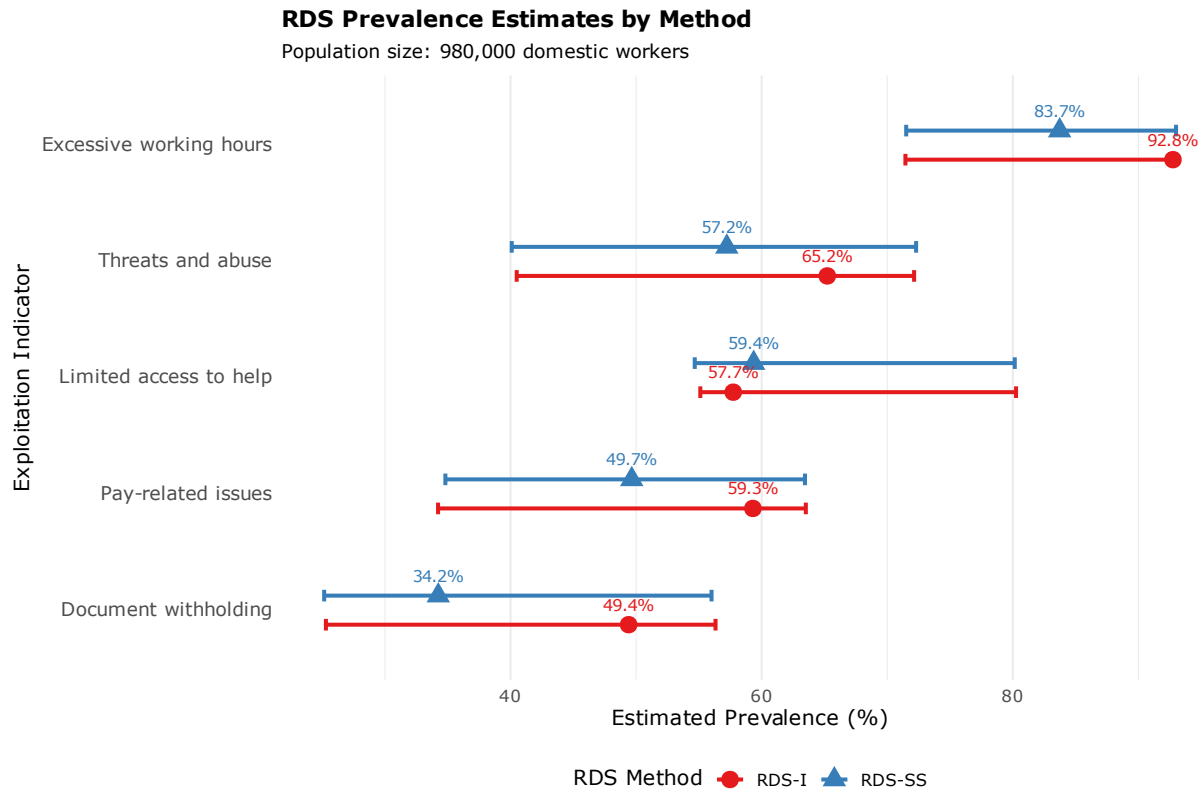
Figure 4: Forest Plot of RDS Estimates: Comparison of RDS-I and RDS-SS prevalence estimates for binary exploitation indicators (N=980,000). Error bars represent 95% bootstrap confidence intervals. Source: Authors' Own Work.

5.2.3 NSUM Estimates

5.3 Comparative Interpretation

Table 1 summarises prevalence estimates across both methods, highlighting points of convergence and divergence. While absolute values differ slightly across RDS and NSUM (reflecting differences in ego- versus alter-based assumptions), the overall pattern is remarkably stable.

Table 6: NSUM Estimates of Binary Exploitation Indicators: Population prevalence estimates using Network Scale-Up Methods with bootstrap confidence intervals. Estimates represent population size of exploited domestic workers. Source: Authors' Own Work.

Table 7: NSUM Estimates of Binary Exploitation Indicators

Exploitation Indicator	NSUM Estimate (95% CI)
Excessive Hours	64.2% (NA-NA%)
Threats/Abuse	59.3% (NA-NA%)
Pay-Related Issues	58.2% (NA-NA%)
Limited Access to Help	45.2% (NA-NA%)
Document Withholding	42.1% (NA-NA%)

Note: Network Scale-Up Method prevalence estimates with bootstrap confidence intervals from enhanced NSUM analysis.

Figure 5 demonstrates the comparison between frequentist RDS estimators (RDS-I, RDS-II, RDS-SS) and Bayesian model-assisted methods across all exploitation indicators. The horizontal spacing of estimates allows clear visual comparison of point estimates and uncertainty intervals, with RDS methods showing bootstrap confidence intervals and Bayesian methods showing credible intervals from MCMC estimation. Both approaches identify similar patterns of exploitation prevalence, with excessive hours showing the highest prevalence across all methods, followed by threats/abuse and pay-related issues.

Method Comparison: RDS vs Bayesian Estimators
Baseline population (980K) with 95% confidence/credible intervals

Note: RDS methods show 95% confidence intervals (bootstrap); Bayesian methods show 95% credible intervals (MCMC)
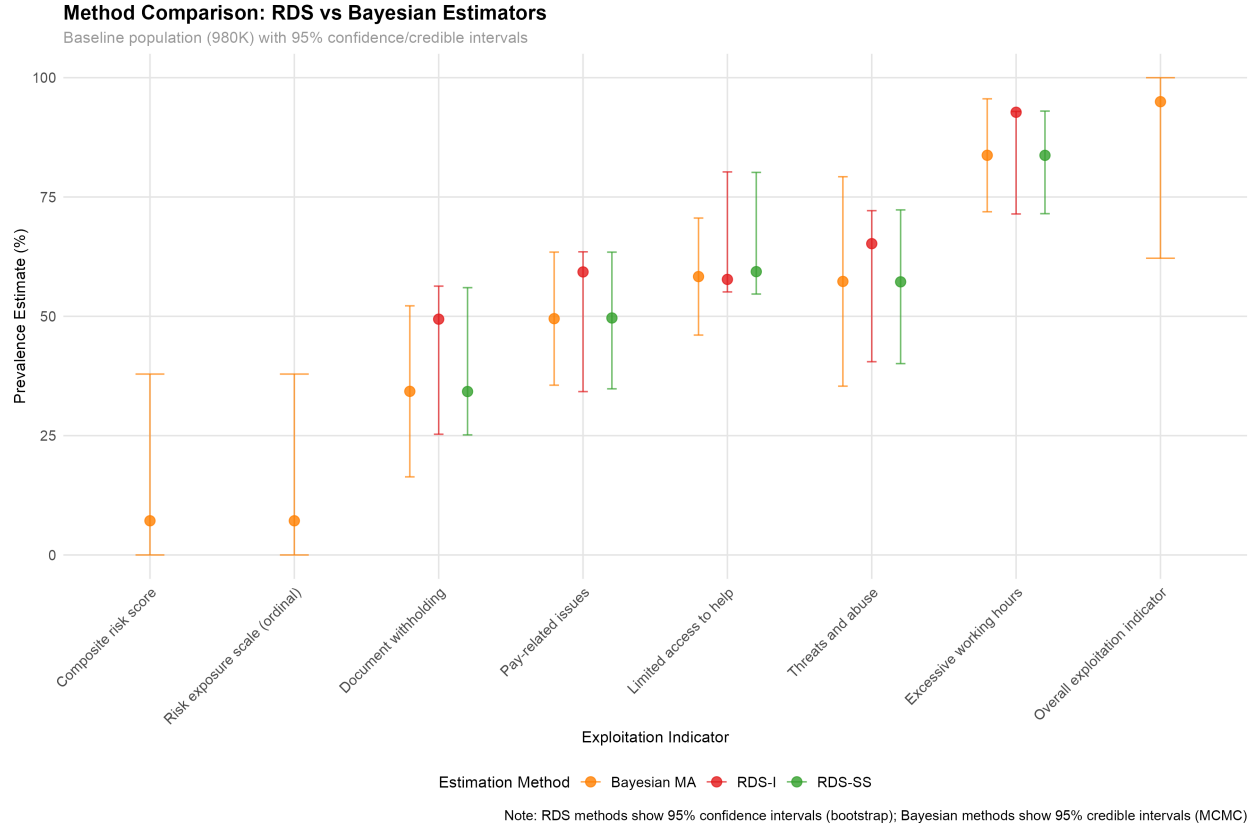
Figure 5: Method Comparison: frequentists RDS vs Bayesian RDS Estimators. Point estimates and uncertainty intervals for prevalence of exploitation indicators across different estimation methods. RDS methods show 95% confidence intervals (bootstrap); Bayesian methods show 95% credible intervals (MCMC). Baseline population assumption: 980,000 domestic workers. Source: Authors' own work.

5.4 Robustness Checks

A series of robustness checks were performed to assess the stability of the findings. **?@fig-forest-plot** presents a comprehensive forest plot comparing all prevalence estimates across indicators and methods, demonstrating remarkable consistency across different analytical approaches. Bootstrap resampling confirmed that the RDS and NSUM estimates remained stable across repeated draws. Sensitivity analyses excluding suspicious datapoints did not materially alter subgroup rankings. Analyses restricted to single subgroups confirmed that the elevated prevalence among Latinx workers was not an artifact of recruitment dynamics.

