United Nations
Educational, Scientific and
Cultural Organization

UNESCO
INSTITUTE
FOR
STATISTICS

Information Paper N. 35
April 2017

open access

Estimation of the numbers and rates of out-of-school children and adolescents using administrative and household survey data

UNESCO

The constitution of the United Nations Educational, Scientific and Cultural Organization (UNESCO) was adopted by 20 countries at the London Conference in November 1945 and entered into effect on 4 November 1946. The Organization currently has 195 Member States and 10 Associate Members.

The main objective of UNESCO is to contribute to peace and security in the world by promoting collaboration among nations through education, science, culture and communication in order to foster universal respect for justice, the rule of law, and the human rights and fundamental freedoms that are affirmed for the peoples of the world, without distinction of race, sex, language or religion, by the Charter of the United Nations.

To fulfil its mandate, UNESCO performs five principal functions: 1) prospective studies on education, science, culture and communication for tomorrow's world; 2) the advancement, transfer and sharing of knowledge through research, training and teaching activities; 3) standard-setting actions for the preparation and adoption of internal instruments and statutory recommendations; 4) expertise through technical cooperation to Member States for their development policies and projects; and 5) the exchange of specialized information.

UNESCO Institute for Statistics

The UNESCO Institute for Statistics (UIS) is the statistical office of UNESCO and is the UN depository for global statistics in the fields of education, science, technology and innovation, culture and communication.

The UIS was established in 1999. It was created to improve UNESCO's statistical programme and to develop and deliver the timely, accurate and policy-relevant statistics needed in today’s increasingly complex and rapidly changing social, political and economic environments.

Published in 2017 by:

UNESCO Institute for Statistics
P.O. Box 6128, Succursale Centre-Ville
Montreal, Quebec H3C 3J7
Canada

Tel: +1 514-343-6880
Email: uis.publications@unesco.org
http://www.uis.unesco.org

ISBN 978-92-9189-207-5
Ref: UIS/2017/ED/TD/4
DataLink: https://doi.org/10.15220/978-92-9189-207-5-en

Table of contents

List of tables

1. Introduction and overview

In the last decade or so, education data from household surveys have been used to complement, supplement and sometimes even substitute for country administrative data on participation and non-participation in schooling.[1] There has been a gradual increase in the use of these household survey data on the demand for schooling, and these data now are used widely in intra- and cross-country comparisons made by the UNESCO Institute for Statistics (UIS), UNICEF, the World Bank and many other providers and consumers of education statistics. Country-level use of household survey education data, however, has been more limited.

The indicators most often produced using household survey data are the net and gross attendance ratios (NAR and GAR), which typically are treated as comparable to the net and gross enrolment ratios (NER and GER) produced using administrative data.[2] However, there are conceptual differences between enrolment and attendance, with both providing imperfect measures of participation in schooling (UIS, 2004).

As the attention to household survey data has increased, and sometimes substantial discrepancies with administrative data have emerged, awareness of the data challenges associated with the use of both administrative and household survey data has grown (see UIS, 2004; UIS, 2010; Omoeva et al., 2013; and Barakat, 2016). In some respects – given the differences in what is being measured by administrative and household survey data (enrolment versus attendance) and the multiple sources of error inherent in each approach, it would be astonishing to find that on the whole enrolment and attendance rates that are comparable. At the same time, given that the attendance and enrolment indicators measure not dissimilar things, the magnitude of difference should not be extreme. Some studies have identified substantial differences between these indicators, however, pointing to serious estimation questions (UIS, 2005; Omoeva et al., 2013).

In addition to providing measures of participation, administrative and household survey data are used to estimate non-participation. The numbers and percentages of out-of-school (OOS) children of primary and lower secondary school age are estimated as a complement to the in-school numbers.

While inevitably there is imprecision in the OOS estimates, the accuracy of the estimates has been improved over time and can be improved further. When more precise population estimates become available, and more accurate approaches to data analysis are developed, estimates of OOS rates are revised retrospectively. For instance, in 2013, as noted by Omoeva et al. (2013), when new population

---

1 Household data also have been used to estimate dropout and repetition rates, as well as adult educational attainment and literacy rates.

2 The NAR may or may not be adjusted to include attendance at any school level, to ensure that, for instance, a child attending secondary school who happens to fall within the age range for primary school, is not counted as out of school in the estimation of the primary NAR.

estimates by the UN Population Division were available, the UIS revised the 2010 estimate of OOS children at the primary level from 61 million to 59 million children.

