Developing and Selecting Measures of Child Well-Being

Howard White and Shagun Sabarwal

UNICEF OFFICE OF RESEARCH

The Office of Research is UNICEF's dedicated research arm. Its prime objectives are to improve international understanding of issues relating to children's rights and to help facilitate full implementation of the Convention on the Rights of the Child across the world. The Office of Research aims to set out a comprehensive framework for research and knowledge within the organization, in support of UNICEF's global programmes and policies, and works with partners to make policies for children evidence-based. Publications produced by the Office are contributions to a global debate on children and child rights issues and include a wide range of opinions.

The views expressed are those of the authors and/or editors and are published in order to stimulate further dialogue on impact evaluation methods. They do not necessarily reflect the policies or views of UNICEF.

OFFICE OF RESEARCH METHODOLOGICAL BRIEFS

UNICEF Office of Research Methodological Briefs are intended to share contemporary research practice, methods, designs, and recommendations from renowned researchers and evaluators. The primary audience is UNICEF staff who conduct, commission or interpret research and evaluation findings to make decisions about programming, policy and advocacy.

This brief has undergone an internal peer review.

The text has not been edited to official publication standards and UNICEF accepts no responsibility for errors.

To consult and download the Methodological Briefs, please visit http://www.unicef-irc.org/KM/IE/

1.   INDICATORS: A BRIEF DESCRIPTION

Indicators provide a signal to decision makers by indicating whether, and to what extent, a variable of interest (such as use of health services) has changed. Indicators can be used at all levels of the results framework from inputs to impact, and should be linked to the programme's theory of change (see Brief No. 2, Theory of Change). For any one programme or policy there should be a comprehensive set of indicators covering all levels of the theory of change. Indicators have an important role for all monitoring and evaluation (M&E) activities, including impact evaluation (see Brief No. 1, Overview of Impact Evaluation).

Most important at the lower levels of the causal chain are monitoring indicators such as inputs (e.g., money spent, immunization kits supplied), activities (e.g., health workers trained, immunization days held) and outputs (e.g., clinics built). For higher-level indicators of outcomes and impact, however, monitoring tells us what has happened but not why it happened. To understand this, impact evaluation must be used. This simply means that while indicators can monitor progress – such as progress made towards the achievement of a particular goal (e.g., reduction in under-five mortality rates) – or provide a warning sign of things going wrong (e.g., 90 per cent of school-aged children in a village do not attend school), understanding the factors behind achieving or not achieving the goal requires more intensive research or evaluation.

Different organizations and documents may use different definitions for an indicator, a measure, a target, a benchmark, a standard or an index, and some of these terms may also be used interchangeably. Hence why it is useful to check the meaning of the terms used in any specific context. This brief uses the following definitions, which are based on the UNICEF core indicators for measuring child protection systems:[1]

- A **measure** is a piece of information about a programme or policy, which may be quantitative (number of children immunized in a region in a year) or qualitative (how well a country's immunization system functions).

- A **variable** is a clearly defined, quantifiable measure of a quantity that may vary across time for a specific unit of observation, or across space, i.e., between different units of observation.

- An **indicator** is a verifiable measure that has been selected by programme or policy management to make decisions about the programme/policy. For example, the proportion of students achieving a passing grade on a standardized test.

- A **target** is the value of, or change in, an indicator expected to be achieved by a specified point in time. Examples include the targets of the Millennium Development Goals such as: "Ensure that, by 2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling."[2]

- UNICEF uses the terms **benchmark** and **index** to mean the same thing – a set of related indicators that provides for meaningful, accurate and systematic comparisons regarding performance. For example, the Human Development Index (HDI) is a composite statistic of life expectancy, education and income used to rank countries into four tiers of human development.

- A **standard** is a set of related benchmarks/indices or indicators that provides socially meaningful information regarding performance. In the above example of the HDI, the standards are the four tiers of human development: very high, high, medium and low human development. In this series of briefs, standards are also referred to as 'rubrics' (see Brief No. 4, Evaluative Reasoning).

---

[1]   United Nations Children's Fund, 'Measuring and Monitoring Child Protection Systems: Proposed Regional Core Indicators for East Asia and the Pacific', *Strengthening Child Protection Series* No. 3, UNICEF EAPRO, Bangkok, 2012.

[2]   United Nations, 'Goal 2: Achieve universal primary education, Target 2.A: Ensure that, by 2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling', web page, Millennium Development Goals and Beyond 2015, UN, http://www.un.org/millenniumgoals/education.shtml.

- A **milestone** captures the achievement of a quantitative or qualitative target, for example, the improvement of enrolment rates as a quantitative target, or the introduction of school-based management as a qualitative target. The milestone can define the variable to be used to measure the achievement of the milestone.

This brief will mainly focus on indicators, including child-specific indicators. When designing M&E activities for child-related interventions, it's important to bear in mind that several aspects of child well-being, including child rights, health and education, require age-specific indicators.

Indicators may be quantitative or qualitative. Quantitative indicators are numerical, for example, the immunization rate in a particular country, the percentage of women attending antenatal classes or the under-five mortality rate). Qualitative indicators use categories of classification, for example, a country has/has not ratified the Convention on the Rights of the Child, or the level of programme implementation is 'high', 'medium' or 'low' (each of these categories is clearly defined).

The SMART criteria[3] are often used to specify what characteristics are required for an indicator to be considered 'good' (see box 1).

Box 1.   SMART indicators

The SMART criteria are a common standard for indicators. Although there are some variations as to what each letter in the SMART acronym stands for, the following are important considerations.

Specific: The indicator is clearly and precisely defined, for example, a specific indicator for 'immunization coverage' would include how many doses are required for maternal injections of tetanus toxoid for someone to count as being covered.

- Is the indicator specific enough to measure progress towards the result?

- Is it clear exactly what is being measured? Has the appropriate level of disaggregation been specified? Is it clear 'which way is up' (i.e., in which direction it is hoped that he indicator will move)?

- Does the indicator capture the essence of the desired result?