The forest plot visualization clearly shows that traditional RDS estimators (RDS-I and RDS-SS) produced results consistent with the model-assisted and Bayesian estimates, albeit with wider confidence intervals in some cases. The pattern of estimates is remarkably stable across methods, with excessive hours consistently showing the highest prevalence, followed by threats/abuse and pay-related issues, while document withholding shows more moderate prevalence levels. This cross-method validation strengthens confidence in our findings and

demonstrates the robustness of the dual estimation approach. Full technical details are reported in Appendices B and C.
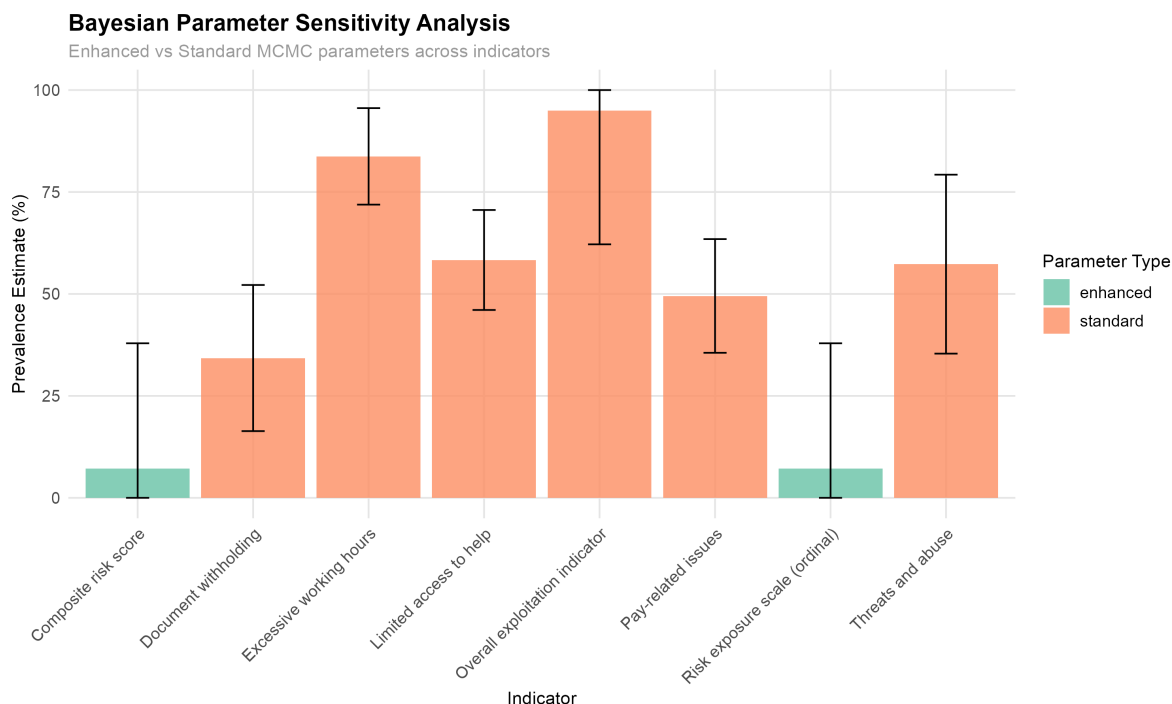
Figure 6: Bayesian Parameter Sensitivity Analysis. Comparison of prevalence estimates using enhanced vs standard MCMC parameters for Bayesian model-assisted estimation. Enhanced parameters use longer burn-in periods and more iterations, particularly important for numeric and ordinal variables like composite risk scores. Error bars represent 95% Bayesian credible intervals. Source: Authors' own work.

6 Discussion

6.1 Implications for Policy

The UK Government has proved reluctant to respond to calls to remove the restrictive, tied, visa conditions currently in force for those migrant workers working in the UK on the Overseas Domestic Workers visa (Gower (2016)). Maintaining these restrictive conditions prevents the ratification in the UK of C189, the International Convention for Domestic Workers (I.L.O. (2012)). If the estimates resulting from our study are correct, these visa conditions place migrant domestic workers at significant risk of serious forms of labour exploitation including, in its most severe form, exploitation that exhibits the characteristics of forced labour—legally considered a form of modern slavery.

To reduce the vulnerability of transnational domestic workers to this—and other—forms of labour exploitation, we urge policy-makers to reconsider these discriminatory visa conditions

and offer the same freedoms to domestic workers that are enjoyed by other groups of workers under UK law.

In addition, given the vulnerabilities experienced by workers due to the private nature of the workplace, we would urge the UK government to consider the regulation of domestic worker employers.

Finally, given the stigma and very real danger of deportation of those migrant domestic workers who may have fallen out of legal migration status, our evidence suggests that there is an urgent need for the UK Government to enforce a firewall between immigration control and labour exploitation if the true scale of abuse is to be made visible and the perpetrators brought to justice.

6.2 Implications for Practice

The UK Visa and Immigration service already offers rights-based training to migrant domestic workers via UK embassies in certain source countries. To reduce migrant domestic workers vulnerabilities, we advocate the expansion of this training both to include explicit training related to employment and labour rights within the UK and to the rapidly expanding range of new source countries from where migrant domestic workers are now drawn.

6.3 Methodological Contributions

This study makes several methodological contributions to the estimation of prevalence in hidden and hard-to-reach populations:

Dual conceptualisation of exploitation: We introduce two distinct approaches to operationalising exploitation. First, we treat exploitation as a binary outcome (exploited versus not exploited), enabling direct comparison of prevalence estimates across RDS and NSUM methods. Second, we construct a continuous risk index, acknowledging that all domestic workers may be exposed to some degree of exploitation risk. To our knowledge, this is the first application of model-assisted RDS estimators to quantify a continuous measure of exploitation risk.

Bayesian parameter sensitivity: Figure 6 demonstrates the importance of appropriate MCMC parameter selection for Bayesian model-assisted estimation. Our analysis shows that enhanced parameters (featuring longer burn-in periods and more iterations) are particularly crucial for numeric and ordinal variables, where standard parameters may suffer from convergence issues. This methodological innovation ensures robust credible interval estimation for continuous risk measures.

Combining RDS and NSUM on the same survey instrument: By designing a survey that captures both ego and alter information, we are able to apply RDS and NSUM to the same sample. This dual approach has rarely been implemented in studies of labour exploitation.

It provides an opportunity to cross-validate results and assess the robustness of prevalence estimates.

Novel bootstrap procedure for NSUM: Recognising the non-random nature of an RDS sample, we developed a three-step bootstrap procedure tailored for NSUM estimation. This resamples respondents, recalculates weights, and re-estimates NSUM prevalence at each iteration. The procedure captures multiple layers of uncertainty and produces more reliable confidence intervals than conventional methods, particularly in small samples.

Application to domestic workers in the UK: Finally, by applying these methods to a population that is both highly stigmatised and under-researched, we demonstrate the feasibility of using advanced network-based estimation techniques in contexts where traditional sampling is impossible. This methodological innovation has potential applications in studies of other hidden labour markets and vulnerable populations.

6.4 Limitations of the Study

As with any empirical research, our study is subject to limitations. In terms of nationality, our sample is not representative of the demographics of those domestic workers employed on Overseas Domestic Worker Visas in 2022 the UK. Due to the increasing number of workers on Overseas Domestic Work visas from the Indian sub-continent, attempts were made also to seed respondents from this community. This proved difficult, with anecdotal information suggesting that domestic workers from this community rarely had access to a personal mobile phone. It is not therefore possible to infer the nature and extent of labour exploitation within this sub-section of the domestic worker population.

As the network structure of our sample demonstrates, even with a well-designed incentive scheme it proved difficult to recruit respondents from these communities of domestic workers in subsequent sampling waves in the time available. Most of our respondents are therefore original sample members draw from the three domestic worker communities used to seed the survey.

6.5 Further Research

We believe that web-RDS combined with statistical estimators such as NSUM offers an important method for the capture and comparison of relative proportions of labour exploitation and abuse in sectors within and beyond the UK. Network scale up methods, and potential enhancements such as Generalised network scale up estimators offer to enhance understanding, not least within operations and supply chain management research, of the extent of labour exploitation in different sectors and across industries.

7 Conclusion

8 References

Abdul-Quader, A., Heckathorn, D. D., McKnight, C., Bramson, H., Nemeth, C., Sabin, K., Gallagher, K., & Jarlais, D. (2006). Effectiveness of respondent-driven sampling for recruiting drug users in new york city: Findings from a pilot study [Number: 3]. *Journal of Urban Health*, *83*(3), 459–76.

Anderson, D. R. (2006). The critical importance of sustainability risk management [Publisher: Risk & Insurance Management Society, Inc.]. *Risk Management (00355593)*, *53*(4), 66–74. Retrieved July 4, 2025, from https://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=20781527&site=ehost-live

Anderson, D. R., & Anderson, K. E. (2009). Sustainability risk management [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6296.2009.01152.x]. *Risk Management and Insurance Review*, *12*(1), 25–38. https://doi.org/10.1111/j.1540-6296.2009.01152.x

Benstead, A. V., Hendry, L. C., & Stevenson, M. (2018). Horizontal collaboration in response to modern slavery legislation: An action research project [Publisher: Emerald Publishing Limited]. *International Journal of Operations & Production Management*, *38*(12), 2286–2312. https://doi.org/10.1108/IJOPM-10-2017-0611

Bernhardt, A., Milkman, R., & N. (2009). *Broken laws, unprotected workers: Violations of employment and labor laws in america's cities* (UCLA IRLE Report.).

Bonnet, F., Carre, F., & Vanek, J. (2022). *Domestic workers in the world: A statistical profile* [WIEGO]. Retrieved January 2, 2025, from https://www.wiego.org/research-library-publications/domestic-workers-world-statistical-profile/

Brown, J., Schonfeld, T., & Gordon, B. (2006). You may have already won...": An examination of the use of lottery payments in research [Number: 1]. *IRB: Ethics & Human Research*, *28*(1), 12–16.

