

Using MICS, MICS Plus, and other Household Survey Data in Impact Evaluations of National/Subnational Programs and Models-to-Scale

A Guidance Note

Author: Yasemin Kisbu, PhD

July 2023

Acknowledgements

This guidance note is authored by Yasemin Kisbu, PhD (Director of Independent Evaluation Laboratory at Koç University in Istanbul Türkiye), with significant and insightful contributions and supervision of José Sierra Castillo (Evaluation Specialist, UNICEF ECARO) and Mirella Hernani (Regional Evaluation Adviser, UNICEF ECARO). We deeply thank Ashley Wax (Evaluation Specialist, UNICEF ECARO), Kathleen Letshabo (Evaluation Specialist, UNICEF HQ), and Deepak Kumar Dey, Siraj Mahmudlu, and Attila Hancioglu from the UNICEF Global MICS team for their valuable feedback and support.

Citation:

Kisbu, Y. Using MICS, MICS Plus, and other Household Survey Data in Impact Evaluations of National/Subnational Programs and Models-to-Scale: A Guidance Note. 2023, UNICEF

Disclaimer:

The contents of this note do not necessarily reflect the policies or views of UNICEF. The designations in this publication do not imply an opinion on the legal status of any country or territory, or of its authorities, or the delimitation of frontiers.

Table of Contents

Acknowledgements.....	ii
Table of Contents.....	iii
List of Figures, Tables, and Boxes	iv
List of Abbreviations	v
1. Introduction	1
1.1 Purpose of the guidance note.....	1
1.2. What impact evaluation questions can be answered using MICS?	2
2. Using MICS and other Household Survey Data for System-Level Impact Evaluations	5
2.1. Possible evaluation designs	5
2.2. Feasibility assessment.....	6
3. Using MICS Plus Process for Evaluations of Models-to-Scale	12
3.1. MICS Plus.....	12
3.2. Which evaluation activities can we use MICS Plus for?	12
3.3. Using MICS Plus for evaluating models-to-scale.....	14
4. The Role of Government.....	19
5. Concluding Remarks and Recommendations	19
ANNEXES	21
Annex A. Quasi experimental designs for evaluations using secondary household survey data	21
Annex B. Flow chart for assessing the feasibility and using MICS for an impact evaluation.....	24
Annex C. Links to MICS, MICS Plus, and example household surveys' documentation	25

List of Figures, Tables, and Boxes

Figure 1. The MICS Plus Process

Figure 2. Modeling for scale-up

Box 1: Example modules on child outcomes from UNICEF MICS

Box 2: Example household surveys

Box 3: Note on causal attribution

Box 4. Linking household survey data with data from different sources and the MICS Link Initiative

Box 5: Using MICS data to evaluate a system-level intervention at the national level

Box 6: Using the EU-SILC household survey data to evaluate a system-level intervention at the subnational level

Box 7: Key advantages of using MICS Plus for impact evaluations

Box 8: Using the MICS Plus process to evaluate a model-to-scale

List of Abbreviations

CCD	Care for Child Development
DHS	Demographic and Health Survey
ECA	Eastern Europe and Central Asia
ECD	Early Childhood Development
EHIS	The European Health Interview Survey
EU-SILC	The European Union Statistics on Income and Living Conditions
GIS	Geographic Information System
IDIA	International Development Innovation Alliance
MICS	Multiple Indicator Cluster Surveys programme
MMAPP	Measuring Mental Health Among Adolescents and Young People at the Population Level
UNICEF	United Nations Children's Fund

1. Introduction

1.1 Purpose of the guidance note

With the vision for the next-generation evaluation function for UNICEF in the Decade of Action, increased efforts are being made to strengthen emphasis in outcome- and impact-level results measurement, as well as to harmonize coordination and collaboration between the data and evidence functions. The recent “*Evaluation of Impact at UNICEF: Strategy and Action Framework*” puts further emphasis on the application of innovative approaches and tools to establish the outcome- and impact-level contributions of the programmes¹.

One valuable resource that can be better utilized as part of evaluations is the [Multiple Indicator Cluster Surveys programme](#), known as MICS. MICS generates statistically sound and internationally comparable data on key indicators on the well-being of children and women worldwide. Implemented in more than 100 low-and-middle income countries worldwide, the nationally representative survey is implemented by trained fieldwork teams conducting face-to-face interviews with household members on a variety of topics that directly affect the lives of children and women. MICS has become a major data source for policy and programming planning for many governments around the world and for monitoring of more than 30 Sustainable Development Goals indicators. Initiated in 2019, UNICEF also rolled out [MICS Plus](#), a telephone survey which uses a representative subsample from any representative household sample frame. Additionally, other national and subnational household surveys conducted by government, UNICEF, government counterparts, and other partners are also available as valued sources of data for evaluation.

MICS and other secondary household survey data can serve valuable purposes in both formative and summative evaluation activities throughout the program cycle. For example, prior to implementing an intervention, MICS can supply disaggregated baseline data, allowing for an informed planning period to better understand the needs and vulnerabilities of the targeted population. During the implementation period, it can be used to establish the effectiveness of an intervention in achieving its expected objectives. By observing the progress in indicators, it can enable a comparative analysis across different geographic regions where the intervention is implemented or not and various implementation strategies, providing early insights into the intervention's effectiveness and implementation. While MICS and other household survey data can have these multiple evaluation applications, this guidance note focuses on their role in impact evaluations, highlighting it as a cost-effective and high-quality data strategy. In the past, this data has been used in impact evaluations primarily only to inform the background/context of the evaluation rather than as the key data source. In that regard, **the main objective of this note is to provide guidance for UNICEF's evaluation community, government counterparts, and development practitioners working on child-related development topics, to be able to:**

- assess when existing MICS, MICS Plus, and household survey data can be used in impact evaluations,
- consider using MICS Plus process in impact evaluations, especially as a cost-effective strategy in evaluating models-to-scale.

¹ UNICEF Evaluation of Impact - Strategy and Action Framework 2022–2025. UNICEF, New York

This guidance note overviews (a) which evaluation questions can be addressed in impact evaluations using MICS, along with its strengths and limits, (b) the feasibility and considerations on using MICS and other household survey data in system-level impact evaluations, and (c) using MICS Plus for evaluations of models-to-scale. Throughout the guidance note, illustrative examples based on specific modules from MICS and other household surveys are provided. Finally, specific recommendations to foster the use of secondary household survey data in evaluations are provided.

1.2. What impact evaluation questions can be answered using MICS?

The evaluation questions we focus on in this note for both the system-level impact evaluations and impact evaluations of models-to-scale are the **cause-and-effect questions** that reveal the effect of an intervention at the outcome/impact levels. Examples of cause-and-effect questions are: “As a result of the learning systems strengthening, do students have better skills than they otherwise would have?”; “Did the early childhood development (ECD) program improve caregivers’ early stimulation and responsive care practices for their young children?”.

The cause-and-effect evaluation questions investigated in these impact evaluation studies can be used to satisfy the following DAC evaluation criteria:

- *Effectiveness*: To what extent were the expected changes observed as a result of the intervention? (e.g., Did the intervention improve changes in access to mental health-care services?)
- *Impact*: To what extent did the intervention produce the primary and secondary long-term effects, whether directly or indirectly? (e.g., Did the intervention cause changes in the mental health care-seeking and well-being of adolescents?)

One should note that when answering the above impact evaluation questions, the use of MICS will offer many opportunities as well as constraints. Below we list some of the strengths and limits of using MICS data in impact evaluations.

Pros

Nationally representative sample. All MICS data are based on a nationally representative sample, thus its use in the evaluation of national programs and policies would provide generalizable results. This presents a unique opportunity to achieve external validity for an evaluation, at no additional cost.

Cost effective, high quality, comparable data. While using available surveys and data in evaluations is a cost-effective strategy, MICS also has the advantage of being a reliable data source as its questionnaire development and data collection meet high quality standards. In that regard, it produces data that are highly trusted by national counterparts. In addition, as the same survey tool is used across countries, it produces comparable data across countries, which allows for benchmarking.