- Does it capture differences across areas and categories of people?

Measurable: The indicator is specified in such a way that it is clear what is being measured, for example, 'the proportion of project schools that have instituted student councils'.

- Can a quantified target be set for the indicator?

- Are changes objectively verifiable?

- Is it a reliable and clear measure of results?

- Is it sensitive to change?

Attainable: The indicator should be appropriate to the programme or policy. For example, an indicator that solely measures information provided on the risks of HIV/AIDS is unlikely to help ascertain whether or not risk behaviour has changed.

- Are the changes anticipated as a result of the programme/policy realistic?

Relevant: The indicator should be relevant to the goal or activity it is intended to capture. For example, under-five mortality is a relevant impact indicator for the outcome of higher immunization coverage but anthropometric indicators would be less so.

---

[3]   Originally developed by Doran, George. T. 'There's a S.M.A.R.T. way to write management's goals and objectives', *Management Review* (AMA FORUM), 70 (11), 1981, pp. 35–36.

- Is the indicator plausibly associated with the programme or policy's sphere of influence?

- Does the indicator capture the essence of the desired result?

Trackable: It must actually be possible to collect data for the indicator over a set period of time.

- Are data actually available at reasonable cost and effort?

Main points

1. Indicators provide a signal to decision makers by indicating whether, and to what extent, a variable of interest has changed.

2. Indicators may be used at all points of the results chain: inputs, activities, outputs, outcomes and impacts.

3. The M&E system for any intervention should include a balanced set of indicators, including a range of these across the causal chain.

4. There are four possible sources of data for indicators: administrative data, existing census or survey data, project monitoring data, or data collected by conducting a new survey. Each of these sources has its own pros and cons

---

2. HOW TO SELECT OR DEVELOP INDICATORS

Indicators may be used at all points of the results chain: inputs, activities, outputs, outcomes and impacts. These are referred to as different 'types' of indicators:[4]

- inputs – the financial, human and material resources used in a programme or policy. For example, training materials produced.

- activities – actions taken or processes through which inputs are mobilized to produce specific outputs. For example, counselling sessions that adhere to quality standards.

- outputs – the immediate effects of programme/policy activities, or the direct products or deliverables of programme/policy activities. For example, the number of vaccines administered.

- outcomes – the intermediate effects of a programme or policy's outputs such as a change in vaccination levels or key behaviours.

- impacts – the long-term, cumulative effects over time of programmes/policies on what they ultimately aim to change. For example, under-five morbidity and mortality.

The M&E system for any intervention should include a balanced set of indicators, including a range of these across the causal chain. The challenge is to select indicators that can capture what matters whilst being realistic in terms of what type of data can be collected.[5] Outcome and impact indicators mostly use internationally accepted standard definitions (this is preferred), whereas indicators lower down the causal chain are more specific to the programme or policy for which the M&E system is being defined.

---

[4]   Joint United Nations Programme on HIV/AIDS, *Basic Terminology and Frameworks for Monitoring and Evaluation, UNAIDS Monitoring and Evaluation Fundamentals*, UNAIDS, Geneva, 2010. See http://www.unaids.org/en/media/unaids/contentassets/documents/document/2010/7_1-Basic-Terminology-and-Frameworks-MEF.pdf.

[5]   United Nations Development Programme, 'RBM in UNDP: Selecting indicators', *Signposts of Development*, UNDP, New York. See http://toolkit-elections.unteamworks.org/?q=webfm_send/105.

Evaluations should use the indicators that were initially selected as indicators of a programme or policy's performance by those designing the intervention. These, in turn, should have been derived with reference to the programme or policy's theory of change. Often, key performance indicators (KPIs) have already been defined and collected at the outcome or impact level, and are a subset of the indicators compiled in the management information system of a particular intervention. A key set of indicators may also be put on an 'indicator dashboard', a management tool that gives a graphical overview of the current status (and often also trends over time) of the KPIs to show 'at a glance' how the intervention is progressing. Where appropriate indicators are lacking, these should be derived ex post, using the programme's theory of change and in consultation with key stakeholders.

To whatever extent possible, it is preferable to select standard indicators using commonly accepted definitions, since the use of standard definitions makes it easier to compare findings across studies. For example, anthropometic data on height and weight are usually expressed as z-scores using international norms produced by the World Health Organization (WHO). Similarly, the design of data collection instruments should be based on existing 'tried and tested' models such as the USAID-supported Demographic and Health Survey (DHS) Program or UNICEF-supported Multiple Indicator Cluster Survey (MICS) initiative. There are, however, disadvantages to using existing indicators, chiefly that they may not be well adapted to the local context. Such drawbacks can be mitigated by using a consultative approach for the selection and/or measurement of indicators to ensure the best fit of existing indicators to local context and preferences.

There may be times when indicators must be customized to the specific cultural setting. For example, a DHS can be used to collect data to construct indicators of female empowerment, but the underlying interview questions (and thus the definition of the indicators) will vary by region. In South Asia, for example, respondents are asked about their ability to move outside of the home, but the same questions are irrelevant in much of sub-Saharan Africa. In such contexts, questions relating to women's control of resources and attitudes to gender-based violence are instead used. It is important to ensure that the data collected on indicators actually correspond to the information that is being sought. Questions and survey tools that collect data on indicators should be pilot tested with a smaller number of respondents in areas where the survey will eventually be administered.

In development, it is often difficult to make objective and exact measurements of complex development changes. Instead, there is a reliance on 'proxy' indicators, which are commonly understood to be closely related to or to 'approximate' the intended changes. For example, the proportion of female parliamentarians may be used as a proxy indicator for the participation of women in national decision making processes. Proxy indicators may also be used when the desired indicator is too difficult or costly to measure, or if what is intended to be measured takes a long time to be achieved.

Once selected, indicators must be operationalized by defining how and by whom the corresponding data will be collected. If using a survey, the specific survey questions must be developed and field-tested.