Brunovskis, A., & Surtees, R. (2010). Untold stories: Biases and selection effects in research with victims of trafficking for sexual exploitation [Number: 4]. *International Migration*, *48*(4), 1–37.

Carter, J. (2025, April 30). *Women and girls at growing risk of modern slavery, data reveals* [Unseen]. Retrieved June 26, 2025, from https://www.unseenuk.org/latest-modern-slavery-helpline-data/

Caruana, R., Crane, A., & Ingram, C. (2025). The boundaries of modern slavery: The role of exemplars in new category formation [Publisher: Academy of Management]. *Academy of Management Journal*, *68*(3), 507–539. https://doi.org/10.5465/amj.2022.0210

Chinander, K. R. (2001). Aligning accountability and awaress for environmental performance in operations [Publisher: SAGE Publications]. *Production and Operations Management*, *10*(3), 276–291. https://doi.org/10.1111/j.1937-5956.2001.tb00375.x

Cobanoglu, C., & Cobanoglu, N. (2003). The effect of incentives in web surveys: Application and ethical considerations [Number: 4]. *International Journal of Market Research*, *45*(4), 1–13.

DeJong, J., Mahfoud, Z., Khoury, D., Barbir, F., & Afifi, R. (2009). Ethical considerations in HIV/AIDS biobehavioral surveys that use respondent-driven sampling: Illustrations from lebanon [Number: 9]. *American Journal of Public Health*, *99*(9), 1562–1567.

Dellana, S., Rowe, W. J., & Liao, Y. (2021). A scale for measuring organizational risk management maturity in the supply chain [Publisher: Emerald Publishing Limited]. *Benchmarking: An International Journal*, *29*(3), 905–930. https://doi.org/10.1108/BIJ-11-2020-0578

European Commission. (2012). Staff working docuemnt on exploitating the employment potential of the personal and household services, SWD (2012) final.

European Union for Fundamental Rights. (2015). Severe labour exploitation: Workers moving within or into the european union [Institution: Koninklijke Brill NV]. https://doi.org/10.1163/2210-7975_HRD-9992-2016018

Frank, O., & Snijders, T. (1994). Estimating the size of hidden populations using snowball sampling. *Journal of Official Statistics-Stockholm*, *10*, 53–53.

Gajic, A., Cameron, D., & Hurley, J. (2012). The cost-effectiveness of cash versus lottery incentives for a web-based, stated-preference community survey [Number: 6]. *The European Journal of Health Economics*, *13*(6), 789–799.

Gile, K. J., Beaudry, I. S., Handcock, M. S., & Ott, M. Q. (2018). Methods for inference from respondent-driven sampling data [Publisher: Annual Reviews]. *Annual Review of Statistics and Its Application*, *5*, 65–93. https://doi.org/10.1146/annurev-statistics-031017-100704

Gile, K. J., & Handcock, M. S. (2015). Network model-assisted inference from respondent-driven sampling data [Number: 3 tex.ids= gile15-networka _eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/rssa.12091]. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *178*(3), 619–639. https://doi.org/10.1111/rssa.12091

Gold, S., Trautrims, A., & Trodd, Z. (2015). Modern slavery challenges to supply chain management [Publisher: Emerald Group Publishing Limited]. *Supply Chain Management: An International Journal*, *20*(5), 485–494. https://doi.org/10.1108/SCM-02-2015-0046

Goodman, L. (2011). Comment: On respondent-driven sampling and snowball sampling in hard-to-reach populations and snowball sampling not in hard-to-reach populations [Number: 1]. *Sociological Methodology*, *41*(1), 347–353.

Gower, M. (2016). Calls to change overseas domestic worker visa conditions. https://researchbriefings.files.parliament.uk/documents/SN04786/SN04786.pdf

Gutierrez-Huerter O, G., Gold, S., & Trautrims, A. (2023). Change in rhetoric but not in action? framing of the ethical issue of modern slavery in a UK sector at high risk of labor exploitation. *Journal of Business Ethics*, *182*(1), 35–58. https://doi.org/10.1007/s10551-021-05013-w

Heckathorn, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations [Publisher: [Oxford University Press, Society for the Study of Social Problems]]. *Social Problems*, *44*(2), 174–199. https://doi.org/10.2307/3096941

Heckathorn, D. D. (2002). Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations [Number: 1]. *Social Problems*, *49*(1), 11–34.

Heckathorn, D. D. (2011). Comment: Snowball versus respondent-driven sampling [Number: 1]. *Sociological Methodology*, *41*(1), 355–366.

Home Office. (2023). Why do people come to the UK to work? https://www.gov.uk/govern ment/statistics/immigration-system-statistics-year-ending-june-2023/why-do-peopl e-come-to-the-uk-to-work

Home Office. (2025). *Modern slavery: National referral mechanism and duty to notify statistics UK, end of year summary 2024* [GOV.UK]. Retrieved June 26, 2025, from https: //www.gov.uk/government/statistics/modern-slavery-nrm-and-dtn-statistics-end-of -year-summary-2024/modern-slavery-national-referral-mechanism-and-duty-to-notif y-statistics-uk-end-of-year-summary-2024

I.L.O. (2012). ILO indicators of forced labour. https://www.ilo.org/global/topics/forced-la bour/publications/WCMS_203832/lang--en/index.htm

I.L.O. (2024, January 28). *What is forced labour?* Retrieved June 30, 2025, from https://w ww.ilo.org/topics/forced-labour-modern-slavery-and-trafficking-persons/what-force d-labour

International Labour Organization, Walk Free, & Institute of Migration. (2022). *Global estimates of modern slavery forced labour and forced marriage.*

Jordan, L., Chui, C., Larmar, S., & O'Leary, P. (2020). Overcoming methodological challenges in prevalence studies in developing contexts with vulnerable children [Number: 3]. *International Social Work, 63*(3), 371–385.

Jüttner, U., Peck, H., & Christopher, M. (2003). Supply chain risk management: Outlining an agenda for future research [Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/13675560310001627016]. *International Journal of Logistics Research and Applications, 6*(4), 197–210. https://doi.org/10.1080/136755603100016 27016

Kalayaan. (2008). The new bonded labour? http://www.kalayaan.org.uk/documents/Kala yaan%20Oxfam%20report.pdf

Laguilles, J., Williams, E., & Saunders, D. (2011). Can lottery incentives boost web survey response rates? findings from four experiments [Number: 5]. *Research in Higher Education, 52*(5), 537–553.

Latin American Women's Rights Service. (2023). Behind closed doors: Experiences of latin american domestic workers in the UK. https://lawrs.org.uk/wp-content/uploads/20 23/08/Behind-closed-doors_domestic_work.pdf

Leverhulme Trust. (2022). *Research project grants* [Grant listings]. https://www.leverhulme .ac.uk/listings?field_grant_scheme_target_id=15

Maltiel, R., Raftery, A. E., McCormick, T. H., & Baraff, A. J. (2015). Estimating population size using the network scale up method. *The Annals of Applied Statistics, 9*(3), 1247–1277. https://doi.org/10.1214/15-AOAS827

Manoudi, A., Weber, T., Scott, D., & Hawley Woodall, J. (2018). *An analysis of personal and household services to support work life balance for working parents and carers - google search* (Synthesis Report ECE Thematic Review). European Commission. Retrieved October 28, 2024, from https://www.google.com/search?q=An+analysis+of+Perso nal+and+Household+Services+to+support+work+life+balance+for+working+par ents+and+carers&rlz=1C1GCEB_enGB932GB935&oq=An+analysis+of+Persona l+and+Household+Services+to+support+work+life+balance+for+working+paren ts+and+carers&gs_lcrp=EgZjaHJvbWUyBggAEEUYOdIBCDIzNjNqMGo3qAIAs AIA&sourceid=chrome&ie=UTF-8

Mantouvalou, V. (2016). Modern slavery? the UK visa system and the exploitation of migrant domestic workers. In *British politics and policy at LSE*. https://blogs.lse.ac.uk/politicsandpolicy/exploitation-of-migrant-domestic-workers-in-the-uk/.

McCreesh, N., Copas, A., Seeley, J., Johnston, L., Sonnenberg, P., & Hayes, R. (2013). Respondent driven sampling: Determinants of recruitment and a method to improve point estimation [Number: 10]. *PLoS ONE*, *8*(10).

McCreesh, N., Johnston, L., & Copas, A. (2011). Evaluation of the role of location and distance in recruitment in respondent-driven sampling [Number: 56]. *International Journal of Health Geography*, *10*(56).

McCreesh, N., Frost, S. D. W., Seeley, J., Katongole, J., Tarsh, M. N., Ndunguse, R., Jichi, F., Lunel, N. L., Maher, D., Johnston, L. G., Sonnenberg, P., Copas, A. J., Hayes, R. J., & White, R. G. (2012). Evaluation of respondent-driven sampling [Number: 1 tex.ids= mccr12-evaluationa publisher: Lippincott Williams & Wilkins]. *Epidemiology*, *23*(1), 138–147. Retrieved April 25, 2023, from https://www.jstor.org/stable/23214188

New, S. J. (2015). Modern slavery and the supply chain: The limits of corporate social responsibility? [Num Pages: 697-707 Place: Bradford, United Kingdom Publisher: Emerald Group Publishing Limited]. *Supply Chain Management*, *20*(6), 697–707. https://doi.org/10.1108/SCM-06-2015-0201

O'Connell Davidson, J. (2015). *The margins of freedom: Modern slavery.* Palgrave.

Platt, L., Luthra, R., & Frere-Smith, T. (2015). Adapting chain referral methods to sample new migrants: Possibilities and limitations. *Demographic Research*, *33*, 665–700.