Improving estimates further requires continued methodological advances and the furthering of efforts to develop authoritative estimates using both administrative and household survey data, rather than having two or more sets of competing and sometimes divergent estimates for a given school year. There are political issues to be addressed in developing joint or blended estimates; this paper, however, focuses on the technical issues.

This information paper characterises and explains variations between administrative and household survey estimates of the numbers and rates of out-of-school children (OOSC) at the primary and lower secondary levels, and suggests ways that data from the two sources might be harmonised. This document also discusses issues surrounding the measurement of exclusion from education for youth of upper secondary age, given this population's competing rights of access to education and the right to work.

Administrative and household survey data efforts, as the UIS and UNICEF (2015) note, differ in purpose, in who collects the data and in what ways, and when and with what frequency. And yet because both data sources provide useful education data on similar topics, the data are examined side by side.

Both the UIS and UNICEF produce estimates on school participation. The UIS produces two sets of participation estimates, using: i) administrative data collected by national governments and reported to the UIS, coupled with data from the United Nations Population Division; and ii) data from multi-topic household surveys – including primarily the Multiple Indicator Cluster Surveys (MICS) sponsored by UNICEF and conducted with national partners, and the Demographic and Health Surveys (DHS) sponsored by USAID and administered by ICF International, with national partners. To date, these household surveys have been used almost exclusively, given the ease of obtaining and analyzing the data, although other surveys also are used. UNICEF, in contrast, uses data from its MICS household survey to produce estimates.

The UIS collects aggregate-level education statistics annually from official sources in a country, using two main surveys: the UIS education survey and the UNESCO, OECD, Eurostat (UOE) education collection. These surveys provide data on education participation, among a host of other topics, and for the ratios, such as NER and adjusted NER, constitute the numerator. For the denominator, the UIS uses data from the UN Population Division for estimates of the school-age population. The UN Population Division estimates are based on data from a range of sources depending on the country – including national censuses, surveys, and vital and population registers.

The two main household survey data sources for education participation are the Demographic and Health Surveys (DHS) and the Multiple Indicator Cluster Surveys (MICS). DHS and MICS both collect data on population, health, nutrition and HIV in developing countries. Data are representative at the national level and generally at the sub-national level as well (by district or region). MICS collects data that are used for monitoring the status of children and women. MICS and DHS cover many of the same topics, although there are variations. For instance, MICS collects birth date data, as well as age in completed years, for all household members, while in the household questionnaire, the DHS collects only age in completed years. DHS does obtain birth date data and reconcile them with age in completed years for selected groups (women aged 15 to 49 years, sometimes men in a similar age range, and children up to age 4). Both the DHS and MICS data can be disaggregated by geographic area (at varying levels of systems), urban/rural residence, household wealth, sex, age and other characteristics.

The UIS presents separate estimates of both enrolment/attendance and OOS children and adolescents rates, by level, derived from administrative data and from household survey data. Data users can access either or both sets of estimates for a given country and period of time. Following the discussion below of differences between administrative and household survey approaches and methodology, this study examines differences between the latest OOS children and adolescents numbers and rates, using UIS-produced data.

2. Data collection methodologies

Differences in definitions: School participation

One key difference between administrative and household survey data is in the definition of what it means to be in school. The UIS estimates the numbers and percentages of children and adolescents who are in and out of school. The age ranges used for primary and lower secondary education in each country are based on the International Standard Classification of Education (ISCED), which may differ from a country's definition of the age range for a given level[3]. Furthermore, only children in formal primary or secondary education are considered to be in school; those in pre-primary or in non-formal education are counted as out of school.

In contrast, household surveys such as MICS and DHS include as in school children attending pre-primary education and may include those attending non-formal education programmes. In fact, while the DHS asks first about whether a household member has attended school, the MICS asks whether the person has attended "school or preschool" to ensure that preschool is understood as part of school. The ways in which the survey questions are worded may differ, as does how levels and grades are specified in each country, but the intent is to capture education participation, regardless of level

[3] UNESCO Institute for Statistics (UIS) (2012). *International Standard Classification of Education: ISCED 2011*. Montreal: UIS. http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf

(including pre-primary education) and including parallel programmes such as accelerated learning and other programmes intended to be equivalent to the formal system[4].