3.   HOW TO COLLECT INDICATORS ON CHILD WELL-BEING

Indicator variables

Indicators may be constructed from the following types of variables:

- Continuous variables can take any value though may fall in a range. For example, age, years of education, distance to nearest health or education facility, expenditure on specific items, and body temperature.

- Categorical variables can take only certain, predefined values. Examples are blood type (A, B, AB or O), region of residence and principal occupation of household head.

- <u>Dichotomous variables</u> have only two possible values, usually 0 and 1 (though 1 and 2 are more commonly used in data collection). Dichotomous variables can be used for personal characteristics (e.g., male = 1, female = 2), programme participation (i.e., non-participant = 0, participant = 1) or an outcome variable such as mortality (i.e., alive = 0, died = 1).

- Ordinal variables are categorical values in which the responses are ordered such as educational attainment or use of health facilities. For example, the values for the question 'What is your highest level of educational attainment?' could be: 1) Did not complete primary; 2) Completed primary; 3) Incomplete secondary; 4) Completed secondary; and 5) Tertiary. Or for the question 'When your child last suffered from diarrhoea from whom did you seek advice or help?', the values could be: 1) Health worker; 2) Traditional healer; 3) Friend or family member; and 4) No one.

- A <u>Likert scale</u> is a question design that yields an ordinal scale based on the respondent's self-assessment. For example, to the question 'How were you treated by the staff at your last visit to a health facility', the response might be: 1) Very well; 2) Well; 3) Satisfactorily; 4) Poorly; or 5) Very poorly.

Data sources for indicators

There are four possible sources of data: administrative data, existing census or survey data, project monitoring data or data collected by conducting a new survey. Table 1 shows the strengths and limitations of each of these sources.

Table 1. Pros and cons of different data sources

	Pros	Cons
Administrative data	- National coverage - Annual basis - Low cost	- May have under- or over-reporting bias - May exclude private sector - May not get full access - Often facility-based, so miss out those not using facilities - Provide only numerator and not also denominator for calculating rates and ratios
Existing census or survey	- Low cost as using existing data - Often have national coverage and may be representative at level of main geographical regions - National surveys often contain very comprehensive data	- Survey may have insufficient <u>sample</u> size in programme areas - May not include questions relating to intervention - Census has limited set of data - May be outdated or unavailable for right time period

Project monitoring data	• Low cost	• Usually no data for comparison group • May have under- or over-reporting bias
New survey	• Can customize to data requirements • Can collect required sample size	• Time and cost • Ensuring adequate data quality can be a challenge

Administrative data are commonly available as most countries have both an education management information system (EMIS) and health management information system (HMIS) with data collected at the facility level (school, hospital and clinic) using an annual census completed by the head of the facility. EMIS and HMIS information may be suitable for impact evaluation, for example, there has been an evaluation of school-based management in the Philippines using the basic education information system (BEIS).[6] EMIS and HMIS may not always have reliable data on outcomes, but EMIS data can often be combined with a separate database of average test scores at the school level. Health facility level data cannot be used to measure health outcomes, however, as facility level data do not capture the health status of those not using the facility (i.e., those resorting to self-treatment, a traditional healer or asking a pharmacist for advice).

Data will often be required from several different surveys, for example, a household survey, a facility survey (of schools or clinics), a survey of service providers (teachers or health workers) *and* a community survey. If measuring children's educational outcomes, data are needed on the child's home environment and parental education. These can be collected from a household survey, but data on school facilities must come from a school survey. The data should be collected in such a way that allows for a child's data to be linked to the data for the school that he/she attends. This linking must be done using pre-assigned community and facility codes, since names and spellings are often non-standard.

The most relevant existing surveys are the MICS and DHS. Both are household surveys rather than facility surveys, and so allow for nationally representative samples, and both include detailed information about maternal and child health, nutrition and education. The household income and expenditure surveys collected under the auspices of the Living Standards Measurement Study (LSMS) contain data on educational attainment, health and education expenditure, and anthropometric indicators for children under five. Some countries also run other child-specific surveys, also collected at the household level, most notably the international Young Lives project, which has data for Ethiopia, Andhra Pradesh (India), Peru and Vietnam, and the Birth to Twenty project in South Africa. Both of these projects follow cohorts of children from birth onwards.

Collecting indicators on child well-being

Collecting indicators for children presents specific challenges.

Firstly, many child-specific issues exist, especially around the unique needs and situation of children. For example, tests of learning outcomes and tests used to ascertain cognitive development and social skills are age-specific, and questions relating to the caregiving environment are child- and age-specific. Indicators must accurately capture cognitive and socio-emotional development during childhood.

---

[6]   Khattri, Nidhi, et al., 'The effects of school-based management in the Philippines: an initial assessment using administrative data', *Journal of Development Effectiveness*, 4(2), 2012, pp. 277–295.

Secondly, the most appropriate measure for assessing child welfare changes according to the child's age. Not only should the indicators be age-specific, as highlighted above, but it may also be necessary to disregard absolute changes in outcomes for monitoring progress and/or measuring impact, as these might not make sense in all situations. For example, literacy is not a relevant question for a 2-year-old, but it is for a 10-year-old.

Finally, there is the challenge of direct data collection. Often data cannot be collected directly from a child if the child is very young or there are ethical issues (discussed below). As such, the primary caregiver(s) must be engaged. Doing so may introduce bias, however, as caregivers may not want to show themselves in a bad light and are therefore more likely to give socially desirable responses (e.g., in relation to violence) or may seek to protect the child or other family members. One possible way to counter this problem is to conduct independent verification wherever possible. For example, immunization estimates are more reliable if calculated for those children for whom immunization cards have been produced. Whatever the approach might be, the indicators used should be well formulated and properly validated.