Robinson, O. (2014). Sampling in interview-based qualitative research: A theoretical and practical guide [Number: 1]. *Qualitative research in psychology*, *11*(1), 25–41.

Romero, A., & Fisher, O. (2025). Blueprint for safer and fairer migration for low-paid work. Retrieved July 11, 2025, from https://labourexploitation.org/app/uploads/2025/06/Flex-Report-Blueprint-for-safer-and-fairer-migration-for-low-paid-work_final_30_June-.pdf

Schwitters, A., Swaminathan, M., & Serwadda, D. (2012). Prevalence of rape and client-initiated gender-based violence among female sex workers: Kampala, uganda [Number: Supply Nr 1]. *AIDS Behaviour*, *19*, 68–76.

Semaan, S. (2010). Time-space sampling and respondent-driven sampling with hard-to-reach populations [Number: 2]. *Methodological Innovations Online*, *5*(2), 60–75.

Semaan, S., Santibanez, S., Garfein, R., Heckathorn, D., & Jarlais, D. (2009). Ethical and regulatory considerations in HIV prevention studies employing respondent-driven sampling [Number: 1]. *International Journal of Drug Policy*, *20*(1), 14–27.

Singer, E., & Bossarte, R. (2006). Incentives for survey participation: When are they "coercive"? [Number: 5]. *American journal of preventive medicine*, *31*(5), 411–418.

Soundararajan, V., Wilhelm, M. M., & Crane, A. (2021). Humanizing research on working conditions in supply chains: Building a path to decent work [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jscm.12260]. *Journal of Supply Chain Management*, *57*(2), 3–13. https://doi.org/10.1111/jscm.12260

Stephens, V., Benstead, A. V., Goworek, H., Charles, E., & Lukic, D. (2024). Theorising worker voice for supply chain justice – communication, representation and recognition [Publisher: Emerald Publishing Limited]. *International Journal of Operations &*

Production Management, 45(3), 653–676. https://doi.org/10.1108/IJOPM-06-2023-0528

Stevenson, M., & Cole, R. (2018). Modern slavery in supply chains: A secondary data analysis of detection, remediation and disclosure. *Supply Chain Management: An International Journal, 23.* https://doi.org/10.1108/SCM-11-2017-0382

Strauss, D., & Cocco, F. (2023). Britain after brexit: The surprising surge in skilled migrants.

Tarei, P. K., Thakkar, J. J., & Nag, B. (2018). A hybrid approach for quantifying supply chain risk and prioritizing the risk drivers: A case of indian petroleum supply chain [Publisher: Emerald Publishing Limited]. *Journal of Manufacturing Technology Management, 29*(3), 533–569. https://doi.org/10.1108/JMTM-10-2017-0218

Thompson, S. (2020). New estimates for network sampling [tex.arxiv: 2002.01350].

Tyldum, G. (2021). Surveying migrant populations with respondent-driven sampling. experiences from surveys of east-west migration in europe [Number: 3]. *International Journal of Social Research Methodology, 24*(3), 341–35.

Vincent, K., & Thompson, S. (2017). Estimating population size with link-tracing sampling [Number: 519 tex.ids= vinc17-estimatinga _eprint: https://doi.org/10.1080/01621459.2016.1212712 publisher: Taylor & Francis]. *Journal of the American Statistical Association, 112*(519), 1286–1295. https://doi.org/10.1080/01621459.2016.1212712

Wang, J., Carlson, R., Falck, R., Siegal, H., Rahman, A., & Li, L. (2005). Respondent-driven sampling to recruit MDMA users: A methodological assessment [Number: 2]. *Drug and alcohol dependence, 78*(2), 147–157.

Wejnert, C., & Heckathorn, D. D. (2008). Web-based network sampling: Efficiency and efficacy of respondent-driven sampling for online research [Number: 1]. *Sociological Methods & Research, 37*(1), 105–134.

Wolf, F. G. (2001). Operationalizing and testing normal accident theory in petrochemical plants and refineries [Publisher: SAGE Publications]. *Production and Operations Management, 10*(3), 292–305. https://doi.org/10.1111/j.1937-5956.2001.tb00376.x

Yauck, M., & Moodie, E. E. M. (2022, May 31). *Neighboot: Neighborhood Bootstrap Method for RDS* (Version 1.0.1).

Zhang, S. (2012). Measuring labour trafficking: A research note. *Crime, Law and Social Change, 58*, 469–482.

Zhang, S., Dank, M., Vincent, K., Narayanan, P., Bharadwaj, S., & Balasubramaniam, S. (2019). Victims without a voice: Measuring worst forms of child labor in the indian state of bihar [Number: 7]. *Victims & Offenders, 14*(7), 832–858.

A Appendix A

Population Parameters

Use ~980,000 as UK domestic worker population estimate (EU data)

Use 44,360 as NRM adult referrals baseline Address treatment of "don't know" responses and zero network size claims Consider separate analysis for Filipino subgroup

A.1 The Survey

(From codebook)

Compare corresponding survey questions between RDS and NSUM methods:

- Q70/Q71 (document withholding)
- Q39+Q42/Q43 (pay issues)
- Q45+Q47+Q48/Q49 (abuse/threats)
- Q61+Q62/Q64 (excessive hours)
- Q78/Q79 (access to help)

Risk Index Implementation

- Clean coding for 13 risk categories with proper weightings
- NRM referral (0.35)
- Forced labor indicators (0.55 total)
- Below minimum wage (0.10)

A.2 Data Processing

The process of coding variables to make them comparable across the Respondent-Driven Sampling (RDS) and Network Scale-Up Method (NSUM) estimations is a multi-step procedure that moves from conceptual mapping to specific data transformations. The core challenge is bridging the "egocentric" questions (about the respondent's own experiences, used for RDS) with the network-based questions (about how many others the respondent knows with certain experiences, used for NSUM). Yes, the sources provide extensive detail on the coding of variables to make them comparable for RDS and NSUM estimation. The fundamental challenge is to create a fair comparison between the "egocentric" RDS questions (about personal experience) and the broader "how many do you know" NSUM questions. This was accomplished by first identifying thematic links between questions and then implementing a specific coding scheme to transform them into a consistent format. Conceptual Mapping of Variables The team first identified sets of questions that cover the same underlying themes, even if their wording differed. These "linking" questions were categorized by the team's confidence in their comparability. Thematic Groups for Comparison: • Document Withholding (Most Confident): Comparing personal experience (Q70) with knowing others

who lack access to their documents (Q71). • Pay/Debt Issues (High Confidence): Linking personal experience with debt (Q39) and withheld pay (Q42) to knowing others with such problems (Q43). • Threats/Abuse (High Confidence): Aggregating personal experiences of being forced, threatened, intimidated, or verbally abused (Q45, Q47, Q48) and comparing them to knowing others who experienced threats or force (Q49). • Excessive Hours (Lower Confidence): Combining questions about inadequate weekly rest (Q61) and excessive overtime (Q62) to compare with knowing others who face similar labor rights issues (Q64). The confidence is lower because the NSUM question (Q64) also asks about annual leave, which is not covered in the corresponding RDS questions. • Access to Help (Least Confident): Matching whether the respondent personally knows where to go for help (Q78) with knowing others who do not (Q79). Coding and Transformation Process The conceptual mapping was then put into practice through a detailed coding process, which is explicitly documented in the project's R script. The primary goal was to convert all variables into a binary (0/1) format, representing "No" or "Yes" to experiencing or knowing someone who experienced the issue.

Key Coding Steps:

1. Handling Ambiguous Answers: For most RDS questions, responses like "I don't know" or "Prefer not to say" were recoded as NA (missing) to exclude them from the binary analysis.
2. Converting Categorical RDS Variables to Binary: Questions with multiple response options (e.g., "Always," "Often," "Sometimes," "Never") were recoded into a 0/1 format. For instance, for Q70 (document withholding), the answers "Always," "Often," and "Sometimes" were all coded as 1 (indicating the presence of the issue), while "Never" was coded as 0.
3. Aggregating Multiple RDS Questions: For themes covered by multiple egocentric questions (like "Threats/Abuse"), each individual question was first recoded into a binary 0/1 variable. These binary variables were then combined using a logical OR (|). This means if a respondent answered "Yes" to any of the grouped questions (e.g., experiencing threats OR intimidation OR verbal abuse), the final composite RDS variable would be coded as 1. This aggregation was necessary to match the broader scope of the single corresponding NSUM question.
4. Converting NSUM Count Variables to Binary: The NSUM questions, which asked for a numerical count (e.g., "how many do you know?"), were also converted to binary. If a respondent knew one or more people (> 0), their answer was coded as 1; if they knew zero, it remained 0. Through this systematic process of recoding, handling missing data, and aggregating variables, the team created a set of parallel binary indicators suitable for a direct comparison between the RDS and NSUM estimation frameworks.