Continuous improvement and added capabilities to meet data demands. The continuous development of MICS involves testing new culturally adaptable and validated measures (e.g., the Measuring Mental Health Among Adolescents and Young People at the Population Level tool, MMAPP) and innovative tools to meet data needs for different purposes (e.g., MICS Plus, MICS GIS, MICS Link initiatives). For example, the recent MICS GIS initiative makes the geospatial data available, which enables the use of different evaluation designs (e.g., instrumental variables). Additionally, the MICS Link initiative allows for combining MICS data with data from other sources (e.g., admin data, facilities

survey data), which provides a range of opportunities for impact evaluations as will be described in the next section.

Cons

Limited control over the questionnaire. While there is no way to make changes to an existing (already collected) MICS data, slight alterations in the questionnaire in the form of adding new items/questions may be possible for an upcoming round of data collection. This would necessitate an understanding of the upcoming MICS timeline as adding new items to the questionnaire is not possible after a certain deadline (e.g., one year before the fieldwork). Moreover, it would require a close collaboration with the MICS team. The questions must meet the overall quality of the survey and follow MICS's quality and ethics protocols. Finally, adding (buying) questions to the survey will have some costs, commonly per question, which may be still very cost-effective if the MICS timeline matches the planned evaluation for the national program of interest.

Variables. Variables covered in MICS (see Box 1 for example child outcomes in MICS) commonly change in longer periods of time (e.g., more than a year) and are more difficult to improve as a result of small interventions. Therefore, available MICS modules may not be appropriate for outcomes changing in shorter time spans and for evaluating small scale interventions. Therefore, using MICS to evaluate national programs is more suitable.

Box 1: Example modules on child outcomes from UNICEF MICS

- **The Child Functioning Module** is used to identify difficulties (e.g., hearing, walking, learning) that a child experiences, according to a range of severity based on a rating scale. The module consists of two questionnaires, one with 16 questions for children aged 2 to 4 years and another with 24 questions for children aged 5 to 17 years. The questions are administered to the mother or primary caregiver of the focal child in question.
- **The Immunization Module** records data on whether a 0 to 2 years old child received BCG, DPT3, OPV3, measles and other immunizations. If a vaccination card or similar document is available, the interviewer fills the records based on the card; if not, information is acquired based on primary caregiver's report.
- **The Child Discipline Module** assesses whether 2 to 4 years old children have been disciplined by any adult in their household in the past month via methods such as "hitting or slapping the child", "calling the child dumb or lazy", "taking away privileges", and "explaining wrong behavior". The module is based on primary caregiver report and consists of 11 questions.
- **The Foundational Learning Skills Module** is directly administered to children 7 to 14 years old in the household by the interviewer and assesses the child's literacy and numerical skills. The reading assessment includes accuracy and comprehension tasks, and the number skills assessment includes number recognition, number discrimination, addition, and pattern recognition and completion.
- **The Early Childhood Development Module** is used to obtain information on to which extent households provide a supportive and stimulating learning environment for the child, if the child has achieved the developmental milestones and exhibits expected behavior of children of his/her age, and whether the child has been left alone or left with another child as a caregiver. The module is based on primary caregiver report for children 2 to 4 years old living in the same household.

Non periodicity. Although MICS is repeated in many of the implementation countries, it does not have specific periodicity. For instance, the survey may happen in some countries every 4-5 years or not. This limitation, partly due to lack of control over government plans, impedes the early planning and feasibility of evaluation studies using MICS data, in addition to the application of some rigorous

evaluation designs described in the next section. In Box 2, we provide examples of international household surveys that are implemented in the ECA region along with examples of child related topics they cover². Furthermore, periodic national surveys can also be located by contacting the relevant ministries and statistical agencies in the country.

Sample design. While MICS being a nationally representative survey is an advantage for the evaluation of national programs, in the case of subnational (e.g., specific provinces) or subgroup (e.g., refugees) populations, the sample may not be suitable for the planned evaluation due to the possible non-overlap with the intervention provinces which will make the evaluation study non feasible or the low prevalence of the subgroup which will lead to sample size issues. In other words, any kind of sub-setting/sub-grouping is likely to have negative implications for the required sample size. One potential remedy for this could be the adjustment in the sample of an upcoming round, which would necessitate a close collaboration between the evaluation and survey teams.

Analysis and specific domains of estimation. Analysis of the MICS data in an evaluation will bring its own complexities as the sampling design and weights should be taken into consideration during the process. Especially, in the case of specific domains of estimation (geographic divisions, subpopulations), adjustments to the weights would be necessary and may further complicate the analyses.

Box 2: Example household surveys

- **Demographic and Health Survey (DHS)** programme collects representative data on population, health, HIV, and nutrition in over 90 countries. Example child relevant topics covered by the survey are mental health (>15 years), immunization, early child development and stimulation, breastfeeding, post-natal care, and domestic violence towards women.
- **The European Union Statistics on Income and Living Conditions (EU-SILC)** survey collects household and individual level data on income, poverty, social exclusion, and living conditions. It provides two types of data: cross-sectional data over a given time and longitudinal data on individual-level changes over time, observed periodically over a 4-year period. Example relevant topics covered by the survey are people with unmet needs for medical treatment (>16 years), housing, and access to childcare services.
- **The European Health Interview Survey (EHIS)** collects representative data on health status, health care use, health determinants, and socio-economic background variables. EHIS targets population above 15 years old living in private households. The survey has currently a 6-year periodicity. Example relevant topics covered by the survey are depressive symptoms (>15 years), health care use, and healthy nutrition.

² More examples of periodic household surveys in the Central Asia region can be found in Brück, T., Esenaliev, D., Kroeger, A., Kudebayeva, A., Mirkasimov, B., & Steiner, S. (2014). Household survey data for research on well-being and behavior in Central Asia. *Journal of Comparative Economics*, 42(3), 819-835. [Link](#)

2. Using MICS and other Household Survey Data for System-Level Impact Evaluations

As MICS produces data on a nationally representative sample and the variables it covers commonly change in longer periods of time and are more likely to improve as a result of strong and sustainable programs, using MICS for system-level evaluations presents a suitable case. System-level evaluations may happen at different levels, such as country programme evaluations, thematic evaluations, and evaluations of individual national/subnational policies and programs. The evaluated (collection of) programs would commonly include system-level interventions seeking to strengthen the building blocks of a system to create lasting changes in a targeted thematic area.

While these initiatives are common in the development world, their evaluations at the outcome- and impact-level have been a challenge. The solution to this challenge has been noted to require the use and adaptation of different evaluation methodologies, from embedding experimental methods within evaluation designs to theory-based approaches. In this guidance note, we illustrate the feasibility of using existing (i.e., secondary) MICS and other household survey data to produce quantitative evidence for system-level impact evaluations. Regarding the purpose, these are considered as *summative* evaluations, aiming to meet accountability and learning needs. Regarding the timing, these evaluations can be conducted both *during implementation* (midline) and *ex-post* (endline) to establish the outcome- and impact-level evaluative evidence.

2.1. Possible evaluation designs

Available MICS/MICS Plus and household survey data can be used under different evaluation designs, if certain requirements are met. In Annex A, we provide requirements of specific quasi-experimental designs^{3,4}, which are most suitable for the use of secondary data in evaluations at outcome- and impact-level. Some of these designs are causally very weak, while others are relatively stronger and aim to establish rigorous *causal attribution* (see Box 3 for a note on causal attribution).

It is central to note that the choice of the most suitable design for the intended evaluation will depend on the evaluation questions, availability of the data, nature of the outcome variable(s), the availability of the control variables, and the implementation of the intervention. For example, if achieving causal inference is high of interest, then designs that reduce selection bias across groups through control variables may be suitable. If the outcome variables are highly variable over time, then having pre-intervention time series data may not be very helpful. Furthermore, evaluators need to be aware that the combination of different methods can sometimes provide the strongest design for the intended evaluation. For example, combining matching methods with other designs such as DiD is frequently employed in evaluation studies, to strengthen causal evidence regarding the impact of the intervention. Furthermore, while the focus of this guidance note is the use of quantitative household survey data in evaluations, qualitative data would also be preferably collected during these evaluations to complement the quantitative survey data with a more in-depth understanding of the intervention's mechanism of change. This would then result in a ***mixed-methods*** evaluation design.

