Sampling Methodology for Malaysia's Phone-Based High-Frequency (HiFy) Monitoring of COVID-19 Impacts on Households

I. SAMPLING FRAME

The Malaysia HiFy survey is a computer-assisted telephone interviewing (CATI) based survey, with a mobile-only frame generated via random digit dialing (RDD). Documentation from the Malaysian Communications and Multimedia Commission (MCMC) was used to identify all active mobile provider codes. All possible subscriber combinations were generated in D Force Sampling's Reactive User Interface Database (DRUID)[1], which houses the complete sampling frame. From this database, complete random telephone numbers were sampled. There was a total of 102,780,000 possible mobile numbers in the sampling frame, which were not stratified. [2]

II SAMPLING METHOD

Target Population: Telephone accessible population of Malaysian citizens, aged 18+

Target Sample Size: 2,200 Malaysians aged 18+ who either make financial decisions, contribute to or are knowledgeable on household finances.

The total number of phone numbers sampled for Round 1 was 33,894. Of these, 2,210 interviews were completed. In Round 2, all of 2,210 respondents who were successfully interviewed in Round 1 were followed up. Of these, 1,047 interviews were completed. In Round 3, all of 1,047 respondents who were successfully interviewed in Round 2 were followed up. Among these, 667 interviews were completed. In addition, 446 fresh respondents (new respondents who did not participate in previous rounds) were added

[1] DRUID is the survey company's (Central Force) SQL database interface developed for the generation, management, and sampling of RDD telephone databases. DRUID houses the complete sampling frame for this study, comprised of over 100 million possible mobile numbers.

[2] There are more than 18 mobile providers in the sampling frame.

and interviewed. These fresh respondents were sampled from the same sample frame as the panel samples.

II A. PANEL SAMPLES

The sampling methodology for the panel samples was as follows:

1. **Simple random sample of numbers**: In Round 1, mobile numbers were selected via simple random sampling from the sample frame through DRUID.
    a. All numbers were sampled without replacement within the survey wave.
    b. However, numbers can be reintroduced into the sample frame for subsequent waves, should there be continued studies.
2. **Preliminary auto-dialer filtering:** In Round 1, an auto-dialer filtering procedure was implemented; once the sample was drawn in the form of replicates (subsamples) of n=10,000, the mobile numbers were filtered using an auto-dialer to determine each number's working status.
    a. Up to two automated call attempts per number were done to determine whether the number was active or not, recording a disposition for each attempt. A minimum of two hours between each attempt was specified in the system.
    b. All numbers that yielded a working call disposition for at least one of the two filtering attempts were then passed to the CATI center human interviewing team.

    Round 2 and Round 3 did not undergo auto-dialer filtering as the sample comprised participating respondents from Round 1.
3. **Respondent selection**: The devices were assumed to be personal, and therefore the person who answered was the selected respondent.
    a. In Round 1, screener questions asking the age of the person who answered were used to ensure that the respondent was at least 18 years old, and within the capacity of either contributing, making or with knowledge of household finances.
    b. In Rounds 2 and 3, screener questions were used to ensure that the respondent of the recorded phone number in each respective round was the same person who participated in the previous round of the survey.

4. **Respondent and interviewer matching**: Each selected respondent was offered the choice to complete the survey with an interviewer of matching ethnic background and/or language capability (Malay, Chinese, or English-speaking interviewers).
5. **Incentives**: Respondents were offered incentive tokens with a small value in the form of either a prepaid phone reload, or postpaid phone reload. The value of the incentive was not to be communicated to respondents prior to the interview to avoid selection bias, but interviewers were permitted to disclose the value amount, if the respondents asked.
6. **Callbacks:**
    a. In Round 1, up to five callbacks were attempted for each filtered telephone number to obtain a completed interview. Therefore, the maximum number of permitted contact attempts totaled 7 (two autodialed filtering attempts and five attempts to establish contact with an eligible respondent and to complete the survey).
    b. In Round 2, up to eight callbacks were attempted for each telephone number to obtain completed interviews. Callback intervals for both rounds were half an hour for busy lines and every three+ hours for answering machines / not reached.
    c. In Round 3, up to five callbacks were attempted for each telephone number to obtain a completed interview. Callback intervals were half an hour for busy lines and every three+ hours for answering machines / not reached.
7. **Respondent substitution:** If the interviewers failed to complete an interview with the selected respondent after all callbacks were exhausted, they were not allowed to perform an in-house substitution. Each call attempt disposition would be properly entered, and interviewers would move onto the next randomly generated telephone number.
8. **Replicate structure and release protocols:**
    a. Phone numbers were sampled and delivered in subsamples called replicates.
    b. The standard of effort outlined in the Callbacks step above was made per replicate before the next was released. This procedure ensured that the

same level of effort was made for all phone numbers, diminishing the bias against households and respondents that are less likely to answer the phone on the first attempt.