A.3 Sample Characteristics

A.3.1 Recruitment Network Structure

```r
library(tidyverse)
# Calculate chain statistics from recruitment network
calculate_chain_stats <- function(data) {
  # Build parent-child relationships
  chains <- data %>%
    dplyr::select(id, recruiter.id) %>%
    dplyr::mutate(
      id = as.integer(id),
      recruiter.id = as.integer(recruiter.id)
    )

  # Find all seeds (recruiter.id == -1 or NA)
  seeds <- chains %>%
    dplyr::filter(recruiter.id == -1 | is.na(recruiter.id)) %>%
    dplyr::pull(id)

  # If no seeds found, return zeros
  if (length(seeds) == 0) {
    return(list(longest = 0, mean_length = 0))
  }

  # Calculate depth for each node (distance from seed)
  get_chain_depth <- function(node_id, chains, depth = 0, visited = c()) {
    if (node_id %in% visited) return(as.numeric(depth))  # Prevent cycles
    visited <- c(visited, node_id)

    # Find children of this node
    children <- chains %>%
      dplyr::filter(recruiter.id == node_id) %>%
      dplyr::pull(id)

    if (length(children) == 0) {
      return(as.numeric(depth))  # Leaf node
    } else {
      # Return maximum depth among children
      child_depths <- vapply(children, function(child) {
        get_chain_depth(child, chains, depth + 1, visited)
      }, FUN.VALUE = numeric(1))
      return(as.numeric(max(child_depths)))
```

```r
    }
  }

  # Calculate longest chain from each seed
  chain_lengths <- vapply(seeds, function(seed_id) {
    get_chain_depth(seed_id, chains)
  }, FUN.VALUE = numeric(1))

  # Calculate average chain length (mean depth from seeds)
  all_depths <- c()
  for (seed_id in seeds) {
    descendants <- chains %>%
      dplyr::filter(recruiter.id == seed_id)

    if (nrow(descendants) > 0) {
      for (desc_id in descendants$id) {
        depth_val <- get_chain_depth(desc_id, chains, depth = 1)
        all_depths <- c(all_depths, as.numeric(depth_val))
      }
    }
  }

  # Ensure we have numeric vectors
  chain_lengths <- as.numeric(chain_lengths)
  all_depths <- as.numeric(all_depths)

  return(list(
    longest = ifelse(length(chain_lengths) > 0, max(chain_lengths, na.rm = TRUE), 0),
    mean_length = ifelse(length(all_depths) > 0, mean(all_depths, na.rm = TRUE), 0)
  ))
}

chain_stats <- calculate_chain_stats(rd.dd)

network_stats <- tibble(
  Characteristic = c("Total Sample Size", "Recruitment Waves", "Average Degree",
                     "Median Degree", "Seeds", "Longest Chain", "Mean Chain Length"),
  Value = c(
    nrow(rd.dd),
    max(rd.dd$wave, na.rm = TRUE),
    round(mean(rd.dd$numRef, na.rm = TRUE), 1),
    round(median(rd.dd$numRef, na.rm = TRUE), 1),
    sum(rd.dd$recruiter.id == -1 | is.na(rd.dd$recruiter.id)),
    chain_stats$longest,
```

Table 8: RDS Sample Network Characteristics

Table 9: RDS Sample Network Characteristics

Characteristic
Total Sample Size
Recruitment Waves
Average Degree
Median Degree
Seeds
Longest Chain
Mean Chain Length

Note: Average and median degree represent the number of other domestic workers known personally (Q...

```
    round(chain_stats$mean_length, 1)
  )
)

network_stats %>%
  kableExtra::kbl(
    col.names = c("Characteristic", "Value"),
    booktabs = TRUE,
    align = c("l", "r"),
    caption = "RDS Sample Network Characteristics"
  ) %>%
  kableExtra::kable_styling(
    latex_options = c("striped", "hold_position"),
    bootstrap_options = c("striped", "hover"),
    full_width = FALSE,
    position = "center"
  ) %>%
  kableExtra::footnote(
    general = "Average and median degree represent the number of other domestic workers
    general_title = "Note:",
    footnote_as_chunk = TRUE
  )
```

A.3.2 Demographic Composition

```
# Create nationality breakdown with proper ordering
nationality_summary <- rd.dd %>%
```

```r
  dplyr::group_by(nationality_cluster) %>%
  dplyr::summarise(
    n = n(),
    percent = round(n() / nrow(rd.dd) * 100, 1),
    .groups = 'drop'
  ) %>%
  dplyr::arrange(desc(n)) %>%
  dplyr::mutate(
    nationality_cluster = factor(nationality_cluster, levels = nationality_cluster),
    label_text = sprintf("%d\n(%.1f%%)", n, percent)
  )

# Create professional bar plot
ggplot(nationality_summary, aes(x = nationality_cluster, y = percent, fill = nationality
  geom_col(width = 0.7, alpha = 0.9) +
  geom_text(aes(label = label_text),
            vjust = -0.3,
            size = 3.5,
            fontface = "bold") +
  scale_fill_viridis_d(option = "plasma", begin = 0.2, end = 0.8) +
  scale_y_continuous(
    expand = expansion(mult = c(0, 0.15)),
    breaks = seq(0, 100, by = 10),
    labels = function(x) paste0(x, "%")
  ) +
  labs(
    x = "Nationality Cluster",
    y = "Percentage of Sample",
    title = "Sample Distribution by Nationality Cluster",
    subtitle = sprintf("Total N = %d", nrow(rd.dd))
  ) +
  theme_minimal(base_size = 11) +
  theme(
    legend.position = "none",
    plot.title = element_text(face = "bold", size = 13, hjust = 0.5),
    plot.subtitle = element_text(size = 10, hjust = 0.5, color = "gray30"),
    axis.title = element_text(face = "bold", size = 11),
    axis.text.x = element_text(angle = 0, hjust = 0.5, size = 10),
    axis.text.y = element_text(size = 10),
    panel.grid.major.x = element_blank(),
    panel.grid.minor = element_blank(),
    panel.background = element_rect(fill = "white", color = NA),
    plot.background = element_rect(fill = "white", color = NA)
  )
```

Sample Distribution by Nationality Cluster

Total N = 85
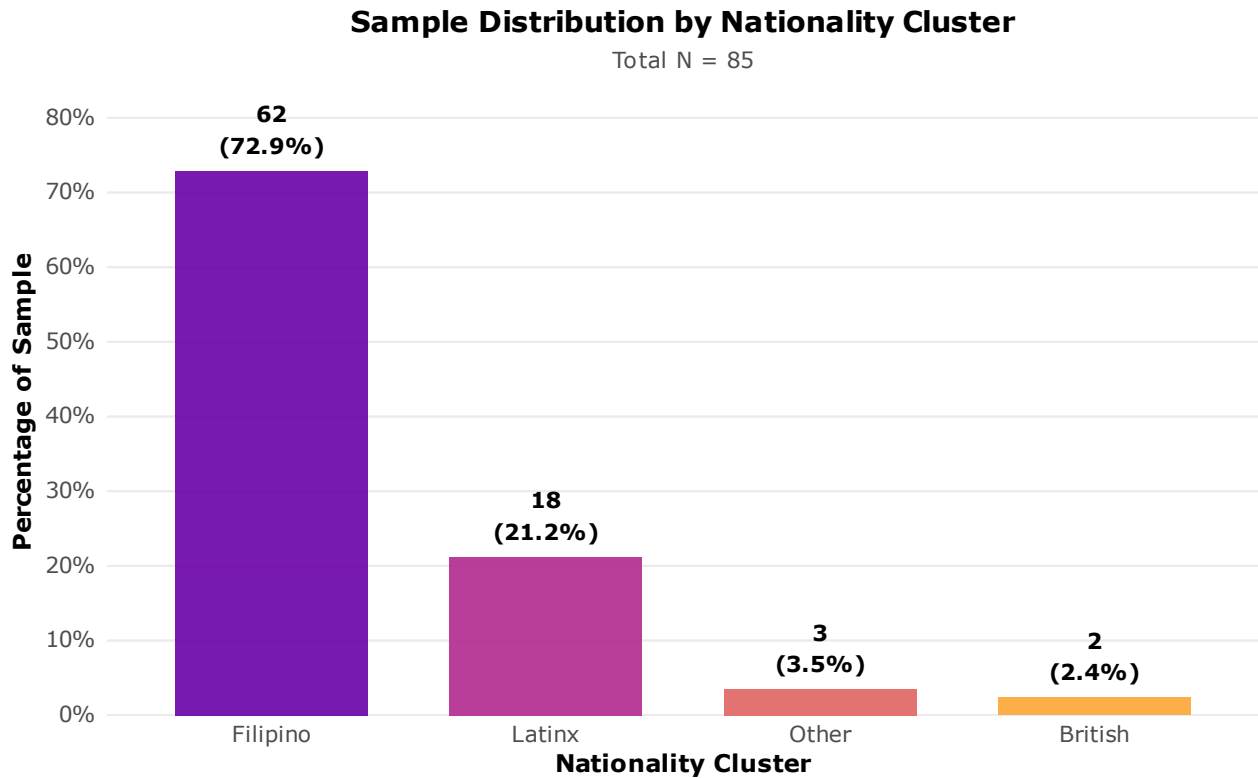
Figure 7: Sample Composition by Nationality Clusters

B Appendix B

This analysis estimates the prevalence of modern slavery among domestic workers in the UK using multiple RDS methodologies. We examine two primary indicators across various population size assumptions and estimation techniques.

B.1 RDS Estimation Results

This section presents comprehensive RDS estimates across all population size scenarios and estimation methods.

Table 10: Comprehensive RDS Estimates: Prevalence estimates for all indicators across population sizes and RDS methods (RDS-I, RDS-II, RDS-SS) with neighborhood bootstrap confidence intervals.

