

## Sampling Frame

The sampling frame for the Childcare Survey in Indonesia was constructed to support a stratified four-stage probability sampling design and was developed separately for each sampling unit. At the first stage, the sampling frame consisted of a complete list of all districts and cities in Indonesia, compiled using the most recent national census and auxiliary data. This frame also includes regional identifiers (Java vs. Outside Java) and information needed to construct an estimated measure of size (MOS) reflecting the number of eligible households. Districts/cities were then selected using probability proportional to estimated size (PPES). At the second stage, the village-level sampling frame was derived from the 2018 Village Potential Statistics (PODES), which provided a comprehensive listing of villages within selected districts, including their urban–rural classification and total number of households. These data formed the basis for village selection using probability proportional to size (PPS).

At the third and fourth stages, the sampling frames were constructed through field-based listing activities. For each selected village, enumerators compiled a complete list of hamlets/RW/*dusun*, along with estimates of the number of households in each, based on information obtained from village officials. One hamlet/RW/*dusun* was then selected using PPS. Within selected hamlets, the list of eligible households was compiled in the field based on information from Kader Posyandu and used for systematic random selection of 15 households per cluster.

## Sample Size

The sample size for the Childcare Survey in Indonesia was determined to support reliable estimation at the national level, as well as for broad regional disaggregation (Java and Outside Java) and by urban–rural residence. The target population was defined as *eligible households*, namely households containing at least one child aged 0–6 years and at least one woman aged 15 years or older. Based on estimates from Susenas 2021, this population was approximately 25.5 million households nationwide. Given the use of a multistage sampling design, the household sample size calculation incorporated a design effect of 2 to account for clustering, and was further adjusted for an anticipated minimum response rate of 80 percent. After evaluating multiple scenarios with different anticipated proportions and precision levels, the study selected a final target sample size of 3,000 households.

To operationalize this household sample under a cluster-based design, the study next determined the number of villages (clusters) to be selected. The number of households per village and the total number of villages were derived based on assumptions about between-village variability, expressed through a coefficient of variation. Assuming moderate heterogeneity across villages (with a coefficient of variation of approximately 20–25 percent), the design specified 15 households per village and a total of 200 villages nationwide. This clustering structure balances

statistical efficiency with fieldwork feasibility while maintaining the overall household target of 3,000 completed interviews.

Finally, the number of districts or cities included in the sample was set for operational and logistical reasons. With a fixed selection of five villages per district or city, the total number of primary sampling units was determined to be 40 districts/cities. Sample allocation across Java and Outside Java, as well as across urban and rural areas, was conducted proportionally to the estimated distribution of eligible households. Under this design, the study reports expected relative precision at the national level ranging from approximately 3.7 to 8.6 percent, with wider precision ranges for regional (4.9% - 11.4% for Java and 5.7% - 13.3% for Outside Java, respectively) and urban–rural disaggregations (4.9% - 11.4% for urban and 5.7% - 13.3% for rural areas, respectively).

## Stratification

The survey employed explicit stratification to improve the representativeness and precision of estimates across key geographic and settlement dimensions. Stratification was applied primarily at the district (primary sampling unit) and village (secondary sampling unit) stages of the sampling design. At the first stage, all districts and cities in Indonesia were grouped into two broad regional strata, Java and Outside Java, reflecting substantial differences in demographic structure, labor markets, and childcare availability across these regions. Within each region, districts were further classified by administrative type (district versus city) prior to selection. This stratification ensured balanced coverage across major regional and administrative contexts while supporting region-level inference.

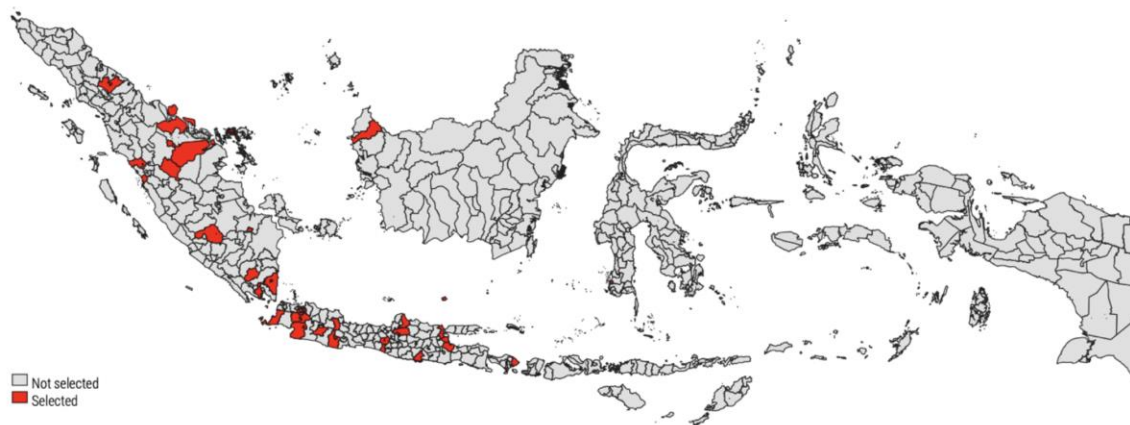
At the second stage of sampling, villages within selected districts were stratified by urban and rural status, based on classifications from the Village Potential Statistics (PODES). Village selection was conducted separately within urban and rural strata using probability proportional to size, with size measured by the number of households. This step ensured that both urban and rural populations were adequately represented in the final sample, consistent with the survey’s analytical objectives. No explicit stratification was applied at the hamlet/RW/*dusun* or household levels; instead, probability proportional to size and systematic random sampling were used at those stages. Together, the regional, administrative, and urban–rural stratification structure supports reliable disaggregation of survey estimates while maintaining a nationally representative sample of eligible households.