9. **Weighting:** No quotas were implemented in the field; rather, the achieved sample was weighted post-field with adjustments including a base weight, non-response weight, raking and trimming, and rescaling. This is explained in full detail in the later section.

II B. FRESH SAMPLES

The sampling methodology for the new/fresh samples which were introduced in Round 3 was as follows:

1. **Simple random sample of numbers**: As in Round 1, mobile numbers were selected via simple random sampling from the sample frame through DRUID. All numbers were sampled without replacement within the survey wave
2. **Preliminary auto-dialer filtering**: Once the sample was drawn in the form of replicates (subsamples) of n=10,000, the mobile numbers were filtered using an auto-dialer to determine each number's working status.
    a. Up to two automated call attempts per number were done to determine whether the number was active or not, recording a disposition for each attempt. A minimum of two hours between each attempt was specified in the system.
    b. All numbers that yielded a working call disposition for at least one of the two filtering attempts were then passed to the CATI center human interviewing team.
3. **Respondent selection**: The devices were assumed to be personal, and therefore the person who answered was the selected respondent. Screener questions asking the age of the person who answered were used to ensure that the respondent was at least 18 years old, and within the capacity of either contributing, making or with knowledge of household finances.

4. **Respondent and interviewer matching**: Each selected respondent was offered the choice to complete the survey with an interviewer of matching ethnic background and/or language capability (Malay, Chinese, or English-speaking interviewers).
5. **Incentives**: Respondents were offered incentive tokens with a small value in the form of either a prepaid phone reload, or postpaid phone reload. The value of the incentive was not to be communicated to respondents prior to the interview to avoid selection bias, but interviewers were permitted to disclose the value amount, if the respondents asked.
6. **Callbacks**: Up to five callbacks were attempted for each filtered telephone number to obtain a completed interview. Therefore, the maximum number of permitted contact attempts totaled 7 (two autodialed filtering attempts and five attempts to establish contact with an eligible respondent and to complete the survey). Callbacks were made within the intervals of half an hour for busy lines and every three+ hours for voicemails / not reached.
7. **Respondent substitution**: If the interviewers failed to complete an interview with the selected respondent after all callbacks were exhausted, they were not allowed to perform an in-house substitution. Each call attempt disposition would be properly entered, and interviewers would move onto the next randomly generated telephone number.
8. **Replicate structure and release protocols**:
    a. Phone numbers were sampled and delivered in subsamples called replicates.
    b. The standard of effort outlined in the Callbacks step above was made per replicate before the next was released. This procedure ensured that the same level of effort was made for all phone numbers, diminishing the bias against households and respondents that are less likely to answer the phone on the first attempt.
9. **Weighting**: No quotas were implemented in the field; rather, the achieved sample was weighted post-field with adjustments including a base weight, non-response weight, raking and trimming, and rescaling. This is explained in full detail in the later section.

III WEIGHTING

A full probability-based sampling methodology was implemented. In Round 1, the complete weight was for the entire sample adjusted to 2019 population estimates from annual intercensal population projections from the Department of Statistics Malaysia (DOSM) on the designated parameters, including gender, age, state, and ethnic group. The estimates excluded the portion of the population that was not part of the target universe, particularly for age group and citizenship status. Assuming a simple random sample, with p=0.5 and n=2,210 at the 95% CI level, yields a margin of sampling error (MOE) of 2.09 percentage points. Incorporating the design effect into this estimate yields a margin of sampling error of 2.65% percentage points.

In Rounds 2 and 3, the complete weight was for the entire sample adjusted to the 2021 population estimates from DOSM's annual intercensal population projections. In Round 2, assuming a simple random sample with p=0.5 and n=1,047 at the 95% CI level, yields an MOE of 3.803 percentage point. Incorporating the design effect into this estimate yields a margin of sampling error of 3.54 percentage points.

In Round 3, among both fresh and panel samples, assuming a simple random sample with p=0.5 and n=1,113 at the 95% CI level, yields an MOE of 2.94 percentage points. Incorporating the design effect into this estimate yields a margin of sampling error of 3.34 percentage points. Among panel samples in Round 3, with p=0.5 and n=667 at the 95% CI level, yields an MOE of 3.80 percentage points. Incorporating the design effect into this estimate yields a margin of sampling error of 4.16 percentage points.

