

## Sampling Method of the Indonesia High-frequency Phone-based Monitoring of COVID-19 Impacts on Households (HiFy) – the Second Sample

(February 16, 2023)

In addition to the analysis of employment status of households' primary breadwinners, analysis of the individuals' employment with a more gender-balanced sample that better represents the Indonesian employment market may be done. Initially, the HiFy survey only collected individual-level information about households' primary breadwinners. Since Round 5 (R5), the HiFy Second Sample (HiFy-SS), which represents the Indonesian adult population of 18 to 64 years old, was designed.

Randomly selected individuals for the HiFy-SS could be the existing primary breadwinner or another household member, who was called the 'second respondent'. The second respondents were asked about their income and employment status, including some retrospective information prior to the pandemic and at the time of survey Rounds 1 and 3. While the primary breadwinner's information can be obtained from the usual employment module, the equivalent questions for the second respondents can be obtained from the questionnaire's 2<sup>nd</sup> respondent module.

A specific data set was prepared that lists all the individuals selected for the HiFy-SS i.e., *hify\_covid19\_ss\_weight\_all.dta* which also contains a variable called "sample\_2\_type" to identify whether he/she is a primary breadwinner or a second respondent. These data need to be utilized to build a complete dataset for the HiFy-SS.

### Sampling Frame

The sampling frame for the HiFy-SS was panel households from HiFy Rounds 1 to 3 consisting of 3,981 households and 17,407 persons. The HiFy-SS individual sampling frame, or eligible sample, were persons living in the household sampling frame aged 18 to 64 years old. About 3,942 (99%) households are eligible samples. The eligible sample in the sampling frame was 10,785 (62%) persons.

### Sample size

The sample size of the HiFy-SS was one eligible member per household. This would result in a sample of 3,942 persons identified as the primary target. The reserved target for HiFy-SS was also drawn as a replacement if the primary target cannot be interviewed. About 3,653 persons were selected as the reserved target. The distribution of the HiFy-SS target samples by sex and age group is in the table below.

Characteristics	Primary Target	Reserved Target
Sex		
Male	1,944 (49.3%)	1,793 (49.1%)
Female	1,998 (50.7%)	1,860 (50.9%)
Age-group		
18 – 29	960 (24.4%)	1,011 (27.7%)
30 – 49	2,134 (54.1%)	1,915 (52.4%)
50 – 64	848 (21.5%)	727 (19.9%)

### Sampling method

In addition to the sampling method of the HiFy survey, one eligible person in each selected household within the sampling frame of the HiFy-SS was randomly selected for the primary target, and another eligible person was randomly selected for the reserved target.