## Sampling Method

The survey adopted a stratified four-stage probability sampling design to produce nationally representative estimates of childcare needs, time use, and children’s wellbeing, with planned disaggregation by region (Java vs. Outside Java) and by urban–rural residence. Sampling was

conducted using probability-based methods at every stage. Districts were selected using probability proportional to estimated size (PPES), villages and hamlets/RW/dusun were selected using probability proportional to size (PPS), and eligible households were selected through systematic random sampling within sampled hamlets.

At the first stage, districts and cities served as primary sampling units and were selected using probability proportional to estimated size (PPES). Districts were stratified by region (Java and Outside Java) and administrative type (district vs. city), and their selection probabilities were based on an estimated measure of size reflecting the number of *eligible households*—defined as households with at least one child aged 0–6 years and at least one woman aged 15 years or older. This measure of size was constructed using auxiliary data to account for variation in both child presence and adult female population across districts. A total of 40 districts/cities were selected nationwide, and these districts/cities are shown by the figure below.



At the second stage, five villages were selected within each sampled district using probability proportional to size (PPS), where size was defined as the total number of households (taken from PODES). Villages were stratified by urban and rural classification prior to selection to ensure balanced representation across settlement types. At the third stage, one hamlet/RW/*dusun* was selected within each sampled village using PPS based on the number of households, with hamlet-level household counts obtained from village officials during field listing. At the fourth stage, 15 eligible households were selected within each sampled hamlet using systematic random sampling from the eligible-household list compiled in the field. In each selected household, one eligible child aged 0–6 years was selected using a Kish method. Before applying the Kish table, eligible children were listed by sex and age: female children first, followed by male children, with each group ordered from oldest to youngest.

## Weighting

Survey weights were constructed to account for the unequal probabilities of selection induced by the stratified four-stage sampling design and to ensure that survey estimates are representative of

the target population of *eligible households* in Indonesia. The final analysis weight for each responding household reflects the inverse of its overall probability of selection across all sampling stages.

Let  $h$  index strata, and let the four sampling stages be indexed in numbers as: district/city (1), village (2), hamlet/RW/*dusun* (3), and household (4). The overall probability of selection for household  $i$  in stratum  $h$  is given by the product of the conditional probabilities at each stage:

$$\pi_{hi} = \pi_{h1} \times \pi_{h2|1} \times \pi_{h3|2} \times \pi_{h4|3}$$

where:

- $\pi_{h1}$  is the probability of selecting a district/city using probability proportional to estimated size (PPES)
- $\pi_{h2|1}$  is the probability of selecting a village within a selected district using PPS
- $\pi_{h3|2}$  is the probability of selecting a hamlet/RW/*dusun* within a selected village using PPS
- $\pi_{h4|3}$  is the probability of selecting an eligible household within a selected hamlet using systematic random sampling

The sampling weight for each household is then defined as the inverse of this overall probability:

$$w_{hi} = \frac{1}{\pi_{hi}}$$

This weighting approach ensures that households from more sparsely populated or lower-probability areas appropriately represent similar households in the population. Each stage's probability components are described as follows.

At the district/city stage, selection probabilities were proportional to an estimated measure of size (MOS), defined as the estimated number of eligible households (households with at least one child aged 0–6 years and at least one woman aged 15 years or older). If  $Z_{hi}$  denotes the MOS for district  $i$  in stratum  $h$ , and  $\sum_h Z_h$  is the total MOS within a stratum, then the district-level selection probability can be expressed as:

$$\pi_{hi} = \frac{n_h Z_{hi}}{\sum_h Z_h}$$

where, in the above,  $n_h$  represents the number of districts in stratum  $h$ .

At the village stage, five villages were selected within each sampled district using PPS based on the total number of households as recorded in PODES 2018. At the hamlet/RW/*dusun* stage, one hamlet was selected per village using PPS based on household counts obtained during field listing. Finally, at the household stage, a fixed number of 15 eligible households were selected per hamlet using systematic random sampling, which gives a household-level selection probability of:

$$\pi_{h4|3} = \frac{15}{D_k}$$

where  $D_k$  denotes the total number of eligible households listed in the selected hamlet  $k$ .

The final household weight used in analysis corresponds to the sampling weight derived from the four-stage selection process. These weights are intended to be applied in all descriptive and inferential analyses to recover unbiased population-level estimates. Variance estimation should account for the complex survey design, including stratification and clustering at the district and village levels.

## Note: Setting up sampling parameter in Stata

Prior to using the survey data for analysis, users should merge the household-level weight information from `HH_Weights.dta` and declare the appropriate survey design using the following Stata command:

```
svyset adm2 [pweight=wgt], strata(strata) fpc(FPC1) vce(linearized) singleunit(centered) || village_code, fpc(FPC2) || hhid, fpc(FPC3)
```

Where:

- `adm2` is the primary sampling unit (district)
- `village_code` is the secondary sampling unit (village)
- `hhid` is the tertiary sampling unit (household)
- `wgt` is the designated weight
- `strata` is the strata for the PSU (region and urban/rural classification)
- `FPC1` is the finite population correction for PSU (number of districts in stratum- $h$ )
- `FPC2` is the finite population correction for SSU (number of villages in district- $i$ )
- `FPC3` is the finite population correction for TSU (number of households in village- $j$ )