WEIGHTING ADJUSTMENTS

The weighting scheme for Round 1 was developed with the following adjustments:

1. **Base Weight:** a base weight was calculated as the inverse of the probability of a number being dialed.
2. **Non-response Weight:** a non-response weighting adjustment was performed using a weighting-class adjustment by the inverse of Response Rate 3 (as defined by AAPOR) by the sample design stratum. Phone numbers that were removed via pulsing were considered Not Eligible for this weighting adjustment.

3. **Raking and Trimming:** a post-stratification weighting adjustment was performed using the aforementioned benchmark source for national population figures. An outlier analysis of the weights was then performed, with those beyond 3 standard deviations of the mean being trimmed. This was done as an iterative process (raking, trimming, and raking again) until weights were stable within a comfortable max weight.
4. **Rescaled:** weights were also delivered in a rescaled format.

The weighting scheme for Rounds 2 and 3 were developed using adjustments in the form of raking and trimming, as well as weights rescaling.

IV ATTRITION

IV A. ATTRITION IN ROUND 2

Attrition occurred when respondents from Round 1 were not able to be re-interviewed in Round 2. The attrition rate for the HiFy survey in Round 2 was 52.6%, with 1,047 respondents who had previously participated in Round 1 successfully re-interviewed in Round 2. Most of the dropouts in Round 2 were due to unanswered calls (62% of all dropouts) or because their phones were uncontactable (14% of all dropouts). 14% of the dropouts refused to participate in the second round. A test to investigate whether attrition was random showed that the dropped respondents were not associated with gender, age group, household income groups, region, or ethnicity. However, there was a weak association between survey participation with respondents' education, as respondents having post-secondary education were more likely to participate in the Round 2 than those with lower education. While such an education bias could not be taken into account in the weight calculation due to the limited data availability in the DOSM's annual inter-censal population projections, the distribution of respondents by education levels remains comparable with the population estimates of household heads' education from the Household Income Survey (given the screening questions ensuring that respondents should be involved in household financial decisions, it is reasonable to expect that, *ceteris paribus*, those with more education are more likely to be involved in such decisions and hence participate in the survey). Further, weights were also implemented to ensure that the dropouts bore no significant impact on the representativeness of our panel samples along key demographics (Figure 1).

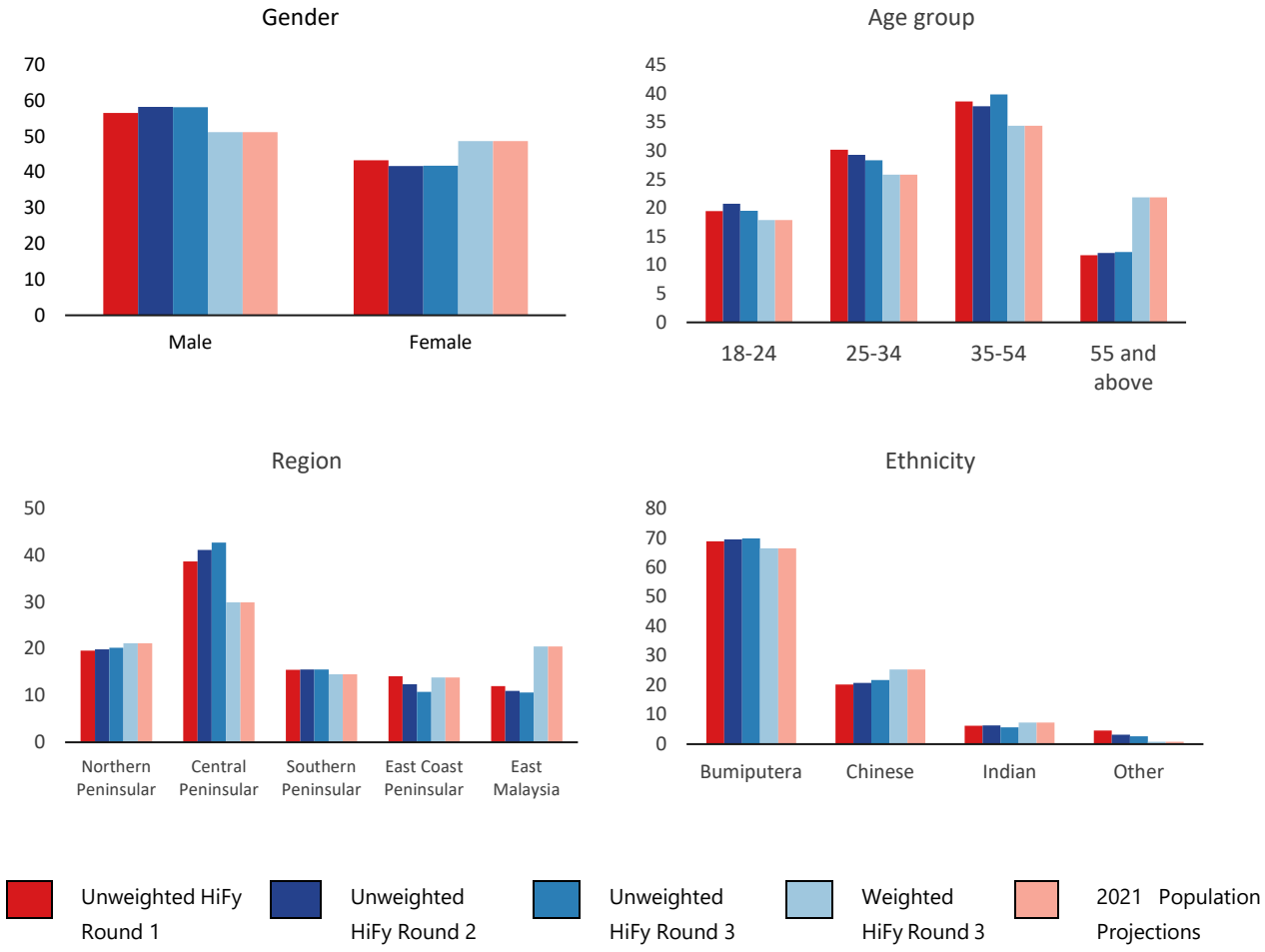
Figure 1 Key demographics in different HiFy survey rounds, unweighted and weighted versions, compared with DOSM's 2021 population projections