Child-specific indicators by sector

Health

The most intuitive and straightforward health indicator is to ask someone how they, or their children, are. The self-rated health question commonly used is: 'In general, would you say your health is excellent, very good, good, fair or poor?'. Self-reported health data are notoriously unreliable, however, and evidence exists that individuals may interpret the categories of 'excellent', 'very good' and so on quite differently.[7] Famously, people in the USA reported they were sicker than those in the Indian state of Bihar.[8] It is therefore likely that there is also *bias* in the parental reporting of child health; the better off are more likely to report sickness, and more so for boys than for girls. [9]

There are two ways to avoid this bias. The first is to be as specific as possible about symptoms. The second is to use vignettes, also known as anchoring vignettes, which were pioneered by WHO for the World Health Survey.[10] Vignettes give short descriptions of the health status and consequent incapacity of three people; respondents are asked to rank the three in terms of their health, and then rank themselves using the same system. Vignettes are a means of measurement and their use culminates in scales that can be used as indicators.

Mortality indicators are commonly used to measure the health status of children across different countries, states within countries and regions. Mortality indicators used for children include the following: neonatal mortality (the probability of dying within the first month of life); infant mortality (the probability of dying before the first birthday); and child mortality (the probability of dying between the first and the fifth birthday). All rates are expressed per 1,000 live births except child mortality, which is expressed per 1,000 children surviving to 12 months of age. Under-five mortality combines infant and child mortality and gives an estimate of the likelihood of dying before the fifth birthday. Premature death is a clear indicator of welfare deprivation and often describes the extent of human development of a country, and these indicators are most often unambiguous. There are, however, two issues that make measuring mortality challenging: sample size and 'censoring', both of which are explained in brief below.

If presented with a group of children under five years of age, it is impossible to directly measure under-five mortality for two reasons. Firstly, data must be captured on those who were born in the reference period and have already died. This is done through a 'birth history' module, as used in the MICS and DHS,

---

[7]   Grol-Prokopczyk, Hannah, et al., 'Using Anchoring Vignettes to Assess Group Differences in General Self-Rated Health', *Journal of Health and Social Behavior,* 52(2), 2011, pp. 246–261.

[8]   Sen, Amartya, 'Health: perception versus observation', *British Medical Journal,* 324(7342), 2002, pp. 860–861.

[9]   Ibid.

[10]   World Health Organization. WHO World Health Survey. http://www.who.int/healthinfo/survey/en/.

although in some settings there may be cultural sensitivities around reporting the birth of children who have since died (stillbirths are omitted, as mortality refers to deaths as a proportion of live births). Secondly, children under five have not yet reached their fifth birthday, and it's impossible to know if they will survive until then. This is a problem known as censoring, which can be resolved by statistical means. A further hurdle is getting the right data on date of birth and date of death – something that is not always easy in a developing country.

Large sample sizes are needed to measure mortality with sufficient precision. For this reason, mortality estimates are frequently reported for five-year ranges rather than for single years. As a result, mortality is generally not used as an indicator in programme/policy impact evaluations as sample size requirements are prohibitive. Power calculations must be conducted for any specific study to see whether the sample size is large enough.

As well as mortality indicators, a number of indicators for measuring childhood morbidity exist (in addition to those on nutrition, discussed above). Morbidity is more common than mortality. A common way of measuring morbidity is to ask caregivers of young children if the child in question had suffered from a particular illness (this is asked for a range of illnesses), usually within a short recall period. For example, the MICS and DHS both ask whether each child under five in the household has had diarrhoea, a cough or fever within the last two weeks. In cases where a child has had such an illness, follow-up questions on the type of treatment given, the type of health facility visited and so on are asked of the child's caregiver. In order to analyse caregiver treatment responses, the data on child illness are needed for the denominator.

HIV/AIDS

Indicators for HIV and AIDS programmes collected from household surveys relate to knowledge, attitudes and sexual behaviour. Knowledge questions can ask about how HIV is transmitted, what can be done to prevent it, misperceptions about HIV, mother-to-child transmission and where to access HIV testing services. Some questions ask about attitudes towards people living with HIV. Sexual behaviour questions ask about age at first sex, number of lifetime and recent sexual partners, use of a condom at last sex, and age of recent sexual partners. Some surveys may also include questions on the use of alcohol and drugs.

HIV and AIDS prevention, care and treatment interventions often target the general population, particularly in high HIV prevalence countries. In countries with low or concentrated epidemics, however, the target is mainly key populations at increased risk of HIV infection such as sex workers, men who have sex with men, and individuals who inject drugs. Collecting accurate data from key populations is challenging as such behaviours are stigmatized and criminalized, and hence are often hidden. Some experiences show that gaining the trust of these groups, and using enumerators from inside them, can improve the quality of data collection.

Finally, data may be collected on a person's HIV status through blood tests or swabs taken during household surveys or at health facilities. There are serious ethical issues in conducting such tests, including the need to establish clear protocols for obtaining consent and for informing the respondent of the test result as well as ensuring the provision of support in cases of an HIV-positive test result.

Nutrition

Nutritional knowledge and beliefs of the primary caregiver can influence key behaviours relevant to ensuring the optimal nutrition status of infants and young children. Standard indicators and related questions exist in regard to key practices such as exclusive breastfeeding or the timely introduction of solid, semi-solid and soft foods. Monitoring of these standard indicators is often part of the evaluated programme. There may also be practices unique to a particular context that can influence the nutrition status of young children and these can also be monitored sometimes. For example, in Bangladesh there is a widespread belief in the benefits of 'eating down', i.e., eating less during pregnancy, which may have negative

consequences for the unborn child. Collecting and analysing data on such practices for which standard methods do not exist requires careful decisions to be made about appropriate methods (i.e., qualitative versus quantitative). In addition, indicator definitions should be developed and associated questions should be validated and field tested.

There is a wide range of nutritional practices that may be asked about. Key standard coverage and practice-related indicators in this field include infant and young child feeding practices as described by WHO,[11] household use of adequately iodized salt; and vitamin A supplement coverage for children aged 6 to 59 months. Depending on the goals of a programme, it may also be useful to monitor indicators that influence nutritional status such as water and sanitation coverage.