³ Cook, T. D., Campbell, D. T., & Shadish, W. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston, MA: Houghton Mifflin.

⁴ Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch. 2016. Impact Evaluation in Practice. Washington, DC: Inter-American Development Bank and World Bank.

Box 3: Note on causal attribution

An impact evaluation assesses changes in the well-being of beneficiaries that can be attributed to a particular intervention. This focus on causal attribution requires us to estimate a *counterfactual*, that is, what the outcome of interest would have been for intervention participants if they had not participated in the intervention. In practice, this necessitates finding a control group to observe what would have happened in the absence of the intervention. Then the intervention group is compared to the control group with regard to the targeted outcomes/impacts. Another requirement to obtain causal estimates is to isolate/control for the potential factors other than the intervention that can also influence the targeted outcomes (i.e., confounders), and to minimize the selection bias. While all the designs mentioned in Annex A aim to estimate intervention effects, some of them are considered as more rigorous quasi-experimental methods as they consider the abovementioned requirements to achieve causal impact estimates; those methods are matching, difference-in-differences, regression discontinuity design, and comparative interrupted time-series.³

2.2. Feasibility assessment

Using MICS and any other household survey data for an impact evaluation will require some conditions to be satisfied. Specifically, a feasibility assessment with three main steps would be followed: (1) First, the evaluation team should determine what data are needed for the anticipated evaluation. Then (2) they should look for the available data in the specific country where the intervention has taken place. And finally, (3) they need to map the data needs against the available data for resolving the feasibility assessment (see Annex B for a feasibility assessment flow chart). Below, we describe specific considerations when conducting a feasibility assessment to determine whether a system-level impact evaluation can be conducted with the existing MICS and other household survey data.

Variables

For impact evaluation purposes, the evaluation team needs to establish the availability of data on the following variables: (a) outcomes/impacts targeted by the intervention, (b) intervention participation status, (c) control variables on the external factors that may also have an effect on the intervention's targeted outcomes, and (d) subgroup membership, if disaggregated effects (e.g., gender-based impact estimates) are of interest. Below are specific considerations regarding evaluation study variables.

Intervention's targeted outcomes/impacts. In the case of planning to work with secondary data, it is helpful to also generate the proxy indicators that are aligned with the main indicators (e.g., "children getting on track in the literacy-numeracy domain" for the main indicator of "children who are developmentally on track in learning").

Intervention participation status. As in most system-level interventions, if the intervention is non-exclusionary, such as community/village level interventions, then geographic indicators can be used to identify beneficiary status. A common challenge at this stage may be that the beneficiary status for the intervention may not be identified at all with the survey data at hand. One potential solution to this challenge is then to explore whether *data linking* is possible (see Box 4). For example, if the beneficiary status (e.g., recipient of the conditional cash transfer program) is available within admin data, it may then be checked whether linking the survey data to the admin data at the necessary level

of observation (e.g., individual, household) is possible. The data linking possibilities can also be investigated for important control variables and subgroup membership indicators stated below.

Control variables will provide the most critical information to obtain reliable evaluation findings, by improving causal attribution. Any data on external factors that can influence the intervention's targeted outcomes should be taken into account in the analyses to reduce the bias in conclusions regarding the impact of the intervention at hand.

Timeline

The evaluation team needs to check whether the timeline of the available MICS or other household survey is in line with the intervention's timeline. In order to get a reliable estimate of the baseline level of the intervention beneficiaries with regard to the intervention's targeted outcomes, the timing of the baseline data collection should not be too far in advance of the intervention's start. This duration may depend on to which extent the targeted outcomes are sensitive to external factors, including seasonality. Similarly, the posttest data should not be collected too far from the time where the intervention's expected benefits will take place. If the evaluation is expected to test gains in medium term outcomes, then the timing of the data collection should match the time when the medium-term outcomes are expected to be achieved as indicated by the intervention's ToC. It is recommended to set the required timing for data collection at pretest and posttest (separate timings to be determined in case of short-term, medium-term, and long-term outcomes of interest) assessments as a *time range* (e.g., baseline data collected during anytime in year 2018 if the intervention activities have started in January 2019) to allow for some flexibility when mapping against the timeline of the available data.

Another main consideration should be given to the feasibility of obtaining data at multiple time points (e.g., baseline, medium-term, long-term) from the same targeted population. This can be feasible through a few options. The first option would be checking whether the same survey, with the necessary

Box 4. Linking household survey data with data from different sources and the MICS Link Initiative

Linking datasets to achieve the data requirements for a potential evaluation can serve as a good solution when the household survey data to be used in the evaluation lacks certain variables or time points. In that sense, datasets are usually merged to serve one of the following two purposes: (1) add the data on the missing but required variables to the main household survey data that will be used for the evaluation, or (2) add more time points on the indicators required for the evaluation to create a time series. In the first case, for example, if the household survey data to be potentially used for the evaluation does not have the indicator on intervention beneficiary status, but the evaluation team knows that this data is available in admin records, then they would check whether it is possible to add that data from the admin records to the household survey data. This data linking activity can only be accomplished though if matching the observations across the admin data and household survey data is possible. For matching households, ideally, household identification codes or household phone numbers or GPS codes if included in the survey dataset would be necessary. For matching individuals within the households, the process will be straightforward if national identification numbers are available and accessible. If not, while demographics such as age and gender can be used to enable matching at the individual level, this may be very difficult to do in practice, also considering that some household members may move out over time.

In the case of using MICS data, the recent **MICS Link** initiative enables combining MICS data with data from other sources (e.g., admin data, facilities survey data) and greatly solves the potential issues making data linking impossible.

modules identified for this evaluation, has been repeatedly administered within the impact time frame of the intervention to be evaluated (e.g., having two rounds of MICS in the same country 4-years apart, which matches the timeline for the intervention's baseline and medium-term outcomes). The second option would be checking whether different data sources can be used to obtain data at separate time points (e.g., baseline and posttest); this would require these different data sources (e.g., two different household surveys) to produce data on the same indicators. For this option, *data linking* with administrative data can also be considered to create a time-series data (e.g., educational attainment data for years 2015 and 2016 coming from admin data to be linked to MICS data collected in year 2017) (see Box 4 for data linking).

Sample and level of observation

For evaluation purposes, one needs the data from program beneficiaries, and also non-beneficiaries if control group is necessary for a more rigorous impact evaluation. Therefore, during a feasibility assessment, first, the eligibility criteria for the intervention beneficiaries should be clearly defined (e.g., place of residence, gender, income level). This will not only guide the feasibility assessment in determining whether these indicators are available in the located datasets, but also whether a control group can be created.

Another important consideration is the **level of observation** needed for the evaluation. This would depend on the type of intervention and the level of targeted outcomes and impacts. Generally, the level of observation is required at the individual (e.g., child learning outcomes), household (e.g., housing quality), community/village, or facility (e.g., health facility infrastructure) levels. This presents another advantage of using secondary household survey data in evaluations, as most survey data provide information at the individual and household levels. It should be noted that, if for the anticipated evaluation using household survey data, data at the facility level is also required (e.g., for matching of intervention and control villages with similar educational infrastructure), facility level datasets, if they already exist or can be primarily collected, may be linked to the household survey data. Similarly, if community level data is also required for the anticipated evaluation, a search for this type of data can be conducted; yet in this case it should be noted that some community-level indicators can also be computed using the existing household survey data (e.g., provincial poverty rate that can be computed using specific indicators in the household survey if the survey had a sample representative of the provinces of interest).

Most national/subnational household surveys, including MICS, use representative samples of the population. Regarding the feasibility assessment on using a household survey for evaluation purposes, **one of the most critical steps is to investigate whether the sampling frame of the identified household survey data adequately overlaps with the targeted sample for the evaluation study.** Below, we present two typical situations for such evaluations, in relation to the sampling feasibility.

Under the first scenario, the intervention is non-exclusionary for a certain geographic area (i.e., all household or individuals in that geographic area can benefit from the intervention). For example, if the intervention is at the subnational level based on a geographic area covering five provinces, in order to use the identified household survey data for the evaluation, its sampling frame should include those provinces.