IV B. ATTRITION IN ROUND 3

The attrition rate in Round 3 was 36%, with 667 respondents who had been previously interviewed in Round 2 were successfully followed up in Round 3. Most of the dropouts in Round 3 were due to unanswered calls (78% of all dropouts) or because of their refusal to participate in the survey (13% of all dropouts). As in Round 2, a test to investigate whether attrition was random indicated that the dropouts were not associated with any particular gender, education levels, household income groups, region, or ethnicity. A test to investigate whether attrition was random indicated that there was a weak association

between the likelihood of dropping out in Round 3 with age group, as respondents aged 35-54 were slightly more likely to participate in the Round 3 than those who were younger. Such potential age group bias was taken into account in the longitudinal weight calculation to ensure that the dropouts bear no significant impact on the representativeness of our panel samples along key demographics (Figure 2).

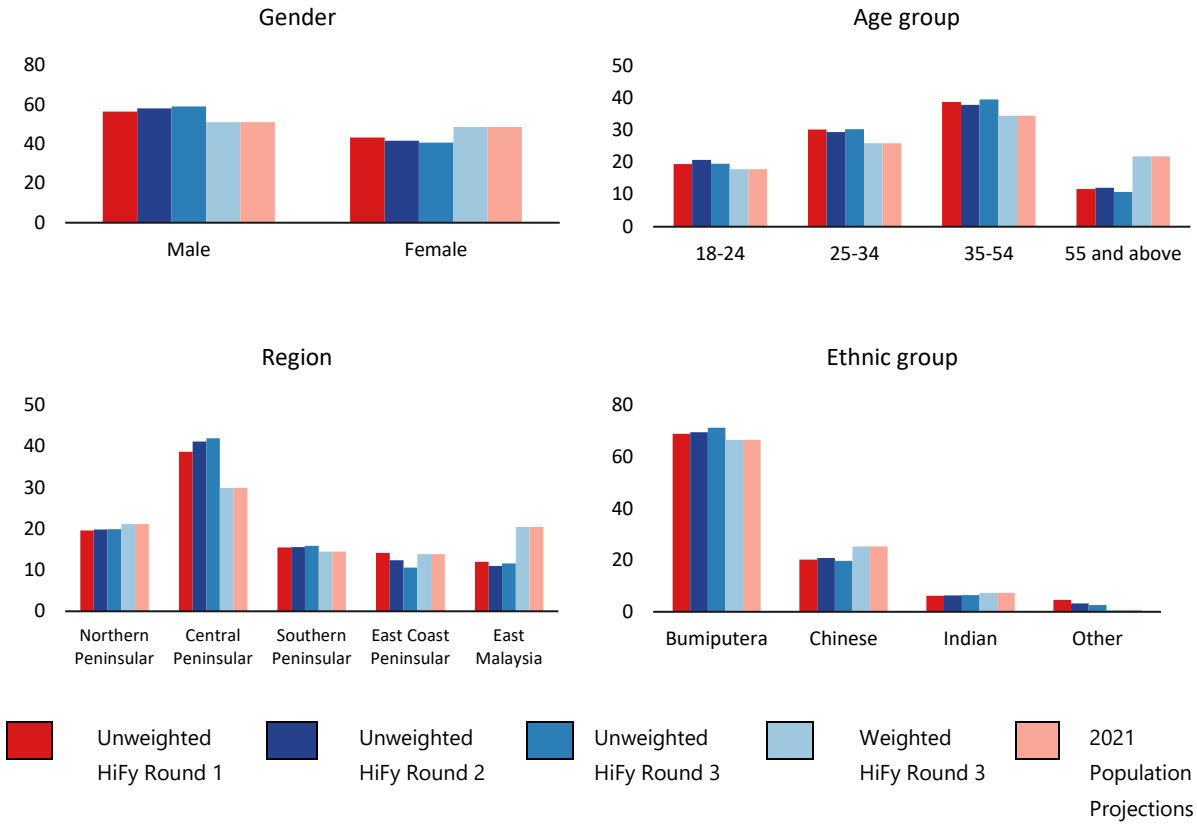
Figure 2 Key demographics among Round 3 panel respondents in different HiFy survey rounds, unweighted and weighted versions, compared with DOSM's 2021 population projections