	RDS-I				RDS-II				RDS-SS			
Indicator	RDS-I 50K	RDS-I 100K	RDS-I 980K	RDS-I 1.74M	RDS-II 50K	RDS-II 100K	RDS-II 980K	RDS-II 1.74M	RDS-SS 50K	RDS-SS 100K	RDS-SS 980K	RDS-SS 1.74M
Document withholding	49.4% (25.5–56.2)	49.4% (25.2–55.9)	49.4% (25.3–56.3)	49.4% (25.4–56.1)	34.2% (25.5–56.3)	34.2% (25.3–56.2)	34.2% (24.7–56.3)	34.2% (25.4–56.5)	34.2% (25.6–56.2)	34.2% (25.1–56.5)	34.2% (25.2–56.0)	34.2% (25.3–56.5)
Excessive working hours	92.8% (71.8–92.8)	92.8% (71.8–92.5)	92.8% (71.4–92.9)	92.8% (71.7–92.8)	83.7% (71.9–93.0)	83.7% (71.5–93.1)	83.7% (71.1–92.9)	83.7% (71.5–93.0)	83.7% (71.5–93.0)	83.7% (71.5–93.0)	83.7% (71.5–93.0)	83.7% (71.6–92.8)
Limited access to help	57.7% (54.7–80.0)	57.7% (54.8–79.8)	57.7% (55.1–80.3)	57.7% (54.9–80.0)	59.4% (55.2–80.0)	59.4% (54.9–79.9)	59.4% (55.0–80.1)	59.4% (54.7–79.9)	59.4% (54.9–80.1)	59.4% (55.1–80.2)	59.4% (54.7–80.2)	59.4% (54.8–80.1)
Pay-related issues	59.3% (34.5–63.5)	59.3% (34.8–63.0)	59.3% (34.2–63.5)	59.3% (34.2–63.6)	49.7% (34.8–63.6)	49.7% (35.2–63.4)	49.7% (34.5–63.6)	49.7% (34.7–63.3)	49.7% (34.0–63.8)	49.7% (34.2–64.1)	49.7% (34.8–63.4)	49.7% (34.6–63.5)
Threats and abuse	65.2% (40.3–71.7)	65.2% (40.2–72.0)	65.2% (40.5–72.1)	65.2% (40.6–72.1)	57.2% (40.3–72.0)	57.2% (40.5–72.1)	57.2% (40.6–72.4)	57.2% (40.3–71.6)	57.2% (40.9–72.5)	57.2% (40.9–72.0)	57.2% (40.1–72.3)	57.2% (40.6–72.6)

Note: RDS-I: Salganik-Heckathorn estimator. RDS-II: Volz-Heckathorn estimator. RDS-SS: Gile's Successive Sampling estimator. Population sizes: 50K, 100K, 980K, 1.74M domestic workers. All confidence intervals from neighborhood bootstrap (n=1000).
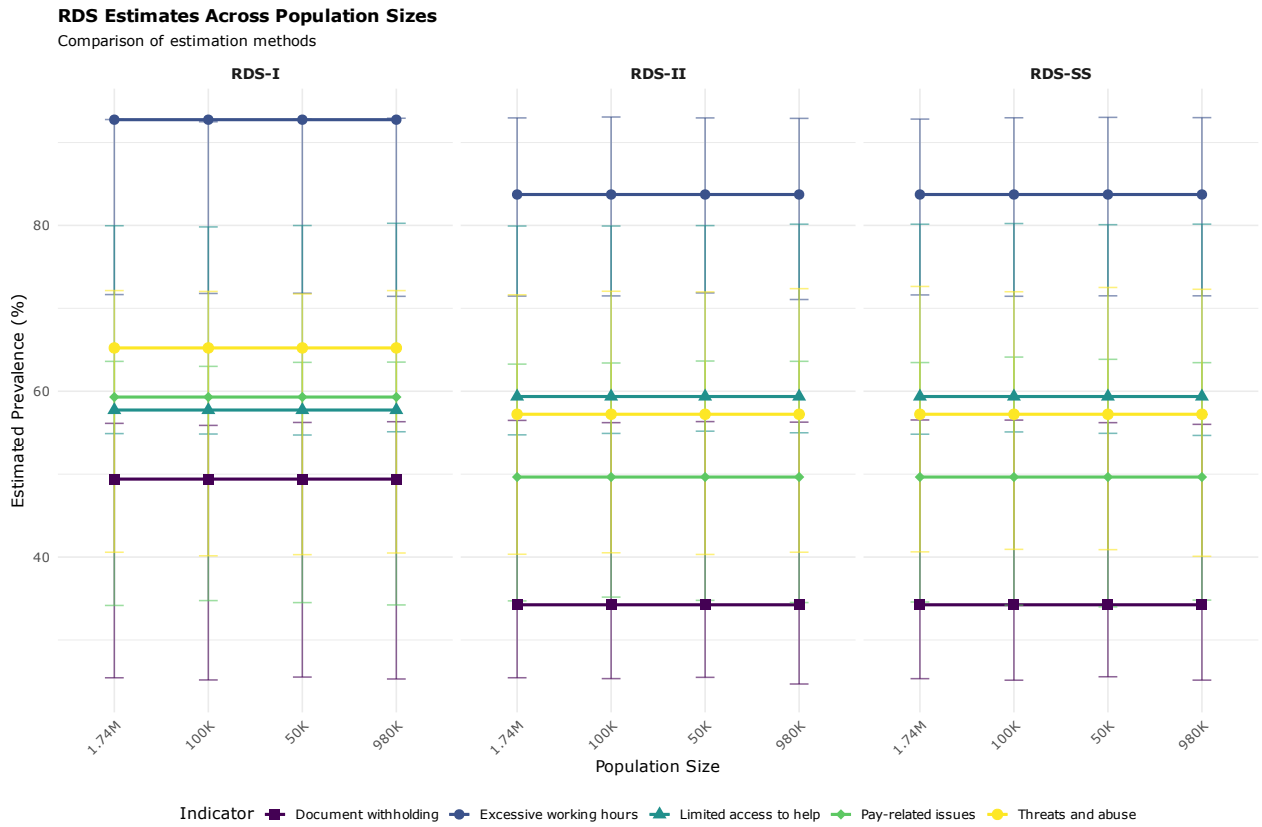
Figure 8: RDS Sensitivity to Population Size: Comparison of RDS estimates across all population size scenarios. Each panel shows one RDS method with point estimates and 95% bootstrap confidence intervals varying by population size.

39

Table 11: Model-Assisted Estimates by Population Size and Seed Selection. Source: Authors'
Own Work.

Seed Selection	Population Size	Prevalence Estimates (95% CrI)			
		Document Withholding	Pay Issues	Excessive Hours	Threats/Abuse
sample	50,000	34.1% (16.1%, 52.1%)	49.3% (35.2%, 63.4%)	84.0% (72.3%, 95.8%)	57.1% (34.6%, 79.6%)
sample	980,000	34.3% (16.4%, 52.2%)	49.5% (35.6%, 63.5%)	83.7% (71.9%, 95.6%)	57.3% (35.4%, 79.3%)
sample	1,740,000	34.3% (16.0%, 52.7%)	49.1% (34.9%, 63.3%)	83.8% (72.1%, 95.6%)	57.0% (34.7%, 79.3%)
random	50,000	34.1% (16.1%, 52.1%)	49.3% (35.2%, 63.4%)	84.0% (72.3%, 95.8%)	57.1% (34.6%, 79.6%)
random	980,000	34.3% (16.4%, 52.2%)	49.5% (35.6%, 63.5%)	83.7% (71.9%, 95.6%)	57.3% (35.4%, 79.3%)
random	1,740,000	34.3% (16.0%, 52.7%)	49.1% (34.9%, 63.3%)	83.8% (72.1%, 95.6%)	57.0% (34.7%, 79.3%)
degree	50,000	34.1% (16.1%, 52.1%)	49.3% (35.2%, 63.4%)	84.0% (72.3%, 95.8%)	57.1% (34.6%, 79.6%)
degree	980,000	34.3% (16.4%, 52.2%)	49.5% (35.6%, 63.5%)	83.7% (71.9%, 95.6%)	57.3% (35.4%, 79.3%)
degree	1,740,000	34.3% (16.0%, 52.7%)	49.1% (34.9%, 63.3%)	83.8% (72.1%, 95.6%)	57.0% (34.7%, 79.3%)

Note: Model-Assisted estimates from Gile & Handcock (2015) Bayesian inference. CrI = Credible Interval from posterior distribution.

Table 12: RDS Estimator Comparison for Binary Indicators. Source: Authors' Own Work.

NULL

B.2 Model-Assisted Estimates by Population Size

B.2.1 Traditional RDS Estimators Comparison

B.3 Model-Assisted Sensitivity Analysis

This section presents comprehensive sensitivity analysis for Model-Assisted inference estimates across all exploitation indicators, examining robustness to population size assumptions and seed selection methods.

Table 13: Model-Assisted Sensitivity Analysis: Point estimates and 95% credible intervals across all parameter combinations. Estimates from Gile & Handcock (2015) Bayesian Model-Assisted inference.