A few limitations may arise under this typical scenario: (1) First, the sampling frame of the household survey may only include three of the five provinces where the intervention has taken place. In that case, the evaluation team should assess whether conducting the evaluation based on those three provinces would provide sufficient evidence for the intervention's effectiveness (i.e., representing the

intervention population). (2) Second, the sample size achieved in those provinces by the household survey may not be adequate to detect impact. If the sample size is not sufficient to achieve required statistical power to detect impacts, then it is not recommended to move forward with that household survey to conduct the evaluation. (3) Another sample size and statistical power related limitation is that, if the intervention focuses on a low-prevalence group such as children with disabilities, it may not be feasible to use the household survey data as the sample size for the targeted group may be very low in such a survey representative of the general population. At this stage, again, the evaluation team may decide that the evaluation of this intervention with the identified household survey data is not feasible because of the lack of statistical power in producing reliable impact findings.

Under the second scenario, the intervention only benefits those who meet certain eligibility criteria. For example, consider a system-level intervention that has been operational at the national level, but only targets refugee families with school-aged children. In this case, assuming that the indicator(s) on the eligibility criteria is available in the current household survey data or through data linking, the household survey with a representative sample at the national level would be suitable for the evaluation of this intervention. However, one major limitation that may arise under this scenario is again related to the sample size. If the sample size for this subgroup is too low in the household survey data, then it will not be possible to conduct the evaluation with this data. Nevertheless, one should also note that, in the case of planning to use an upcoming round of MICS for an evaluation study, the evaluation function can also consider the possibility of influencing the survey sample (e.g., by paying for larger sample sizes in some domains); this will require early planning and close consultation about its feasibility with the MICS team though.

Box 5: Using MICS data to evaluate a system-level intervention at the national level

Background:

In this hypothetical example, Turkmenistan Ministry of Education, in partnership with the UNICEF CO, has taken an initiative to strengthen the learning assessment and evaluation system to improve the learning outcomes of students, especially for students in early grades. The system-level intervention aims to (a) establish a well-functioning assessment system at the classroom level, (b) ensure that a stable national country-wide standardized examination across all grades is in place, and (c) warrant country participation in international learning assessments. Furthermore, the intervention aims to ensure that students with disabilities take part in the national learning assessment by establishing the necessary regulations.

Brief feasibility assessment for conducting an evaluation using secondary household survey data:

Step I. Identify data needs:

Evaluation questions: An evaluation is aimed to be conducted to answer the following questions:

EQ 1. Is there an improvement in foundational learning skills of primary school children since the start of the learning assessment and evaluation system strengthening initiative?

EQ 2. From an equity perspective, is the improvement in foundational learning skills of primary school children similar for children with or without disability?

Variables: Outcome variable is the students' learning skills. Targeted beneficiaries are students attending primary, middle, and high schools. The subgroup variable is the disability status of the student.

Timeline: The intervention implementation started in year 2016 and medium-term outcomes are expected by 2019.

Sample and level of observation: Sample representing all students enrolled in elementary, secondary, and high schools in the country. Individual-level data is needed.

Step II. Determine data availability:

The country participated in MICS in the years 2015 and 2019; the data is available and accessible.

Step III. Assess whether conditions are met to use the available household survey data in the anticipated evaluation:

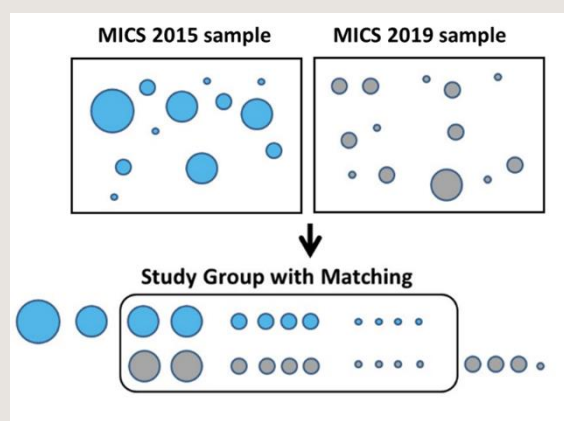
Timing of the survey data: MICS data collected in year 2015 can be used as pretest and MICS data from year 2019 can be used as posttest for the targeted medium-term outcomes.

Availability of required variables in the survey dataset: MICS data included the measurement of foundational learning skills (i.e., initiative's targeted outcome), functional difficulties (i.e., disability status), and child and family background characteristics at the individual level of observation (i.e., control/potential matching variables).

Sample: In both waves, MICS was implemented with country-representative samples.

A suggested evaluation study design:

A repeated cross-sectional design with matching can be chosen for this evaluation. Under this design, the primary school children's foundational learning skills in reading and numeracy on average (per grades) will be compared across years 2015 (i.e., before the start of the initiative) and 2019 (i.e., 3 years into the intervention). The samples from both waves will be first matched on observed background characteristics (e.g., demographic, socio-economic) to ensure comparable groups regarding potential confounders that may also influence children's learning outcomes (e.g., family background characteristics, province, age, sex, grade). Findings will also be reported across children with and without disability, using the functional difficulties module, to investigate the second EQ.



Assumptions and limitations: Limited level of control for preexisting factors causing changes over time and population differences. Assumes successful matching to attain comparable samples over time and a sample size (for the matched final evaluation sample) that attains sufficient statistical power to detect impacts.

Box 6: Using the EU-SILC household survey data to evaluate a system-level intervention at the subnational level

Background:

In this hypothetical example, to ensure access of refugee children and young people (CYP) to health services provided by the Bulgaria health system, the Bulgaria Ministry of Health adopted a health system strengthening approach. An agreement with the National Health Insurance Company is signed in year 2020 to reimburse the medical expenses for outpatient and inpatient services provided to refugee CYP aged 0-24 years old. The package of health services provided to refugee CYP under this insurance covers all primary health care facilities. As part of the initiative, a refugee CYP Health Tracking System has also been activated. The initiative has been piloted in the province of Sofia and 6 other pilot provinces with high refugee population. The medium-term outcome is increased rates of refugee CYP receiving medical treatment when needed and the long-term outcome is improved health of refugee CYP.

Brief feasibility assessment for conducting an evaluation using secondary household survey data:

Step I. Identify data needs:

Evaluation question: An evaluation is aimed to be conducted to answer the following evaluation question regarding the intervention's medium-term outcome:

EQ: Did the intervention improved the rates of refugee CYP receiving medical treatment when needed?

Variables: Outcome variable is receiving a medical treatment when needed. Targeted beneficiaries are 0-24 years old refugee CYP living in Sofia and 6 other pilot provinces.

Timeline: The intervention implementation started in year 2020 and medium-term outcomes are expected by 2021.

Sample and level of observation: Refugee CYP living in the pilot region. Individual-level data is needed.

Step II. Determine data availability:

The country participates yearly in the EU-SILC household survey; the data is available and accessible.

Step III. Assess whether conditions are met to use the available household survey data in the anticipated evaluation:

Timing: EU-SILC data collected in year 2021 can be used as posttest data for the targeted medium-term outcome.

Availability of required variables in the survey dataset: EU-SILC data includes indicators on unmet needs for medical treatment for individuals over 16 years old (i.e., proxy outcome variable), age and refugee status (i.e., beneficiary status variables), and household and individual background characteristics (i.e., control/potential matching variables). Since the EU-SILC data provides data on the "unmet need for individuals" for individuals over 16 years old, the evaluation study can only test impacts for the overlapping age group of 16-24 years old CYP.

Sample: EU-SILC is implemented at the country level with country-representative samples and the sampling frame covers Sofia and 5 of the 6 pilot intervention provinces. Therefore, the evaluation study can only use those overlapping provinces (i.e., Sofia and 5 pilot provinces) for the intervention group.

A suggested evaluation study design:

Two-group posttest only design, with matching can be used for the evaluation. Only the 16-24 years old CYP's unmet needs for medical treatment at posttest will be compared across the intervention (i.e., Sofia and 5 other pilot provinces) and control groups (i.e., non-pilot provinces). The samples from both groups will first be matched on observed background characteristics (e.g., individual and family demographics, economic status) to ensure comparable groups regarding potential confounders that may influence unmet medical treatment needs.