### Weighting

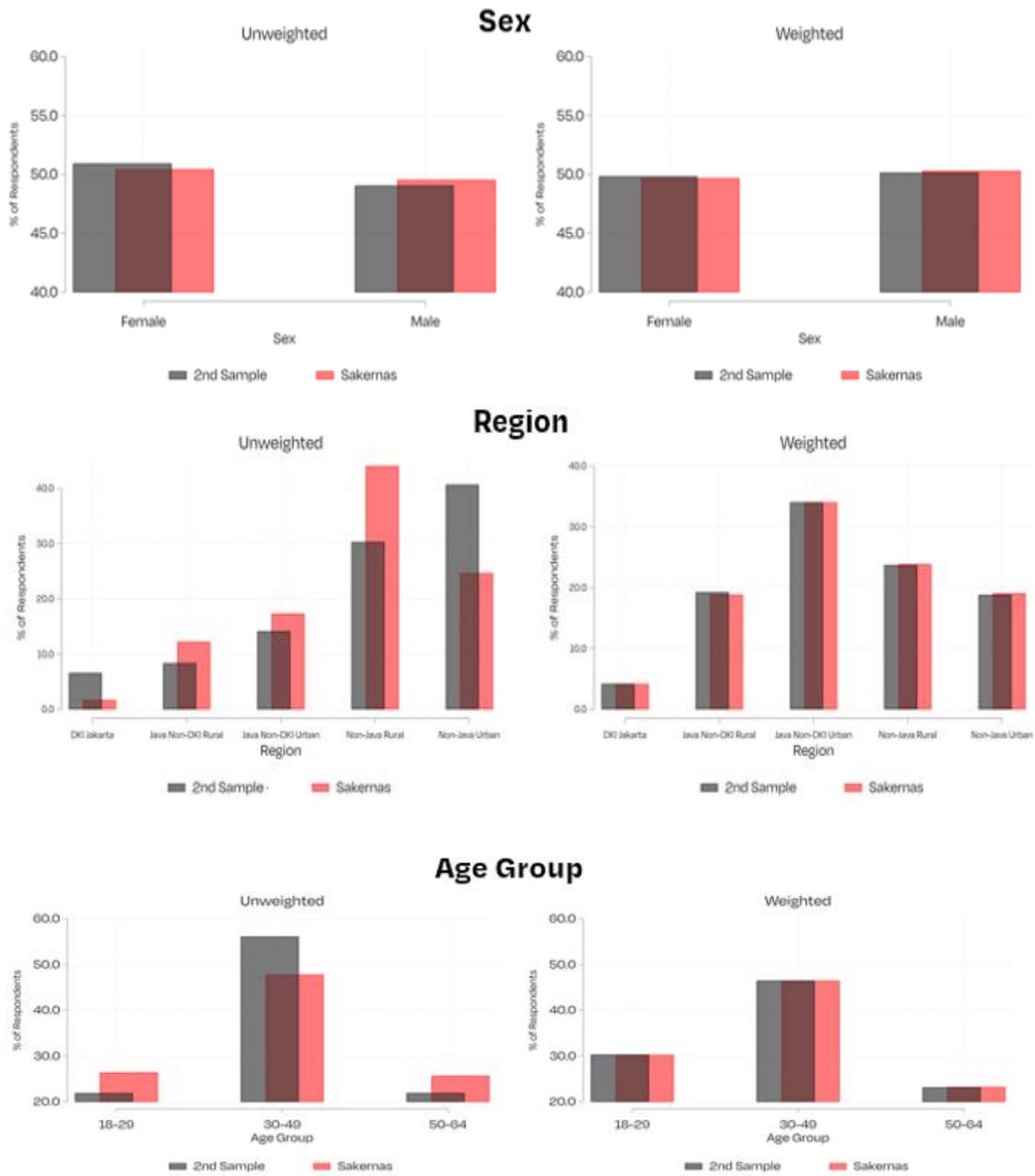
The selection probability for each eligible sample in each selected household was calculated by one divided by the number of eligible persons in the household. The overall sampling weight for the HiFy-SS in each selected household was calculated by multiplying the number of eligible persons by the HiFy survey weight for the same household.

### Response rate

Out of 3,942 persons as the target sample in R5, 3,629 (92%) responded, and they became the actual sample of the HiFy-SS. About 3,506 (96.6%) observations were from the primary target and 123 (3.4%) observations were from the reserved target. Furthermore, the 3,629 individuals were revisited during R6 and R7; the response rate of the HiFy-SS was 3,231 (89%), and 2,994 (82.5%) respectively. The distribution of the HiFy-SS actual sample by sex and age group are presented in the table below.

Characteristics	R5 sample	R6 sample	R7 sample	R7 balanced panel
Sex				
Male	1,772 (48.8%)	1,585 (49.1%)	1,484 (49.6%)	1,437 (48.9%)
Female	1,857 (51.2%)	1,646 (50.9%)	1,510 (50.4%)	1,502 (51.1%)
Age-group				
18 – 29	827 (22.8%)	708 (21.9%)	598 (20.0%)	592 (20.1%)
30 – 49	1,998 (55.1%)	1,814 (56.2%)	1,727 (57.7%)	1,692 (57.6%)
50 – 64	804 (22.1%)	707 (21.9%)	669 (22.3%)	655 (22.3%)

## The distribution of the HiFy-SS vs National Labor Force Survey (SAKERNAS)



#### **NOTE: Setting-up sampling parameter**

Prior to using the HiFy-SS data for analysis, a sampling set-up needs to be done using the following commands in STATA:

```
svyset psu [pweight=fwgt], strata(strata) fpc(N1h) vce(linearized) singleunit(certainty) || ssu,  
fpc(N2hi) || tsu, fpc(Zhij) || _n, fpc(n_elig)
```

where,

- *psu* is the primary sampling unit (district)
- *ssu* is the secondary sampling unit (village)
- *tsu* is the tertiary sampling unit (household)
- *\_n* is the quaternary sampling unit (person)
- *fwgt* is the designated weight
- *strata* is strata for PSU (region)
- *N1h* is the finite population correction for PSU (number of districts in stratum-h)
- *N2hi* is the finite population correction for SSU (number of villages in district-i)
- *Zhij* is the finite population correction for TSU (number of households in village-j)
- *n\_elig* is the finite population correction for QSU (number of eligible persons in household-k)