

### **Exploratory analysis of survey responses for GNSUM estimation (based on Feehan and Salganik, 2014)**

1. Number in sample frame – number of respondents in domestic work in the UK within the past 12 months.

Calculation:

**Total number of respondents**, excluding incomplete responses and any respondents not in DW in the past 12 months.

2. In total, how many domestic workers are known by our survey respondents?

Calculation:

**Sum** of response to Q. 2f : ‘How many domestic worker contacts do you have on your mobile phone?’ from all legitimate respondents.

Use: This gives the alter sample size and can be used along with the government estimate of overseas domestic worker visa numbers at 18,553 in 2022 to estimate the absolute value of the hidden population.

3. Total number of the frame sample reporting awareness of domestic workers experiencing all types of exploitation

Calculation: **sum** of every respondents’ answer to Q 2.g : how many domestic workers do you know that have experienced exploitation?

Use: Numerator for estimate of hidden population :this gives aggregated relational data about connections of the frame sample to the hidden population of exploited domestic workers e.g. the sum of all items in the frame sample ( $y_i, H$ ) number of out reports of links to exploited domestic workers from person  $i$ .

4. Weighted average number of connections

Calculation: **Average (mean)** number of reported connections to everyone in the frame sample (i.e. the number of domestic workers each respondent knows) i.e. **Sum** of number of connections to domestic workers each respondent knows Q [please insert appropriate question reference here] divided by number of respondents who gave an answer to this question

Use: Denominator: This gives an estimate of  $d_i, F$  – the number of undirected network connections each respondent has to everyone in the sample frame ( $F$ )

5. GNSUM according to equation 23 (Feehan and Salganik, 2014)

Calculation: Answer 3 divided by Answer 4 divided by Answer 1

Use: Estimate of the hidden population in the sample according to equation 23 (Feehan and Salganik, 2014)

6. To calculate the absolute number of those exploited in the population,

Calculation: Official estimate of Overseas Domestic Workers (18,553) divided by Answer 2 (size of the alter sample) multiplied by Answer 5 (estimate of the hidden population).

We need also to state how we have calculated the following:

7. Degree ratio:

Calculation: For each respondent who self-identifies as being exploited, **sum** of the number of domestic workers with whom each respondent knows **divided** by the **number of respondents** who self-identify as being exploited.

Use: Equal to the average number of connections from a member of the hidden population (H) to the Sample Frame population (F) (Feehan and Salganik 2014 equation 18)

8. True positive rate:

Calculation: **Sum** of the number of domestic workers who are reportedly known to be exploited by all qualifying respondents **divided by count** of the total number of self-reports of exploitation (Need to check that this interpretation is correct).

Use: Equal to the number of in reports to H (hidden population of exploited domestic workers) from F (frame sample) divided by the number of edges connecting H and F.

9. Variance estimation and confidence limits

Need to establish a way to automate and run the following procedure - perhaps in R? [Selim, take a look at this, but do say if this is something for which we need additional support]

i) generate B [we would need to decide on the value of B] replicant samples by randomly sampling with replacement from Frame sample sF]

ii) use these replicate samples to produce a set of replicate estimates Estimate NH1 ... Estimate NHB

iii) combine to produce a confidence interval, for example by the percentile method which chooses the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles of the B estimates (Fig F.1) Efron and Tibshirani, 1993).

10. Measures of robustness

We may also need some help with these.