\begin{table}[!h] \caption{Model-Assisted Sensitivity Analysis: Point estimates and 95% credible intervals}

label	Degree: 50K	Random: 50K	Sample: 50K	Degree: 980K	Random: 980K	Sample: 980K	Degree: 1.74M	Random: 1.74M	Sample: 1.74M
Document With-holding	34.1% (16.1- 52.1%)	34.1% (16.1- 52.1%)	34.1% (16.1- 52.1%)	34.3% (16.4- 52.2%)	34.3% (16.4- 52.2%)	34.3% (16.4- 52.2%)	34.3% (16.0- 52.7%)	34.3% (16.0- 52.7%)	34.3% (16.0- 52.7%)
Pay-Related Issues	49.3% (35.2- 63.4%)	49.3% (35.2- 63.4%)	49.3% (35.2- 63.4%)	49.5% (35.6- 63.5%)	49.5% (35.6- 63.5%)	49.5% (35.6- 63.5%)	49.1% (34.9- 63.3%)	49.1% (34.9- 63.3%)	49.1% (34.9- 63.3%)
Threats/Abuse	57.1% (34.6- 79.6%)	57.1% (34.6- 79.6%)	57.1% (34.6- 79.6%)	57.3% (35.4- 79.3%)	57.3% (35.4- 79.3%)	57.3% (35.4- 79.3%)	57.0% (34.7- 79.3%)	57.0% (34.7- 79.3%)	57.0% (34.7- 79.3%)
Excessive Hours	84.0% (72.3- 95.8%)	84.0% (72.3- 95.8%)	84.0% (72.3- 95.8%)	83.7% (71.9- 95.6%)	83.7% (71.9- 95.6%)	83.7% (71.9- 95.6%)	83.8% (72.1- 95.6%)	83.8% (72.1- 95.6%)	83.8% (72.1- 95.6%)
Limited Access to Help	58.2% (46.0- 70.4%)	58.2% (46.0- 70.4%)	58.2% (46.0- 70.4%)	58.3% (46.1- 70.6%)	58.3% (46.1- 70.6%)	58.3% (46.1- 70.6%)	58.4% (46.2- 70.7%)	58.4% (46.2- 70.7%)	58.4% (46.2- 70.7%)
Any Ex-ploitation	94.9% (61.5- 100.0%)	94.9% (61.5- 100.0%)	94.9% (61.5- 100.0%)	95.0% (62.2- 100.0%)	95.0% (62.2- 100.0%)	95.0% (62.2- 100.0%)	94.8% (59.2- 100.0%)	94.8% (59.2- 100.0%)	94.8% (59.2- 100.0%)

Note: Estimates from Gile & Handcock (2015) Bayesian Model-Assisted inference. Population sizes: 50K, 100K, 980K, 1.74M domestic workers. Seed methods: Degree (by network degree), Random, Sample (from data). CrI = Bayesian credible interval.

\end{table}

Sensitivity of MA Estimates to Population Size and Seed Selection

Gile & Handcock (2015) Bayesian Model-Assisted Inference

MA Estimates Across Indicators (N = 980,000)
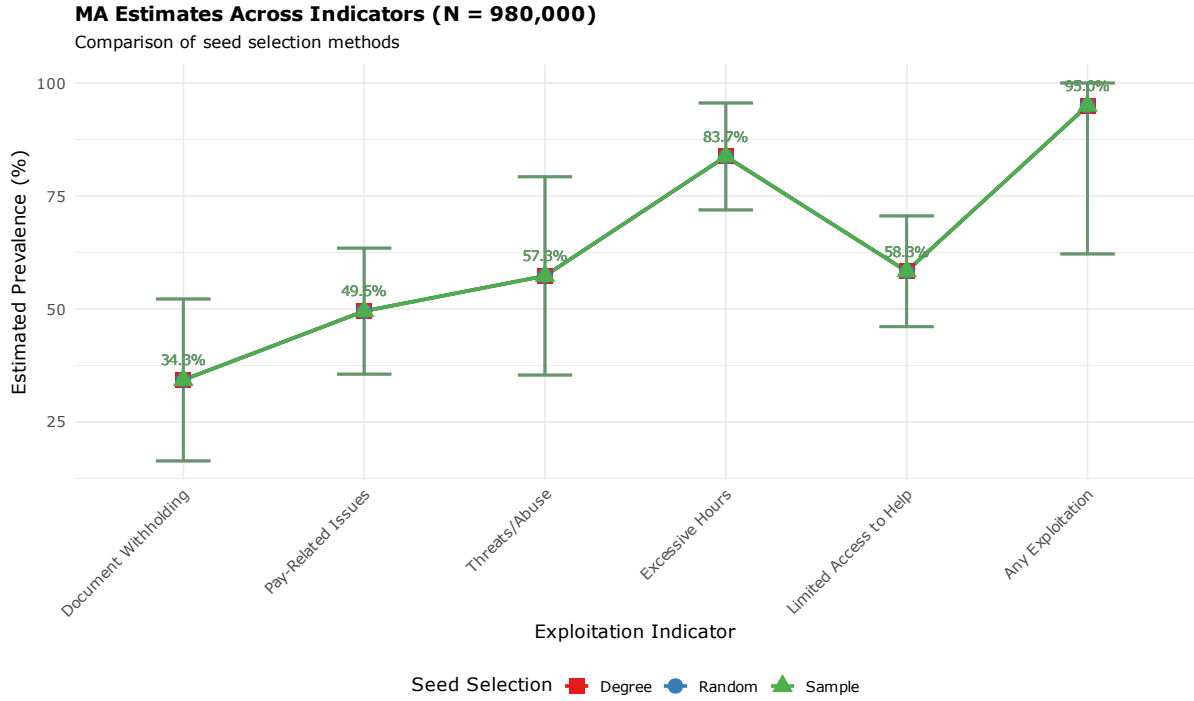Comparison of seed selection methods

Figure 10: Model-Assisted Point Estimate Comparison: Direct comparison of point estimates across all indicators, fixing population size at 980K. Lines connect estimates using the same seed selection method, showing consistency across indicators.

C Appendix C: Bootstrap Estimation for Network Scale-Up Using RDS Data

C.1 Introduction

When estimating the size or characteristics of a hidden population using the Network Scale-Up Method (NSUM), researchers typically assume a probability sample from the frame population. However, in many applied settings—including hard-to-reach populations—data are collected via*Respondent-Driven Sampling (RDS). RDS introduces specific structural dependencies and inclusion probabilities that violate the assumptions of simple random sampling.

This presents a challenge: how can we correctly estimate uncertainty for NSUM estimates derived from an RDS sample? As shown in **feeh16-generalized<empty citation>** and **salg06-variance<empty citation>**, the NSUM estimator depends crucially on inclusion weights π_i, which must reflect the sampling design. When data are RDS-based, these weights are typically derived from known degree-based estimators such as RDS-II or Gile's Successive Sampling (SS) weights.

To address this challenge, we propose a **three-step bootstrap procedure** for NSUM estimation using RDS data. This approach is flexible, modular, and applicable across several classes of NSUM estimators. It separates the issues of: 1. How to resample an RDS chain (Step 1), 2. How to recalculate sample-specific weights (Step 2), and 3. How to apply a chosen NSUM estimator (Step 3).

C.2 Step 1: Resampling the RDS Sample

We begin by resampling from the observed RDS sample in a way that mimics the original recruitment structure. Let the original sample be:

$$\mathcal{S} = \{i_1, i_2, ..., i_n\}$$

with recruitment chains and wave indicators. Let d_i denote self-reported degree for respondent i, and let the recruitment tree structure be encoded via seed/recruiter IDs.

C.2.1 Options for RDS Resampling

- **Tree Bootstrap**: Sample entire recruitment trees (originating from seeds) with replacement. This respects the hierarchical recruitment structure and allows design effect estimation (**salg06-variance**).
- **Successive Sampling Bootstrap (SSB)**: Sample with replacement according to inclusion probabilities derived from the SS model (**gile11-improv<empty citation>**).
- **Neighborhood Bootstrap**: Use ego-network topology to preserve recruitment ties and neighborhood structure (**yauc22-neighboor<empty citation>**).
- **Chain Bootstrap:** Each bootstrap replicate samples chains or subchains with replacement, preserving recruiter–recruitee links. This is implemented in `surveybootstrap::rds.boot.draw.chain()` and used in studies like **weir12-comparison<empty**

Let $\mathcal{S}^{(b)}$ denote the sample drawn in bootstrap replicate b.

C.2.2 Resampling Procedure

We selected our first wave of par5cipants nonrandomly by convenience sampling. Mobile phone numbers were used both to iden5fy seed par5cipants and to act as a unique iden5fier for those whom they referred. We used the publicly available 'Neighboot' bookstrapping procedure to es5mate confidence limits from our sample (Yauck and Moodie, 2020). Bootstrapping contributes to overcoming the high degree of uncertainty surrounding es5mates from samples generated through respondent-driven sampling and allows beLer sta5s5cal inference, especially related to sampling variability (Baraff et al., 2016). These authors develop

a mul5-level tree bootstrapping procedure based on resampling recruitment trees that depends solely upon the structure of the sampled recruitment trees rather than the aLributes being measured on the respondents. This allows es5ma5on of variability and correla5ons between aLributes. Their tree bootstrap method can produce interval es5mates that are able to account for the high variability of the RDS process, even when the design effects are very large, producing es5mates that are slightly conserva5ve – which is preferable in this situa5on. The process involves a sample, without replacement from ini5al seed par5cipants, followed by resampling from each of the selected seeds recruits and further resampling form second-level recruits' recruits and con5nues un5l there are no recruits remaining (Baraff et al., 2016; Yauck et al., 2022).More recently, however, Yauck et al. (2022) have raised concerns that the tree bootstrapping method proposed by Baraff et al., 2016) severely over es5mates uncertainty and, instead, propose a neighborhood bootstrap method to quan5fy uncertainty in RDS. This is designed to capture the 'local' neighborhood depdendence by including the selected recruiters' recruits. Yauck et al., (2022) argue that this approach improves upon previously advocated approaches because the 'local' neighbors' 5es are present in any full, unobserved network of the target popula5on. This method is par5cularly suitable for our study, since it offers computa5onal efficiencies which yield confidence intervals with conserva5ve coverages for small sample sizes and is based on resampling recruited individuals and their observed neighbors -defined as those individuals with whom they are directly connected within the RDS tree. Yauck et al.'s method relies on the following assump5ons: i) social connec5ons are reciprocal and ii) an individual within the network can reach another individual through a finite set of connec5ons. Their bootstrap method is based upon sequen5ally resampling recruiters and their recruits and can be described in two main steps: step 1 involved the random selec5on, with replacement of the number of recruiters within the RDS sample. Step 2 then includes the selected recruiters and their corresponding recruits in the bootstrap sample. The method allows the construc5on of calibrated interval es5mates. Unlike other methods, the advantages of this approach include its ability to derive es5mates from a single bootstrap sample for any number of aLributes. Limita5ons Baraff et al. (2016)'s approach relies on a series of simplifying assump5ons, namely: i) the social network is finite and connected ii) network connec5ons are reciprocal, not directed [I believe this to be true of the respondents in our study] iii) recruits accurately report their network degree [this is facilitated by the concrete request to provide a count of domes5c workers contacts within their mobile phone – thereby minimising recall error[iv) recruitment coupons are distributed uniformly at random to neighbours in the network [in our study, the research team managed the distribu5on of the survey link to those suggested by our seeds and respondents from earlier waves] v) that individuals may be recruited into the sample more than once. Baraff et al., note that this laLer assump5on par5cularly does not relate to the reality of RDS in prac5ce.