A second alternative for creating a matched control group is matching at the province-level rather than individual-level, which requires determining comparison provinces that are similar to the intervention provinces on several characteristics such as health infrastructure and overall and refugee population.

Assumptions and limitations: Limited level of control for preexisting differences across the intervention and control populations. Assumes successful matching to attain comparable samples. Study is limited to the 16-24 years old CYP. As the sample is at the subnational level, a good representativeness of the targeted population would not be achieved.

3. Using MICS Plus Process for Evaluations of Models-to-Scale

3.1. MICS Plus

MICS Plus, an initiative of the MICS programme, serves as a tool to collect longitudinal data on the situation of children and their families and households. For its implementation, a representative subsample from any representative household sample frame, including a register or a completed household survey such as MICS, is selected. Longitudinal data collection then consists of regular (e.g., 2-3 times per year) interviews conducted over the phone with the same households, over a period of one year or more. Operationally, MICS Plus requires setting up a digital data collection system including a call center and a centralized data management system. MICS Plus can be used for collecting specific program indicators for monitoring purposes, opinion polling, obtaining real-time data for crisis monitoring, and for questionnaire testing.

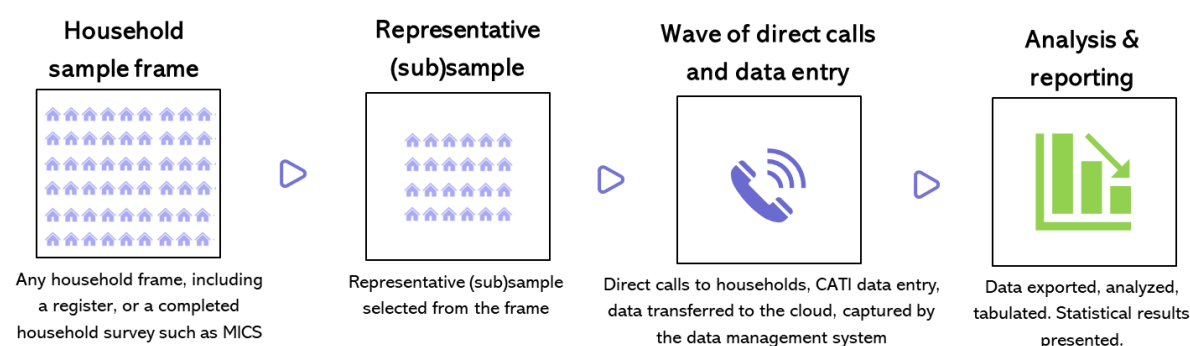


Figure 1. The MICS Plus Process⁵

3.2. Which evaluation activities can we use MICS Plus for?

While MICS Plus has been developed mainly as a monitoring tool, it can also be used for evaluation activities. In that regard, using MICS Plus in evaluation studies can satisfy the following **DAC evaluation criteria**: The *Relevance* criterion answering the question of “To what extent the intervention is suited to the priorities and needs of the target group?”; The *Effectiveness* criterion answering the question of “To what extent were the objectives of the intervention achieved?”; and the *Impact* criterion answering the question of “What difference has the intervention made to the beneficiaries?”.

In the case of impact evaluations, as it is the focus of this guidance note, MICS Plus can have several uses, as we describe below:

Secondary data source for outcome/impact evaluations. The available MICS Plus data can be used as secondary survey data in impact evaluations, in which all the considerations, designs (e.g., difference-

⁵ Figure source: Innovations in Household Surveys: MICS PLUS as a Longitudinal Real-time Data Collection Tool. Hancioglu & Karaulac, February 2021, presented at UN Statistical Commission Side Event.

in-differences, matching, pretest-posttest) and requirements listed in the previous Section 2 and Annex A would be applicable.

Testing new scales and questions for outcome/impact evaluations. One bottleneck for outcome/impact evaluations is the availability of valid and reliable scales to measure the targeted outcomes of the intervention of interest. In the case of non-availability of established scales/items that suit the intervention outcomes, the development and testing of new items can be a costly and time-consuming process. The MICS Plus process can become an efficient solution for this evaluation related activity. In close collaboration with the MICS Plus team, the scales/items to be tested can be added to one of the waves of an ongoing MICS Plus and thus quickly be tested in a quality sample, without much extra cost. As the MICS Plus survey relies on a representative sample, it can even allow for collecting population norm data for some inventories.

Primary data collection for outcome/impact evaluations.

The MICS Plus process can also be planned ahead to provide primary data for an upcoming evaluation. This approach, called as *piggybacking*⁶ on planned data collections, builds on adding new questions or modules to a scheduled survey questionnaire. These added questions will aim for the collection of required variables listed in section 2, in line with the evaluation questions to be answered. While piggybacking is an efficient possibility, evaluation teams can also design their own MICS Plus sample for an upcoming evaluation, especially in the case of an intervention targeting specific geographic areas or groups for which we would like to achieve adequate sample sizes, still at a low cost. It should be noted that this would require early planning and close consultations with the MICS Plus and Monitoring, Evaluation and Planning teams.

Box 7. Key advantages of using MICS Plus for impact evaluations

- Longitudinal nature of data, allowing to assess change over time
- Flexibility in the selection of topics to be covered based on the data needs (e.g., non-MICS modules can be included, in addition to established MICS modules)
- A representative sample at the national/subnational level (e.g., random selection of n households from each of the recently collected MICS clusters, systematic random selection in each stratum from a population and household register database), with a coverage of urban/rural areas
- Being low cost (e.g., phone-based data collection)
- Being able to generate real-time results, if needed for the evaluation.

Using MICS Plus process for impact evaluations is also quite useful and resource-efficient for the testing of **models-to-scale**, which may not yet be ready for a large evaluation investment. While MICS Plus is not a tool for testing early-stage pilots, it is suitable to test models that are already backed up by some evidence but may not yet be ready for scale-up. Imagine a model that is being tested in a subregion of a country, and the planned MICS plus includes that region in its sampling frame. In that case some modules specific to the targeted outcomes of the model being tested can be integrated into a MICS Plus process. While all considerations for primary data collection for an evaluation would apply here, there would be further considerations specific to the MICS Plus process. Below we address these concerns in addition to the different uses of the MICS Plus process in the case of evaluating models-to-scale.

⁶ Wassenich, P. (2007) Data for Impact Evaluation. Doing Impact Evaluation Series no. 6. Washington DC: World Bank.

3.3. Using MICS Plus for evaluating models-to-scale

Scaling is defined by The International Development Innovation Alliance (IDIA) as “to support the widespread replication and adaptation of an innovation across large geographies and populations for transformational impact.” By adopting this definition, UNICEF has been working on a process-based approach for models to scale, in which a country office and its government counterparts (a) identify an intervention with the highest probability to succeed at scale, then (b) implement the key intervention

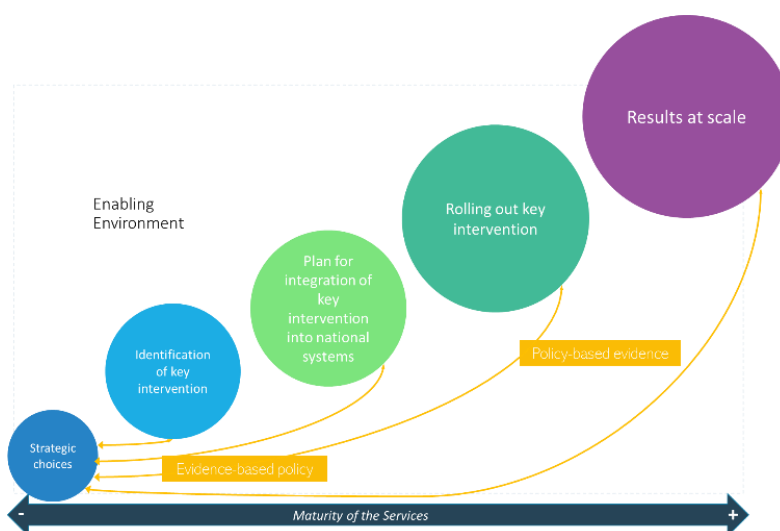


Figure 2. Modeling for scale-up⁷

(modelling phase) at a low-scale (e.g., subnational), and finally (c) scale-up the successful intervention (Figure 2)⁷. This approach requires evaluating the model to be scaled in order to validate its success where it was employed and avoid false positives at low scale. Testing of the models may involve low-cost experimental or quasi-experimental evaluations. **In that regard, using the MICS Plus process offers a cost-effective and high-quality data opportunity to assess the effectiveness/impact of the model intervention, with the intention of expanding its implementation, thereby facilitating large-scale testing.**