Household surveys are not always the recommended method for collecting data on the above indicators. For example, vitamin A supplementation is generally reported via administrative data sources. Given the number of factors that influence nutrition status, there may be an eagerness to collect information on numerous indicators when trying to monitor nutrition programmes. It is important to recognize, however, that data quality suffers as more questions are added to household surveys, and the number of indicators included in household surveys must be balanced with the feasibility of collecting data in the field.

The most common nutrition impact indicators for children under the age of five years use anthropometrics based on weight, height and age: (1) weight-for-height can be used to ascertain *wasting prevalence*, which represents how many children are too thin for their height – it can change quite quickly and is often seasonal; (2) height-for-age can be used to ascertain *stunting,* which represents how many children are too short for their age and is a good indicator of cumulative nutritional insults during the most critical periods of growth; and (3) weight-for-age is a composite measure of nutritional status that comprises elements of both wasting and stunting. Each of these measures is expressed as a prevalence of children falling below minus two standard deviations from the median of the reference population (WHO Child Growth Standards). The proportion of children falling below minus three standard deviations from the median of the reference population is considered severely malnourished (i.e., severely wasted, stunted or underweight).

Anthropometric measures are usually taken for children up to the age of five. For older children, other measures may be used such as upper arm circumference or body mass index.

Water, sanitation and hygiene (WASH)

Lack of access to improved drinking water sources and to improved sanitation facilities, and poor hygiene practices are important determinants of child health, most notably diarrhoeal diseases, which cause 9 per cent of all under-five deaths and are also an important contributory factor to stunting.

Standardized household survey questions on water, sanitation and hygiene (WASH) in the MICS and DHS typically ask about: (1) the use of different types of drinking water sources and different types of sanitation facilities (they also track the practice of open defecation – the worst form of poor excreta disposal, with the biggest impact on public health); (2) the presence on the premises of a specific place for handwashing, where water and soap are present; and (3) the use of household water treatment, if any.

The latter questions on household water treatment, like questions on knowledge of and actual handwashing behaviour, have been found to result in vast over-reporting. Their use should be discouraged in favour of the much more robust questions on the use of specific drinking water sources and sanitation facilities, and the presence of a specific place for handwashing with soap – which is recorded through observation.

The most common outcome measure for water and sanitation interventions is child diarrhoea. Attributable fractions are difficult to estimate, however, and confounding factors are common, which makes reduction in diarrhoea incidence a less recommended indicator of success unless a randomized controlled trial (RCT)

---

[11]   World Health Organization, *Indicators for assessing infant and young child feeding practices, Part 1, Definitions*, WHO, Geneva, 2008. See http://www.who.int/nutrition/publications/infantfeeding/9789241596664/en/.

design is used. A more accurate indicator used in recent years is stunting. The chronic malnutrition that causes stunting is very strongly correlated with open defecation practices and poor sanitation. Cholera outbreaks provide good indicators of WASH interventions in children. They are sparse, however, and cholera is often foodborne rather than waterborne. Time savings, measured by round trip time (i.e., the time it takes to go to the source, get water – including waiting in line – and return) is another important outcome indicator for women and children living in households with no water on the premises, and it should be included as an outcome in any impact evaluation of WASH interventions.

Education

Most countries have an EMIS, which is an annual census of schools that is usually completed by each school itself for the ministry or department of education. The EMIS collects basic data on school facilities and numbers of teachers and pupils. EMIS data are thus an example of administrative data, which, if sufficiently comprehensive, may be used for impact evaluations. EMIS data have their advantages: (1) these data are supposed to cover all schools, rather than a sample, and therefore they allow for disaggregation at lower levels (e.g., district and school); and (2) EMIS data provide information on the school environment and the quality of education provided.

Sometimes an EMIS will exclude the private sector, however, or only include it on a voluntary basis, which may be a source of bias. Another limitation of EMIS data is that they cannot be used to estimate enrolment rates since: (1) the private sector is excluded; (2) in some countries there may be biases to encourage over-reporting of student numbers; and (3) student numbers are the numerator for the enrolment rate but the denominator is not captured by the EMIS.

One way of overcoming some of the problems of EMIS data is by using household surveys. Standard household surveys (such as the MICS) collect data on whether children are currently in school and their current grade/class, and the number of completed years of education for those who have left school. Data are also collected on children's school attendance in the previous year, enabling the computation of flow rates (i.e., repetition, dropout, survival, transition, etc.). Household surveys thus provide high quality data on both gross attendance rates (the number of students attending school at several different grade levels) and net attendance rates in a particular school year (attendance of the official age group for a given level of education expressed as a percentage of the corresponding population), and possibly also information on attendance in the past week (care is required in analysing these data so absence from school during official holidays is not counted as non-attendance). Household data can also be used to analyse transition from one stage of schooling to another, completion, dropout and repetition.

A further problem with EMIS data for many study designs is that the data are collected at the facility level. Two issues arise as a result. Firstly, as with all facility data, data are not captured on those not using the facility. Secondly, the data are not linked to household data, despite the home environment being important in determining education outcomes.

While information on enrolment and attendance is important for understanding the status of the education of children, there is generally also an interest in learning outcomes. There may be national test score data available either in the EMIS, or which can be linked to the EMIS at the school, class or individual level. An advantage of these data is that they are national and so can be disaggregated down to district level. Data from national test systems may be unsuitable for use in impact evaluations, however, for a number of reasons: (1) data with the required identifiers to link to specific schools are unavailable; (2) data are not an accurate measure of learning outcomes because of corruption or other abuse in the exam system (e.g., teachers give students the answers so their school gets good results); and (3) national exams are not carried out in every grade, so may not apply to the grades covered by the intervention. If any of these problems is present then it will be necessary to collect new learning outcome data.

An impact evaluation will usually require data from both household and facility. A school-level RCT is an exception, although even in this case the analysis that can be done is limited if data from the household are unavailable.

Early child development

Another aspect of the measurement of child well-being that until recently had received far less attention is the importance of preschool attendance along with other aspects of early childhood development (ECD).

