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

I suggest that we aim initially for two sets of point estimates – a) using all valid responses and b) focused on respondents from the Filipino/a community. I have reflected these two sets of estimates in the calculations below.

A) Number in sample frame – number of respondents in domestic work in the UK within the past 12 months. (sum of Q5==0)

B) Calculation of the number of respondents in a sub-sample of only Filipo domestic workers (sum of (Q5==0 AND Q8==(Filipina, Filipino, File, Pinay and any other similar variants)

Calculation:

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

1. In total, how many domestic workers are known by our survey respondents? (sum of Q13)

Calculation:

**a) Sum** of response to Q13: 2 f for all Q5==0 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 size of the exploited population by multiplying the sample proportion estimated through the basic network scale-up method by the size of the known population. I believe that we should half this figure to get an estimate for the number of valid visa holders at any one time. (Visas are issued for a 6 month period only)

b) Sum of response to Q13 for (Q5==0 AND Q8==(Filipina, Filipino, File, Pinay and any other similar variants))

This gives the alter sample size for Filipino domestic workers, which can be used along with the government number of Overseas Domestic visas issued in 2022 (10,186) to estimate the number exploited within the overall population. Again this figure should be halved to represent an estimate of the valid visas in a six-month period.

1. Total number of the frame sample reporting awareness of domestic workers experiencing all types of exploitation (sum of Q35 or Q88, Q89, etc.)

~~Calculation:~~ **~~sum~~** ~~of every respondents’ answer to Q 14 2.g : how many domestic workers do you know who are dissatisfied with their working conditions?~~

For each respondent group (a and b), calculate the following

Total numbers in the frame sample reporting awareness of domestic workers experiencing the following types of exploitation:

Abuse of vulnerability (number of unique respondents for whom Q29: 5a1== (0, 1,2) or Q51:5d2-a==0) i.e. ((Number Q29 union Number Q51) –(Number Q29 intersect Number Q51))

Deception (number of respondents for whom Q45:5c2==0)

Restriction of movement (number of unique respondents for whom Q32:5a4==(0,1,2) or Q 46:5c3==(2,3))

Isolation (number of respondents for whom Q65==(1)

Physical and sexual abuse (number of unique respondents for whom Q44: 5c1==(0) or Q44 5c1-b=0)

Intimidation (number of respondents for whom Q47==(0,1,2))

Retention of identity documents (number of respondents for whom Q70==(0,1,2))

Withholding of wages (number of respondents for whom Q42 5b==(0,1,2))

Debt bondage (number of respondents for whom Q39: 5b4==0)

Abusive working and living conditions (number of respondents for whom Q48: 5c5==(0,1,2) or Q63 :5d10 ==(0,1,2) or Q72: 5f4==(2,3) or Q76: 5f8 ==(1,2,3))

Excessive overtime (number of respondents for whom Q16:3b==(6,7) or Q61: 5d8==(2,3) orQ62:5d9==(0,1,2))

Use: Numerators for multiple or index estimates 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 (yi,H) number of out reports of links to exploited domestic workers from person I.

1. Weighted average number of connections

Calculation: **Average (mean)** number of reported connections to everyone in the frame samples a and b (I.e. the number of domestic workers each respondent knows) i.e. **Sum** of number of connections to domestic workers each respondent knows Q13: 2f [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 di,F – the number of undirected network connections each respondent has to everyone in the sample frame (F)

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

Calculation: Answer 3 divided by Answer 4 divided by Answer 1 for samples a and b

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

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

a) Overall sample Calculation: Official estimate of Overseas Domestic Workers (19,780/ 2) divided by Answer 2a (size of the alter sample) multiplied by Answer 5 (individual estimates/ index of the hidden population).

b) Calculation for Filipino workers only: Government estimate of Overseas Domestic Workers (10,186/2) divided by Answer 2b multiplied by Answer 5 (individual estimates/ index of the hidden population)

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

1. Degree ratio (equation 17):

Calculation for both samples a and b

For each of those who have been exploited (Q80: 5f12 (0,1)), the average number of their connections to members of the sample F (i.e. number of times that their mobile number appears in the contacts of the Frame sample).

Divided by:

The average number of connections from a member of F to the rest of F (for each respondent, the average number of times a mobile phone number appears in the contacts of other members of the frame sample).

Perhaps the argument that we make for focusing on the degree ratio is that it gives us a measure of the relative isolation of domestic workers in the hidden population (i.e. are they more isolated on average than other workers)?

We would be looking for a degree ratio of less than 1 - If the degree ratio is less than 1 then ‘the hidden populations members have, on average, fewer connection to the frame population than frame population members’. (p.168)

8. True positive rate:

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)

Calculation:

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 2.5th and 97.5th percentiles of the B estimates (Fig F.1) Efron and Tibshirani, 1993).

10. Measures of robustness

We may also need some help with these.

11. How many people do you know who have been advised to enter the NRM? Q 83. 5f15: Do you know personally of any other domestic workers who went through the NRM system? If so, how many? (sum of Q83)

Use: This gives us a probe alter since we know that, from April 2009 to March 2023, there have been 73,019 referrals to the NRM.

Calculation: Sum of Q. 83: 5f15 divided by number of NRM referrals (73,019)