In addition to the advantages of using MICS Plus for evaluations that we mentioned in the previous subsection (e.g., testing new items/scales), in the case of testing models-to-scale, MICS Plus offers other possible solutions. In particular, the MICS Plus process can be further helpful for the following:

- (1) **Testing the stages of a ToC.** The MICS Plus process will allow us to obtain the early-stage results for the intervention of interest at a low cost. This will give us the opportunity to test whether the intervention has changed the targeted intermediate outcomes, and thus to corroborate the ToC of the model-to-scale. Imagine a model-to-scale which integrates the Care for Child Development (CCD) package into existing health and education service delivery platforms in a country, and the package aims to strengthen caregivers’ responsive care and early stimulation practices for their young children via increasing their parenting knowledge and skills (i.e., the medium-term outcomes). The MICS Plus process can test whether the intervention was successful in changing parental knowledge and skills by incorporating the relevant items/scales for these outcomes (See Box 8 for this example). Commonly, in a ToC, the path from the intervention activities to the medium-term outcomes represents the actions taken toward the realization of these intermediate outcomes (i.e., the Action Theory path). And then, in theory, it is expected that the change in the medium-term outcomes will

⁷ From Models to Scale: Managing models to give them the best chance to scale up successfully. 2022, UNICEF. Figure directly taken from this source.

lead to a change in the long-term outcomes (i.e., the Conceptual Theory path)⁸. The use of MICS Plus to test the Action Theory of the ToC will inform the project team whether the intervention activities need to be revised before the scale-up.

- (2) Testing the assumptions of a ToC.** The MICS Plus process can also allow us to test the assumptions of the intervention ToC. In that regard, MICS Plus will provide essential data for a pivotal aspect of impact evaluation. If assumptions are not met, even perfect implementation of all other aspects of the intervention will not lead to the desired impact.
- (3) Testing different interventions across domains.** Models-to-scale can involve different (variations of) interventions to improve outcomes across different domains. For example, an ECD focused model may include different combinations of interventions targeting the four early developmental domains (i.e., language-cognitive, physical, social-emotional, and learning). The model implementation and the MICS Plus process can then be planned and designed in a way to test which of those different interventions or their combinations are the most successful across domains. This kind of an evaluation design will grant valuable information to review the current model and lessons learned for future similar models. One should also note that the same strategy can also be used to comparatively test different (variations of) interventions across time and intervention assignment rules.
- (4) Using the survey's sample as a frame for designing interventions.** MICS Plus sample relies on a high-quality representative sample, which is commonly the last round of MICS in that country. Thus, for evaluation purposes, it provides a unique opportunity to rely on that sample for designing and implementing the interventions in accordance with an evaluation design plan. This will enable more rigorous impact evaluations through the early planning of control and (multiple) intervention groups (e.g., randomized at province level).

Special considerations when designing and conducting an impact evaluation of a model-to-scale with MICS Plus

Sample. The same considerations for the sample mentioned in Section 2 would apply here as the MICS Plus process is bound by an existing sampling frame such as the one adopted by the most recent MICS round. In the case of using the piggybacking approach, the evaluation team would be checking whether the original sampling frame from which a subsample of households will be selected for the implementation of MICS Plus includes the territory (e.g., provinces) of the model intervention. In the case of partial overlap between the intervention and the MICS Plus implementation provinces, if the overlapping provinces would provide sufficient evidence on the effectiveness/impact of the model intervention, the evaluation study can then be adapted/revised accordingly. It should be also noted that in the case of using this subsample, there may be more emphasis on the problem of acquiring adequate data if minority groups are targeted by the intervention, due to the already low prevalence of this group in the general population.

Another concern regarding the sample may arise from the phone-based data collection in MICS Plus. Evaluators should take into account that this may lead to the exclusion of the most vulnerable groups, who may be the target of the model intervention at hand, as phone numbers for the selected households would be required. One potential remedy in this situation can be to distribute phones to

⁸ See Kisbu, Y., MacKinnon, D. P., & O'Rourke, H. P. (2015). Statistical models of mediation for drug program evaluation, for an example of how Action and Conceptual Theories are represented in a ToC and can be analyzed separately and together. [Link](#)

households without phone numbers, if resources are available. Furthermore, the MICS sample can be used to check and validate the subsample representativeness.

Finally, ideally, following the MICS Plus design, the survey items would be answered by any respondent in the household. However, if there is a need to focus on one respondent (e.g., mother or main caregiver of child), then this specific choice of respondent may potentially lead to higher attrition over time. However, for evaluation purposes, keeping the attrition at a tolerable rate is essential. At the same time, this choice of respondent would still depend on the nature of the outcome variables of interest. Therefore, in the case of reliable information on the outcome indicator can only be obtained from the main caregiver in the household, the evaluation team should be aware that this may risk having higher attrition rates; the risk getting amplified as the number of waves to be used for evaluation purposes increases.

Variables and measurement. One important advantage of the MICS Plus process is that the survey is not limited to existing modules of MICS, and questionnaires on new topics of interest may be added to the survey. For example, if the model intervention has a focus on attitudes toward child marriage, a set of questions to measure this variable can be added to the survey with the guidance and support from the global MICS team. This flexibility regarding the questionnaire also helps in measuring other key variables of interest for the evaluation, such as beneficiary status (e.g., usage of services provided as part of the model intervention), subgroup membership and control variables.

The phone-based data collection in MICS Plus may bring important considerations regarding which indicators, how reliably, and how many of them can be measured in this survey. It should be noted by the evaluation team that if the outcomes of interest require test and direct observation based assessment methods (e.g., assessing early numeracy skills of children living in the household through a test directly administered to the child) and if other means of assessment (e.g., parent report) are not reliable for measuring this indicator, then it would not be possible to obtain data on those outcomes in this evaluation through the use of MICS Plus. The bias of measurement in this case may also arise from the potential respondent bias, as in most cases the knowledgeable adult household member is selected for the interview in MICS Plus, who may not necessarily be the main caregiver of the child. The second limitation would be the limited time for the phone-based interviews, which would restrict the number of questions to be covered in the questionnaire. The duration of a phone interview in MICS Plus typically takes 5-15 minutes on average per household. One flexibility of the MICS Plus process that alleviates this difficulty though is that it allows for including/excluding some modules at different data collection waves and also shortening existing MICS modules for use in MICS Plus. A third limitation is the privacy issues that would prevent asking some questions on the phone due to ethical concerns, or if asked may lead to inaccurate responses given by the respondent.

Timeline of data collection. While the modules of interest can be included in all waves of MICS Plus, it is also possible to include only some selected modules on specific waves, which also allows for mixing variables across different time frames. This gives great flexibility to the evaluation team in determining the evaluation design. For example, it is possible to only include the relevant modules at two specific waves of data collection, which correspond to before and after intervention implementation; this will create a pretest-posttest evaluation design (with or without a non-equivalent control group design options). A third wave can be used for obtaining follow-up data as well. Or it is also possible to create a time series by collecting data at multiple data points before the intervention and after the implementation; this will allow for an interrupted time-series design (with or without control group), a more rigorous design regarding causal inference. Nevertheless, the choice of the number of data

points will still primarily depend on the timeline of the intervention and its' ToC regarding when the changes in outcomes/impacts are expected to happen. Considering that the MICS Plus data collection timeline is recommended to be up to 12 months (although it can be sometimes longer depending on the needs), for outcomes that are expected beyond this duration, data linking options can be considered for obtaining data on the same indicators beyond the MICS Plus timeline.