The standard MICS questionnaire includes a comprehensive module on ECD. The module collects information on several aspects of relevance to ECD, including information on the quality of care that children receive, their home environment (such as the availability of reading and play materials), their attendance in early childhood education and adult support for learning. The module also includes a specific set of questions intended to assess children's social-emotional, physical and cognitive (learning and numeracy/literacy) development.

Child protection

The MICS provides a good range of indicators related to child protection. Questions in the standard questionnaire cover attitudes towards wife-beating, child labour, child marriage and female genital mutilation/cutting. The MICS also collects data on child discipline and children with disabilities. There are other important issues that come under the category of child protection and these relate to sexual exploitation and violence against children. Some of these topics such as violence are sensitive, thus raising ethical and practical concerns in data collection. (Other topics such as marriage are not sensitive.)

Structured surveys may be an inappropriate means for collecting data on sensitive topics, since the relatively formal approach prevents the creation of a strong rapport and limits the space for much probing. Qualitative methodologies such as in-depth interviews, focus group discussions (see Brief No. 12, Interviewing) and other innovative approaches (e.g., use of maps) may be more appropriate and are often used to collect data on sensitive topics, for example, as in a study of sexual abuse of girls in schools in Ghana.[12]

Income poverty and asset indexes

'Poverty' is often taken to mean income poverty, although it is widely accepted that poverty is a multidimensional concept. Income poverty is measured either through income data or expenditure; the latter measure is generally regarded as more reliable. Asset indexes have become more common as they are relatively easy to collect and develop. Many surveys – like the DHS and national LSMS surveys – collect data on assets but not on income or expenditure.

Income poverty is measured at the household level, so the incidence of child poverty can be estimated by calculating the proportion of children living in poor households. Poverty calculations need to adjust for economies of scale in consumption (some goods are shared by household members, so two people living together live more cheaply than two people living alone) and household composition (adult equivalence scales are used to allow for the fact that children consume less than adults, and females less than males). Larger households are generally poorer than smaller ones, so the incidence of child poverty will be higher than that amongst adults of the national population as a whole.

There is some child-specific spending which can be analysed for comparison purposes. For example, differential spending on health and education for boys and girls can be an important finding. But it is

---

12   White, Howard, 'Using Household Survey Data to Measure Educational Performance: The Case of Ghana', *Social Indicators Research*, 74(2), 2005, pp. 395–422.

empirically difficult to estimate the 'total expenditure' per child, and so it makes more sense to talk of 'children living in poor households' than of 'poor children'.

Poverty estimates are usually calculated using household survey data. A limitation of household surveys, however, is that they miss out those individuals not living in households, which includes nomadic populations, homeless individuals, street children, people in refugee camps, and those living in institutions such as facilities for people with disabilities and prisons.

Inequality

Indicators that measure how income or any other resource is distributed among individuals or households in a population are referred to as 'measures of vertical inequality' and give an idea of relative well-being. These measures can be applied to the distribution of income or expenditure or other impact indicators such as mortality, access to services, and assets.

Measuring the distribution of income or expenditure requires a measure of household income per capita adjusted for household size and composition to allow for economies of scale in consumption and the differential consumption requirements of men and women, and adults and children.

Inequality can be shown graphically in a Lorenz curve, which plots the cumulative share of income against the cumulative share of population.

There are various measures, of which the most common is the Gini coefficient. The Gini coefficient is a measure of statistical dispersion of income distribution, which ranges from 0 (perfect equality) to 1 (extreme inequality, which would occur if one person accounted for all income). More transparent measures are the income shares by population quintile. In most cases, UNICEF uses income or wealth quintiles to measure income inequality, giving particular emphasis to the lowest two quintiles.[13] One can also specifically examine the share of income or another indicator accruing to the poorest group (and the wealthiest group) and monitor how it changes over time.

Measures of access to basic services such as education, health and nutrition are often presented by income or asset quintile.

Inequality between groups (referred to as 'horizontal inequality') such as between men and women or between different religious groups is important for identifying groups suffering from social exclusion.

All of the above measures can be used to measure the level of inequality at a particular point in time or the change in the level of inequality over time. Measures such as the Gini coefficient, and to some extent quintile income and asset share, take a long time to register substantial changes.

Finally, a distinction is made between inequality of opportunity and inequality of outcome. Equality of opportunity is captured by equality of access to good education, health and employment opportunities or credit and other requirements for self-employment. Equality of outcome focuses on equity or otherwise in key outcome and impact indicators such as income, mortality and rights.

4.   ETHICAL ISSUES AND PRACTICAL LIMITATIONS

The ethical issues around conducting evaluations in general, and specifically around data collection and reporting related to children, have been addressed in other briefs (see Brief No. 1, Overview of Impact Evaluation; Brief No. 3, Evaluative Criteria; and Brief No. 12, Interviewing). Further information can also be found on the dedicated online resource Ethical Research Involving Children.

---

[13]   Stewart, Frances, 'Approaches towards Inequality and Inequity: Concepts, measures and policies', *Office of Research Discussion Paper* No. 2013-01, UNICEF Office of Research, Florence, 2013. See http://www.unicef-irc.org/publications/pdf/stewart%20inequality_inequity_layout_fin.pdf.

There are also particular ethical and practical issues related to collecting and using indicator data, including around:

- inaccurate or missing data

- data corruption – errors in data that occur during the writing, reading, storage, transmission or processing of computerized data, which introduce unintended changes to the original data

- data gaming – ways of intentionally distorting the indicator definition and/or the data to achieve performance standards

- measuring what can be easily measured rather than what matters – some outcomes and impacts, especially those related to child development and rights, are more difficult, and sometimes costly, to measure. But this should not cause researchers to resort to measuring what can be measured rather than what should be measured.

All of these issues may lead to false conclusions being drawn about the programme or policy (or components thereof).