IV C. ADDITION OF FRESH SAMPLES IN ROUND 3

As new/fresh respondents were introduced in Round 3, it is important to investigate whether these respondents have similar characteristics with those of the panel

respondents. A test was conducted for this purpose, and the results indicated that there were no significant differences between the fresh and panel respondents across gender, education levels, and regions. While fresh samples contained fewer proportions of individuals aged 55 or above and those from Chinese ethnic backgrounds, these relative demographic differences were taken into account in the cross-sectional weight calculation, ensuring the representativeness of our cross-sectional samples to the national population benchmark figures (Figure 3).



V. ADDITIONAL NOTES

Sampling Parameter:

Prior to using HiFy Round 1 data for an analysis, the following sampling set-up can be used in the Stata software:

```
gen stratif = (state * 10000) + (sex * 100) + (age_group * 10) +
ethnic_group

svyset s0_m_01 [pw=finalwgt2], strata(stratif) vce(linearized)
singleunit(certainty) clear
```

where,

- *finalwgt2* is the overall weight
- *stratif* is the combination of several sampling units used, including state, gender, age group, and ethnic group

For HiFy Round 2 data, one can implement the following sampling set-up prior to analysis:

```
gen stratif_w2 = (state * 10000) + (sex * 100) + (age_group * 10) +
ethnic_group

svyset s0_m_01 [pw=finalwgt2], strata(stratif_w2) vce(linearized)
singleunit(certainty) clear
```

where,

- *finalwgt2* is the overall weight
- *stratif_w2* is the combination of several sampling units used, including state, gender, age group, and ethnic group

For HiFy Round 3 analysis using full observations of both fresh and panel observations, one should first append both datasets (which are currently separated into *mys_r3_data_fresh.dta* and *mys_r3_data_panel.dta*) into a single dataset. Afterwards, the following sampling set-up could be implemented on the appended dataset prior to analysis:

```
gen stratif_w3 = (state * 10000) + (sex * 100) + (age_group * 10) +
ethnic_group

svyset S0_M_01 [pw=finalwgt_cs], strata(stratif_w3) vce(linearized)
singleunit(certainty) clear
```

where,

- *finalwgt_cs* is the overall cross-sectional weight
- *stratif_w3* is the combination of several sampling units used, including state, gender, age group, and ethnic group

For HiFy Round 3 analysis using only panel observations, one can use *mys_r3_data_panel.dta* and implement the following sampling set-up prior to analysis:

```
gen stratif_w3 = (state * 10000) + (sex * 100) + (age_group * 10) +
ethnic_group

svyset S0_M_01 [pw=finalwgt_panel], strata(stratif_w3) vce(linearized)
singleunit(certainty) clear
```

where,

- *`finalwgt_panel`* is the overall longitudinal weight
- *`stratif_w3`* is the combination of several sampling units used, including state, gender, age group, and ethnic group