One operational consideration under the study timeline would be that the MICS Plus data collection process may create irregular intervals between the waves selected to benefit the evaluation study and should be recognized by the evaluation team. This mostly may happen due to the substitution process for households non-responding to the phone calls. MICS Plus employs a rigorous model-based substitution process to handle non-response in order to ensure both the representativeness of the subsample and the targeted sample sizes⁹. For instance, this process, leading to the calling of more households during the first wave may double the duration of the first wave data collection (e.g., four weeks of implementation rather than two weeks) as compared to the second wave.

⁹ More details on the substitution can be found at the following MICS Plus link: <https://mics.unicef.org/mics-plus/methodology-and-use>

Box 8: Using the MICS Plus process to evaluate a model-to-scale

Background:

In this hypothetical example, following the strong national demand for supporting early childhood development (ECD) and implementing quality parenting programs, the government of Georgia and UNICEF Georgia CO implement a model to integrate Care for Child Development (CCD) package into existing health and education service delivery platforms. The package aims to strengthen caregivers' responsive care and early stimulation for their young children. The model comprises CCD Training of Trainers sessions for ECD multi-sectoral teams representing different ministries; CCD training (by the ECD multi-sectoral teams) to frontline service providers, including early childhood education teachers, nurses, medical officers, community health workers, nutritionists and social welfare officers, and provision and adaptation of job aids (including CCD counselling cards) to support service delivery. After a piloting stage at low scale, the model will be considered for future scale up at the national level. As stated in the intervention's ToC, the medium-term outcome is caregivers' parenting knowledge and skills, and the long-term outcome is improved caregiver responsive care and early stimulation practices. Changes in medium-term outcomes are expected after the six months into the implementation.

A planned MICS Plus data collection in Georgia:

The same year, UNICEF Georgia CO plans to implement a MICS Plus data collection for the purpose of post-COVID19 situation monitoring across different areas such as education and housing.

Considerations for integrating the evaluation of the piloted model into the planned MICS Plus process:

The evaluation team would like to benefit from the planned MICS Plus process to evaluate the new CCD model in improving caregivers' responsive care and early stimulation practices. They take the following considerations into account and take actions:

Sample: The evaluation team checks whether the planned MICS Plus sampling frame based on the latest round of MICS in Georgia, overlaps with the intervention pilot region and if yes, they inquire about the planned sample size in the pilot region. As they learn that the intended sample can be achieved through the MICS Plus process, they next evaluate whether the targeted outcomes can be measured via the phone-based survey.

Measurement: The evaluation team adds a new module on parenting knowledge and skills, considering that phone-based data collection is feasible to collect data on these items.

Timeline: The MICS Plus data collection is planned to start in January 2022 and end in December 2022 with 4 quarterly waves. Considering the timeline of the intervention, the evaluation team decides to integrate the relevant module for this evaluation into the first and last waves to collect pretest and posttest data.

Evaluation study design: The evaluation team decides to adopt a difference-in-differences (DiD) study design for this evaluation. Under this design, they will compare the change in caregiver responsive care and early stimulation practices over time (i.e., from February to December 2022) between the MICS Plus sample from the provinces where the model intervention was implemented and the provinces where the model intervention was not implemented.

Assumptions and limitations: This design controls for changes over time that affect both populations, but not for potential differences in preexisting trends across the groups. In case the sample size for beneficiaries is small, another design strategy may be to compare treatment and control provinces.

4. The Role of Government

So far, the document has mostly focused on the technical aspects of using MICS and other household survey data for impact evaluation. However, one should also note that, commonly, MICS and most of other nationally representative complex household surveys tend to be government owned. Therefore, even in the case of an evaluation led by a non-governmental agency, the government would have an important role in the process. As such, the evaluation team should be aware of the political complexities of the process and plan accordingly.

On the other hand, the use of MICS also creates opportunities for national capacity development. In most cases, it is likely that the survey owner is a different agency than the program owner and the evaluation unit (within or outside of the government). By participating in the process, survey owners will better understand evaluation data demands and will be seen as valuable partners by other government agencies. The program owners, on the other hand, will gain a better understanding of national data plans and how best to use them. This easy flow of information between evaluation and data functions within government is key to a strengthened national evaluation capacity and to properly scale up national programs.

5. Concluding Remarks and Recommendations

Household surveys can be a good source of data for evaluations, if certain requirements are met depending on the type of evaluation and evaluation questions of interest. Yet, the first requisite for such use of this survey data in evaluations is its availability. Therefore, **this guidance note further highlights the necessity for countries to invest in good quality, population representative household surveys such as MICS.**

When such survey data is available, further *planning* efforts are necessary to ensure it benefits the evidence needs. Below, we make specific recommendations:

- Mapping of existing and/or upcoming, especially periodic, household survey data in the country is a valuable common exercise. It is recommended to make sure that the upcoming evaluations consider these data plans and assess if the data are suitable for the evaluation needs.
- An important point for the use of household survey data in evaluations is its consideration during the planning stage of a new programme. Preidentified actions, making sure that available household survey data are also considered for use in future evaluations, are highly recommended.
- Using an upcoming round of MICS for an evaluation will require to coordinate with many parties such as the MICS team and the government counterparts. Engaging in early is crucial to capitalize on the cost-effective nature of this data strategy instead of proceeding with a new data collection activity.
- To properly use MICS Plus, when evaluating models-to-scale, early planning activities which include a thorough feasibility assessment regarding the sampling frame and choice of respondents, in close consultation with the MICS team are recommended.

Additionally, the use of MICS for evaluations creates synergies between those utilizing data and those generating it. The former allows data users to improve their evaluation capacities and gain a better understanding of national data plans and how to shape them. The latter fills a new niche in the

country's data landscape and allows data producers to adjust their operations to meet the growing demand. This demand-supply cycle, if adequately supported, can lead to an overall improvement in both national evaluation and data capacities.

ANNEXES

Annex A. Quasi experimental designs for evaluations using secondary household survey data

One-group pretest-posttest. This design obtains one pretest (i.e., before the intervention) and one posttest (i.e., after the intervention) observation of the outcome variable(s) from the same participants. It allows for investigating change over time (i.e., posttest minus the pretest scores on the outcome variables) but cannot make a causal attribution to the intervention as alternative explanations for the change cannot be ruled out.

Different samples pretest-posttest (repeated cross-sectional). This is similar to the abovementioned pretest-posttest design; however, the pretest and posttest observations are not obtained from the same respondents. For example, under this design, to evaluate a system-level intervention that was implemented in year 2018 with impacts expected to occur by year 2021, the evaluator uses the MICS 2017 round data as pretest and MICS 2021 round as posttest data, where the two rounds observed different representative samples from the targeted population. As this design does not control for population differences and other external factors that may also cause changes over time, it does not provide a rigorous test of causal hypotheses.

Two-group posttest only, with matching. This design obtains only posttest data from two groups: the intervention group and the control group who was not assigned to get the intervention and then computes the difference between their scores on the targeted outcome variable(s). To achieve intervention and control groups that are similar at baseline, groups are matched on pre-intervention or time-invariant variables.

Two-group pretest-posttest (Difference-in-differences, DiD). This design obtains data at both pretest and posttest and from intervention and control groups. To estimate impact, the change from pretest to posttest for the control group is subtracted from the change from pretest to posttest for the intervention group; this allows testing which group improved regarding the intervention's targeted outcomes. This is a frequently employed quasi-experimental design with intermediate control for confounding due to having both a pretest and a control group. However, the intervention and control groups can differ at baseline (i.e., selection bias may occur), which may provide alternative explanations for the obtained DiD impact estimates. Causality can be claimed if the assumption that all unobserved differences between groups do not vary over time; in that regards MICS rich data set provides many covariates. Additionally, one generalization of the DiD method is the *synthetic control* design, which uses a weighted combination of groups as control.