5.   HOW TO USE INDICATORS IN IMPACT EVALUATION

The purpose of well selected indicators in results-based management is to monitor routinely key conditions in the implementation of a programme or policy as well as specific results arising from the intervention (usually direct outputs, and often also outcomes) in order to inform decisions about the programme or policy. As noted above, an impact evaluation can draw on existing data supplemented by other sources. Ideally, ongoing performance measurement and in-depth evaluation should be designed to function as an integrated system.[14]

For impact evaluation, it is sometimes thought that only data on outcome or impact indicators are required. Causal chain analysis also requires data on intermediate variables such as changes in knowledge and behaviour, however. In addition, it has been acknowledged that the specific processes (and the quality thereof) by which programme activities convert inputs (different types of resources) into results (outputs, outcomes and impacts) are often overlooked in a results chain – an oversight referred to as the 'missing middle'. These are the institutional arrangements (including government) and processes that determine decision making and the process by which those decisions are, or are not, implemented (referred to as 'governance'). While measures to assess inputs, outputs and outcomes have been elaborated for most sectors including child protection, the tools available for measuring governance indicators are still at a nascent stage.[15]

A major challenge in evaluation is that many so-called measures, even when measured carefully and accurately, are only indicators of important variables, and therefore must be complemented by triangulated data. In other words, indicators only indicate; they do not explain. Determining that change has occurred does not tell you why it has occurred. Indicators constitute only one part of the logical and substantive analysis needed to understand the context, the programme/policy and its effects, including impact. Brief No. 1, Overview of Impact Evaluation provides more information on mixing methods at the design and analysis stages.

---

[14]   Barrados, Maria, and J.S. Blain, 'Improving Program Results Through the Use of Predictive Operational Performance Indicators: A Canadian Case Study', *American Journal of Evaluation*, 34(1), 2013, pp. 45–56.

[15]   United Nations Children's Fund, 'Measuring and Monitoring Child Protection Systems: Proposed Regional Core Indicators for East Asia and the Pacific'*, Strengthening Child Protection Series* No. 3, UNICEF EAPRO, Bangkok, 2012.

6.   EXAMPLE INDICATORS USED IN UNICEF STUDIES

UNICEF impact evaluations use a range of indicators, some of which are common for development interventions, like income per capita and schooling, and others of which are quite specific to the interventions, like questions on substance abuse amongst teenagers (table 2). All of these indicators are outcome and impact indicators that appear appropriate to the intervention being evaluated. As mentioned above, it is best to use existing indicators rather than design new ones: the use in Pakistan of the Bayley Scales of Infant and Toddler Development is a good example of this practice.

Table 2.   Indicators used in selected UNICEF studies

Study	Indicators
'Innovative Features in Conditional Cash Transfers: An impact evaluation of Chile Solidario on households and children', Martorano, Bruno, and Marco Sanfilippo, *Innocenti Working Papers* No. 2012-03, UNICEF Innocenti Research Centre, Florence, 2012.	Monetary: household income per capita (thousands of USD) Labour: a) employment for the household head (%); b) employment in the informal sector for the household head Education: (a) the highest number of years of education in the family; (b) school enrolment of school-aged children Health: enrolment in public health system
The Pakistan Early Child Development Scale Up Trial: Outcomes on Child Development, Growth and Health, Yousafzai, A.K., et al., PEDS Trial Outcome Data Report, written for UNICEF Pakistan, Karachi, 2012.	Early childhood development: scores on cognitive, language, motor and socio-emotional development assessed using Bayley Scales of Infant and Toddler Development (third edition) Nutrition: growth as indexed by length/height-for-age, weight-for-age and weight-for-height z-scores.
The South African Child Support Grant Impact Assessment: Evidence from a survey of children, adolescents and their households, UNICEF South Africa, Pretoria, 2012.	Education: a) grade attainment; b) scores on mathematical ability tests and reading and vocabulary tests; c) average days absent from school Child labour: adolescents employed outside the home (days of employment in the previous month) Health: likelihood of illness (as measured during a 15-day period prior to the survey) Risky sexual behaviours among adolescents: a) sexual activity ('never had sex'); b) number of sexual partners, pregnancy ('ever pregnant') Substance use: a) alcohol ('never had alcohol'); b) drugs ('never had drugs')

7.   KEY READINGS AND LINKS

Readings

Barrados, Maria, and J.S. Blain, 'Improving Program Results Through the Use of Predictive Operational Performance Indicators: A Canadian Case Study', *American Journal of Evaluation*, 34(1), 2013, pp. 45–56.

Child Protection Monitoring and Evaluation Reference Group, *Ethical Principles, Dilemmas and Risks in Collecting Data on Violence against Children: A review of available literature*, UNICEF, Statistics and Monitoring Section/Division of Policy and Strategy, New York, 2012. See http://www.childinfo.org/files/Childprotection_EPDRCLitReview_final_lowres.pdf. (Accessed 1 January 2014.)

Kusek, Jody Zall, and Ray Rist, Ten Steps to a Results-Based Monitoring and Evaluation

System: A Handbook for Development Practitioners, World Bank, Washington D.C., 2004. Specifically chapter 3, 'Selecting Key Performance Indicators to Monitor Outcomes', pp. 65–79. See http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2004/08/27/000160016_20040827154900/Rendered/PDF/296720PAPER0100steps.pdf.

Lippman L.H., et al., 'Positive indicators of child well-being: A conceptual framework, measures and methodlogical issues', *Innocenti Working Paper* 2009-21, UNICEF Office of Research – Innocenti, Florence, 2009. See http://www.unicef-irc.org/publications/pdf/iwp_2009_21.pdf. (Accessed 2 January 2014.)

McCormick, Marie C., 'Issues in Measuring Child Health', *Ambulatory Pediatrics*, 8(2), 2008, pp. 77–84.

Perrin, Burt, 'Effective Use and Misuse of Performance Measurement,' *American Journal of Evaluation*, 19(3), 1998, pp. 367–379.

Stewart, Frances, 'Approaches towards Inequality and Inequity: Concepts, measures and policies', *Office of Research Discussion Paper* No. 2013-01, UNICEF Office of Research, Florence, 2013.

UNESCO Institute for Statistics, *Children Out of School: Measuring Exclusion from Primary Education*, UNESCO Institute for Statistics, Montreal, 2005. See http://www.uis.unesco.org/Library/Documents/oosc05-en.pdf.(Accessed 2 January 2014.)