References Baraff, Aaron J., Tyler H. McCormick, and Adrian E. Ra'ery. 2016. "Es5ma5ng Uncertainty in Respondent-Driven Sampling Using a Tree Bootstrap Method." Proceedings of the Na5onal Academy of Sciences of the United States of America 113 (51): 14668–73. hLps://doi.org/10.1073/pnas.1617258113. Feehan, Dennis M., and MaLhew J. Salganik. 2016. "Online Appendix: Generalizing the Network Scale-up Method." Sociological Methodology 46 (1): 153–86. hLps://doi.org/10.1177/0081175016665425. Mamadou Yauck, Erica E M Moodie, Herak Apelian, Alain Fourmigue, Daniel Grace, Trevor A Hart, Gilles Lambert, Joseph Cox, Neighborhood Bootstrap for Respondent-Driven Sampling,

Journal of Survey Statistics and Methodology, Volume 10, Issue 2, April 2022, Pages 419–438, https://doi.org/10.1093/jssam/smab057 Yauck, Mamadou, and Erica E. M. Moodie. 2022. Neighboot: Neighborhood Bootstrap Method for RDS (version 1.0.1). https://cran.r-project.org/web/packages/Neighboot/index.html.

C.3 Step 2: Recalculating Weights

NSUM estimators require inclusion weights π_i or their inverses $w_i = 1/\pi_i$. Because bootstrap samples differ in composition and recruitment pattern, these weights must be **recomputed for each replicate**.

C.3.1 General Structure

For each replicate b, construct: - $\mathcal{S}^{(b)}$: resampled respondent IDs - $d_i^{(b)}$: degree reports in replicate - $\pi_i^{(b)}$: estimated inclusion probabilities

C.3.2 Weighting Options

- **RDS-II Weights (volz08-simple<empty citation>)**:

$$w_i^{(b)} \propto \frac{1}{d_i^{(b)}}$$

- **SS Weights (gile11-improv<empty citation>)**: Incorporate sampling fraction and frame size N_F. Computed numerically via successive sampling approximation.

Let $\mathbf{X}_i$ denote covariates (e.g. traits, degree, indicator of hidden population membership), which are retained from the original data and passed to Step 3.

C.4 Step 3: NSUM Estimation

C.4.1 Step 3: NSUM Estimation on Resampled and Reweighted Data

Given a bootstrap sample $\mathcal{S}^{(b)}$, we estimate the size of the hidden population N_H using one of several NSUM estimators. All estimators rely on weighted sums of out-reports and degree.

C.4.1.1 (a) Generalized NSUM (GNSUM)

As described in **feeh16-generaling<empty citation>**, GNSUM estimates:

$$\hat{N}_H^{(b)} = \left(\frac{\sum_i \frac{y_{i,H}}{\pi_i^{(b)}}}{\sum_i \frac{d_i}{\pi_i^{(b)}}} \right) \cdot N_F$$

Where:

- $y_{i,H}$: number of known alters in the hidden population,
- d_i: degree (known network size, e.g. from Q13),
- N_F: size of the frame population (e.g. domestic workers in UK).

C.4.1.2 (b) GNSUM with Symmetric Visibility (for Hidden Members in RDS)

In your context, some RDS respondents are ex post identified as members of the hidden population. Under the **symmetric visibility assumption**, if $i \in H$, we assume the probability that i knows j is the same as j knows i. This allows **in-reports** to be incorporated.

Let:

- $I_H(i) = 1$ if $i \in H$, 0 otherwise
- y_i^{in}: number of other hidden members who report knowing i

Then the estimator becomes:

$$\hat{N}_H^{(b)} = \left(\frac{\sum_i \frac{y_{i,H} + I_H(i) \cdot y_i^{\text{in}}}{\pi_i^{(b)}}}{\sum_i \frac{d_i}{\pi_i^{(b)}}} \right) \cdot N_F$$

C.4.1.3 (c) Modified Basic Scale-Up (MBSU)

This estimator adjusts for key reporting biases via three correction factors:

$$\hat{N}_H^{(b)} = \left(\frac{\sum_i \frac{y_{i,H}}{\pi_i^{(b)}}}{\sum_i \frac{d_i}{\pi_i^{(b)}}} \right) \cdot N_F \cdot \frac{1}{\delta \cdot \tau \cdot \rho}$$

Where:

- δ: **Transmission bias** (probability respondent is aware of alter's status),
- τ: **Barrier effect** (social mixing between hidden and frame population),

- ρ: **Popularity bias** (relative visibility of hidden population members).

These values can be:

- Estimated using **known subpopulations** (e.g., alter groups with known size),
- Set by expert **elicitation**,
- Scanned in **sensitivity analysis**.

C.4.1.4 (d) Model-Based NSUM

Bayesian implementations (e.g. Maltiel et al. (2015)) model:

- Degree distributions,
- Reporting error,
- Group visibility,
- Hidden population size.

They yield a **posterior distribution** over N_H, and uncertainty is embedded within the model.

This step applies an NSUM estimator to the bootstrap sample $\mathcal{S}^{(b)}$ using the recalculated weights $w_i^{(b)}$ and responses $y_{i,H}$, where $y_{i,H}$ is the number of known contacts respondent i has in hidden population H.

Let N_F denote the frame population size (assumed known), and let d_i be the degree of respondent i.

C.4.2 Generalized NSUM Estimator (GNSUM)

The weighted GNSUM estimator is:

$$\hat{N}_H^{(b)} = \frac{\sum_{i \in \mathcal{S}^{(b)}} \frac{y_{i,H}}{\pi_i^{(b)}}}{\sum_{i \in \mathcal{S}^{(b)}} \frac{d_i}{\pi_i^{(b)}}} \cdot N_F$$

This estimator assumes proportional mixing and equal visibility.

C.4.3 Symmetric Visibility Variant

In our setting, some RDS respondents can be identified *ex post* as members of the hidden population H. Denote this set $\mathcal{H} \subseteq \mathcal{S}$. Under the assumption of **symmetric visibility**, we define:

$$\hat{v}_H = \frac{1}{|\mathcal{H}|} \sum_{j \in \mathcal{H}} \frac{y_{j,F}}{d_j}$$

That is, the average proportion of alters known by members of H who are in the frame population F. Incorporating this, the symmetric visibility GNSUM becomes:

C.4.4 Step 4: Aggregating Bootstrap Estimates

After B replicates:

$$\hat{N}_H^{(1)}, ..., \hat{N}_H^{(B)}$$

We compute:

- **Point estimate**: $\bar{N}_H = \frac{1}{B} \sum_b \hat{N}_H^{(b)}$
- **Standard error**: sample SD
- **95% CI**: empirical percentiles (e.g., 2.5%, 97.5%)

C.4.5 Step 2: Recalculate Inclusion Probabilities

RDS produces samples with **unknown and unequal probabilities of inclusion**, which must be corrected when used in NSUM estimation. This is achieved by estimating the probability $\pi_i^{(b)}$ that each individual i in bootstrap replicate b is included in the sample, conditional on their degree and position in the recruitment tree.

These probabilities are used to generate **sampling weights** $w_i^{(b)} = 1/\pi_i^{(b)}$, which are passed into NSUM estimators.

C.4.5.1 Estimation Methods

(a) RDS-II (Volz-Heckathorn) Weights
Assumes the probability of selection is proportional to the respondent's degree:

$$\pi_i \propto d_i \quad \Rightarrow \quad w_i = \frac{1}{d_i}$$

These weights are normalized post hoc. This method is simple but does not account for homophily or finite population correction (**volz08-rds\<empty citation\>**).

(b) Gile's Successive Sampling (SS) Weights
This method assumes RDS approximates successive sampling without replacement. The inclusion probabilities are computed by simulating from a known or assumed population size N. This method is implemented in `RDS::gile.ss.weights()` and adjusts for:

- Finite population effects,
- Degree-based sampling,
- Sample depletion over waves.

(c) User-Defined or Model-Based Weights
Researchers may define weights using alternative models or Bayesian simulations. This includes post-stratification or fitting full generative models of the RDS process.

C.4.5.2 Software Example (R)

```
library(RDS)
rd <- as.rds.data.frame(boot_sample, id = "id", recruiter.id = "recruiter.id")
boot_sample$ss_weight <- gile.ss.weights(rd, N = 980000)
boot_sample$vh_weight <- rds.weights(rd, weight.type = "RDS-II")
```

The output is a new dataset with respondent traits, degrees, and updated $\pi_i^{(b)}$, which are used in Step 3.

C.4.6 5. References

- **feeh16-generalizing<empty citation>**
- **salf06-variance<empty citation>**
- **gile11-inference<empty citation>**
- **volz08-rds<empty citation>**
- Maltiel et al. (2015)
- Yauck and Moodie (2022)
- **weir12-comparison<empty citation>**
- **salg06-variance<empty citation>**
- **rust96-rescaled<empty citation>**
- **rao88-resampling<empty citation>**