Regression discontinuity design (RDD). RDD requires posttest data from participants under the intervention and control groups. Under this design, participants are assigned to their conditions on the basis of a cutoff score on an assignment variable. For example, imagine that households with monthly income (i.e., the assignment variable) under a certain cutoff score (e.g., 400 USD per month) get assigned to the cash transfer intervention (i.e., the intervention group). And households with income above the cutoff score do not receive the cash transfer (i.e., the control group). By incorporating the assignment variable into the analysis, RDD achieves causal impact estimates comparing observations on each side of the cutoff. RDD has also variants such as comparative RDD,

which overcomes the main limitations of conventional RDD by adding data on a comparison function, in the case of large sample sizes¹⁰.

Comparative interrupted time series. The simple interrupted time series (ITS) design requires an intervention group with several observations on the outcome variable before and after the intervention. The comparative ITS (CITS) design adds a control group time series to the simple ITS. It is a strong design that is able to test cause-and-effect evaluation questions and capture both immediate and delayed effects. However, when implementing CITS, the evaluators should carefully consider other interventions/events that can affect the treatment and control groups differently through the time span of the observations.

It should be noted that all quasi-experimental designs have specific **assumptions** that should be tested by the evaluator. For example, propensity score matching relies on the assumption that intervention and control groups are comparable, if data on covariates that are related to both the outcome and the intervention exposure are available and included in the propensity score model. If DiD is used without matching, then the assumption of parallel trends, i.e., trend in outcomes in intervention and control groups was similar before the intervention implementation, need to be satisfied. We refer the reader to a technical document on causal impact evaluation designs for the design-specific assumptions.

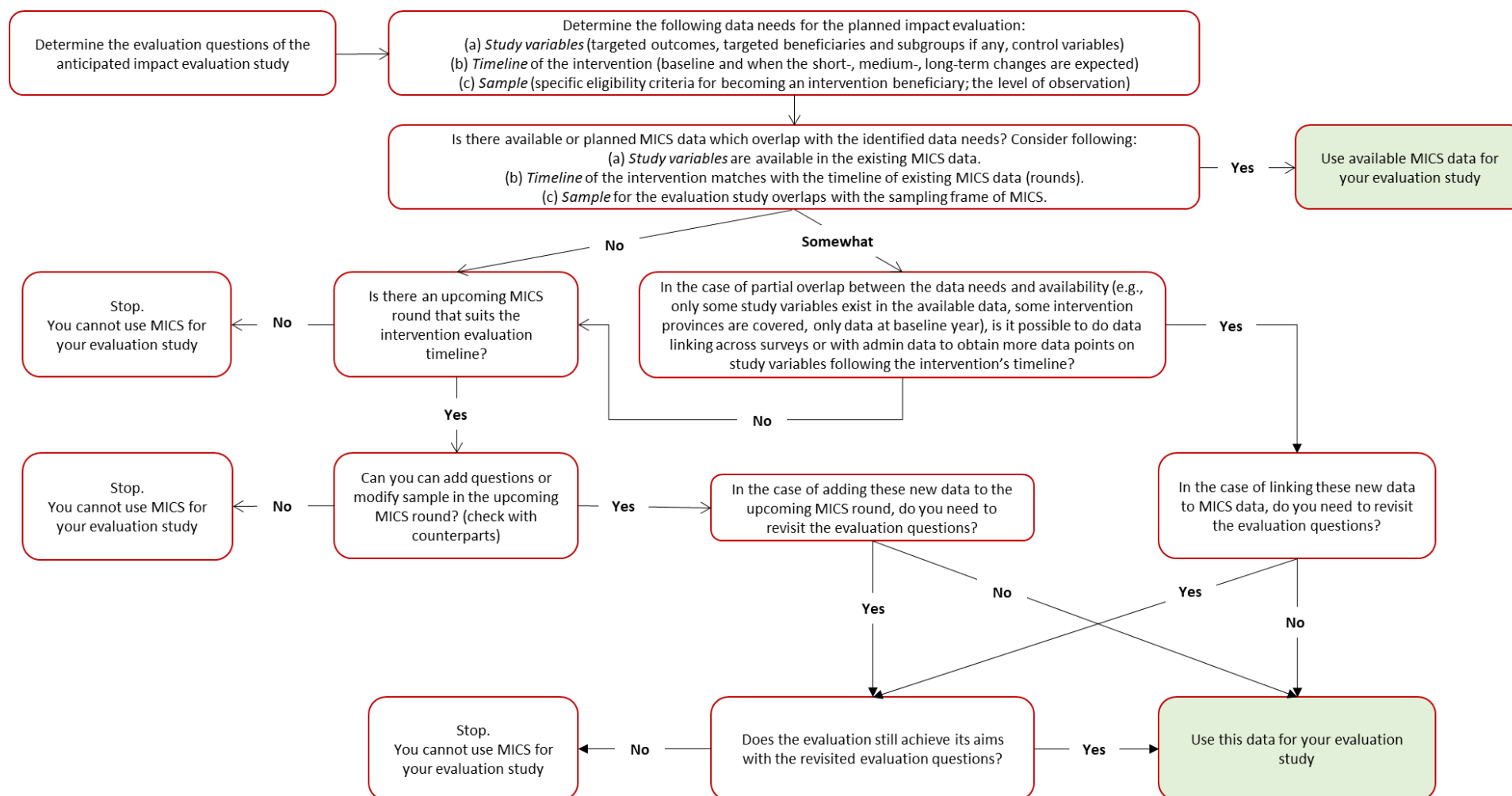
Type of Design	Design diagram	Comparison group data required	Baseline data required	Control for confounding	Strengths and limitations
One-group pretest-posttest	$O_1 \times O_2$	No	Yes	Weak	Do not control for other factors causing changes in the treatment group over time.
Different samples pretest-posttest (repeated cross-sectional)	$O_{1,sampleA} \times O_{2,sampleB}$	No	Yes	Weak	Do not control for population differences and other factors causing changes over time
Two-group posttest only, with matching	T: $\times O_2$ C: O_2	Yes	No	Intermediate	Preexisting differences across the treatment and control populations may confound the findings
Two-group pretest-posttest (difference-in-differences)	T: $O_1 \times O_2$ C: $O_1 O_2$	Yes	Yes	Intermediate	Control for changes over time that affect both populations, but not for potential preexisting trend differences or time-variant confounders.

¹⁰ Kisbu, Y., Cook, T. D., Tang, Y., & Clark, M. H. (2018). Comparative regression discontinuity: A stress test with small samples. *Evaluation Review*, 42(1), 111-143.

Regression discontinuity design	T: $O_A \times O_2$ C: $O_A - O_2$ with cut-off-based assignment to conditions based on a baseline assignment variable O_A	Yes	No	Strong	Limited generalization of impact estimates for populations beyond the cutoff. Can be confounded if there are other interventions that are implemented according to the same cutoff.
Comparative interrupted time series	T: $O_1O_2O_3 \times O_4O_5O_6$ C: $O_1O_2O_3 - O_4O_5O_6$	Yes	Yes, multiple times	Strong	Especially strong when trends are similar in both groups. Can be confounded if there are other interventions/events not captured by the time series that affect groups differently.

Notes. O= observation; X=treatment; T=treatment group; C=Control group.

Annex B. Flow chart for assessing the feasibility and using MICS for an impact evaluation



Annex C. Links to MICS, MICS Plus, and example household surveys' documentation

UNICEF MICS tools (sampling, modules and indicator lists, questionnaires):

<https://mics.unicef.org/tools?round=mics6#survey-design>

UNICEF MICS Plus methodology and use:

<https://mics.unicef.org/mics-plus/methodology-and-use>

Demographic and Health Survey (DHS) data, variables and definitions, indicators:

<https://dhsprogram.com/data/>

<https://dhsprogram.com/data/Data-Variables-and-Definitions.cfm>

<https://dhsprogram.com/data/Survey-Indicators.cfm>

European Union Statistics on Income and Living Conditions (EU-SILC) methodology, implementation by country, modules:

<https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions>

<https://ec.europa.eu/eurostat/web/income-and-living-conditions/database/modules>

European Health Interview Survey (EHIS) country data availability and time series, guidelines, questionnaires, variables:

<https://ec.europa.eu/eurostat/web/microdata/european-health-interview-survey>

Independent Evaluation Laboratory

Koç University
Rumelifeneri Yolu 34450 Sarıyer
İstanbul, Türkiye

Visit iel.ku.edu.tr