White, Howard, et al., 'Comparative Perspectives on Child Poverty: A Review of Poverty Measures', *Working Paper* No. 1, Young Lives, 2002. See http://www.younglives.org.uk/publications/WP/comparative-perspectives-on-child-poverty-a-review-of-poverty-measures. (Accessed 28 December 2013.)

Links

Anchoring Vignettes, website, http://gking.harvard.edu/vign.

CE-DAT, 'Complex Emergency Database', web page on data for a range of indicators for countries subject to complex emergencies, Centre for Research on the Epidemiology of Disasters, http://cedat.be/.

Childinfo, 'Multiple Indicator Cluster Surveys (MICS)', web page on country reports prepared using MICS data, UNICEF, http://www.childinfo.org/mics.html.

The DHS Program, home page for website on Demographic and Health Survey data and reports, http://www.measuredhs.com/.

United Nations Children's Fund, 'Measuring and Monitoring Child Protection Systems: Proposed Core Indicators for East Asia and the Pacific', *Strengthening Child Protection Series* No. 3, UNICEF EAPRO, Bangkok, 2012. See http://www.unicef.org/eapro/Measuring_and_monitoring.pdf.

United Nations Children's Fund, 'Multiple Indicator Cluster Survey (MICS)', web page, UNICEF, http://www.unicef.org/statistics/index_24302.html.

The World Bank, 'About LSMS', web page on the Living Standards Measurement Study, The World Bank, http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/0,,contentMDK:21478196~menuPK:3359066~pagePK:64168445~piPK:64168309~theSitePK:3358997,00.html.

GLOSSARY

Anthropometric	*The measurement of the size and proportions of the human body, for example weight-for-height scores.*
Ethics/ Ethical research	*Ethical research is research that follows widely held guidelines about what is ethical, moral and responsible in research settings (e.g. not plagiarizing others' work, not submitting questionable data, not destroying or concealing sources, etc.). Ethical research is reflexive and explicitly considers its impact on both participants and the broader community throughout the research cycle.*
Evaluation	*A systematic and objective effort to determine the relevance, appropriateness, effectiveness, efficiency, impact and sustainability of development efforts, based on agreed criteria and benchmarks among key partners and stakeholders. It involves a rigorous, systematic and objective process in the design, analysis and interpretation of information to answer specific questions. It provides assessments of what works and why, highlights intended and unintended results, and provides strategic lessons to guide decision-makers and inform stakeholders.*
Demographic and Health Survey (DHS)	*Nationally-representative household surveys that provide data for a wide range of monitoring and impact evaluation indicators in the areas of population, health, and nutrition. (DHS definition)*
Impact evaluation	*An evaluation that provides information about the impacts produced by an intervention. It can be undertaken of a programme or a policy, or upstream work – such as capacity building, policy advocacy and support for an enabling environment. It goes beyond looking only at goals and objectives to also examine unintended impacts.*
Incidence	*The number of new or newly diagnosed cases of a disease that occurs during a given period. An incidence rate is the number of new cases of a disease divided by the number of persons at risk for the disease.*
Index	*A set of related indicators that provides for meaningful, accurate and systematic comparisons regarding performance. For example, the Human Development Index (HDI). Related term: benchmark.*
Interview	*A method of data collection where the interviewer asks the interviewee questions. This methodology is common in qualitative research.*
Median	*A measure of central tendency. The midpoint of a distribution when all the scores are arranged from highest to lowest.*
Multiple Indicator Cluster Survey (MICS)	*International household survey initiative developed by UNICEF to fill data gaps for monitoring the situation of children and women around the world.*
Monitoring	*The collection and analysis of data as the project progresses to assure the appropriateness of the evaluation, its design and participant protections.*
Operationalization	*The process of defining concepts in terms of simple, observable procedures. An operational definition states a concept or variable in such a way that it can be measured and observed. For example, religious devotion may be operationally defined in terms of frequency of attending church. Use of constructs and scales are other common forms of operationalization.*

Randomized controlled trial (RCT)	*A research or evaluation design with two or more randomly selected groups (an experimental group and control group) in which the researcher controls or introduces an intervention (such as a new programme or policy) and measures its impact on the dependent variable at least two times (pre- and post-test measurements). In particular RCTs – which originated in clinical settings and are known as the 'gold standard' of medical and health research – are often used for addressing evaluative research questions, which seek to assess the effectiveness of programmatic and policy interventions in developmental settings.*
Reliability	*Data are measured and collected consistently according to standard definitions and methodologies; the results are the same when measurements are repeated.*
Research	*The systematic approach to the collection and analysis of data and information, in order to generate new knowledge through and answer a specific question or test a hypothesis. Research adds new knowledge and is documented in formats that permit scrutiny and replication.*
Rubric	*An evaluation tool which uses a set of criteria and a rating/scoring guide predetermined by the evaluator(s). Rubrics can be used for evaluating presentations, projects, portfolios, and so on.*
Sample	*The population researched in a particular study. Usually, attempts are made to select a "sample population" that is considered representative of groups of people to whom results will be generalized or transferred.*
Statistical dispersion	*The variability or spread in a variable or a probability distribution. Common examples of measures of statistical dispersion include the variance, standard deviation and interquartile range. Related terms: statistical variability or variation. (Princeton University definition)*
Survey	*A research tool that includes at least one question which is either open-ended or close-ended and employs an oral or written method for asking these questions. The goal of a survey is to gain specific information about either a specific group or a representative sample of a particular group.*
Theory of change	*Explains how activities are understood to produce a series of results that contribute to achieving the final intended impacts. It can be developed for any level of intervention – an event, a project, a programme, a policy, a strategy or an organization.*
Validity	*The degree to which a study accurately reflects or assesses the specific concept that the researcher is attempting to measure. A method can be reliable, consistently measuring the same thing, but not valid. See: reliability.*
Z-score	*A standard score, expressed in terms of standard deviations from the mean.*