It is important to note that all of these numbers and ratios are gross measures of participation, answering the question whether children go to school at all, or are enrolled, during a given school year. There has been considerable discussion of the limitations of these measures for estimating meaningful participation in schooling. For instance, if a child is recorded as having enrolled in a school but never actually attends, or if a child attends school for only a week in a given school year, he or she still typically is captured as having participated – despite the fact that the child had no meaningful exposure to education in that year. These points are inarguable, and yet the measures of enrolment and attendance are what they are and provide important information. To obtain data on the intensity and consistency of participation, more detailed data on student attendance and absenteeism are required, but these data would measure a different phenomenon than does a binary participation/ non-participation measure.

Methodological approaches

Attendance and enrolment rates and ratios, and their associated OOS rates, are derived from different data sources, using different data collection, sampling and other methods. As a consequence, each set of estimations faces different challenges and is prone to different types of error.

UIS estimations using administrative and population data

Country administrative sources, often the ministry of education or national statistics office, submit education data to the UIS by responding to UIS surveys. The UIS reviews these data for completeness and consistency with international definitions and standards, including ISCED. The Institute also runs checks for errors and compares data within a time series, and sends a report to in-country counterparts to request clarification and updated figures as needed. Sometimes the UIS needs to generate revised estimates, including imputing missing data, and in this process may use "information in national publications, official websites and other reliable sources of data (e.g. household surveys)" (UIS and UNICEF, 2015).

To estimate the OOS rate, the UIS needs the numbers of children enrolled in primary or secondary education and the population estimate for the given age group. The UIS calculates the number and

[4] As defined in the DHS Interviewer's Manual: The term "school" means formal schooling, which includes preschool, primary, secondary, and postsecondary schooling, and any other intermediate levels of schooling in the formal school system. This definition of school does not include daycare, Bible school or Koranic school, or short courses like typing or sewing. However, it does include technical or vocational training beyond the primary school level, such as long-term courses in mechanics or secretarial work. (ICF International, 2015).

percentage of a given age group, including those of primary school age, attending either primary or secondary school. The remaining number of children in that age group, and 100% minus the percentage in either primary or secondary school, constitutes the OOS population.

The intent of administrative data is to capture the entire school-going population. However, in practice, administrative data often do not capture data on students in private and unregistered schools, which may lead to an underestimate of the numbers of students.

Data may be missing and/or inconsistent for the numerator or denominator or both. As the UIS explains in its Frequently Asked Questions (FAQs)[5], data may be missing for a variety of reasons, including not receiving all the necessary data from the country level for the calculation of an indicator for a given year, or the UIS or the country itself identifies data inconsistencies that raise questions about the estimates. Omoeva et al. (2013) note that in each year since 1999, there are missing data on the OOS numbers for almost 40% of the countries captured in the UIS Data Centre, although the UIS does have unpublished estimates for many of these countries. Missing data can present a challenge in estimating both regional and global trends in OOS children and adolescents, particularly when many of the countries with missing data are populous, though the UIS typically does use available estimates in its regional and worldwide OOS estimations, even when those estimates are not precise enough for use at the national level.

As the UIS and UNICEF (2015) point out, problematic population estimates complicate efforts to estimate the numbers of school-age children in a number of countries. This is the case with China, for which there is no recent published estimate of the rate and number of OOS children and adolescents. There are issues in other countries as well, particularly when the latest census was 25 years ago or more, as in the Democratic Republic of Congo. For that country, adding to the difficulty of the outdated census data is the fact that it has been unsettled by ongoing conflict, and related migrations. The most recent available estimate of the OOS primary-aged population dates from 1999.

Because there is error in both numerator and denominator, and because the data come from separate sources (administrative data and population data), sometimes net enrolment rates can exceed 100%. The UIS (2010) found that nearly 25% of the 1,000 values for adjusted NER (ANER) for 200 countries over five years exceeded 100%. The UIS found, though, that these instances of exceeding 100% were by and large in countries with participation rates very close to 100%, so that variability in either numerator or denominator resulted in a rate that exceeds what is possible. The study notes that in these instances, it makes sense to cap the participation rate at 100%.

[5] See http://www.uis.unesco.org/Education/Pages/FAQ.aspx

Administrative data can suffer from the same age data quality issues as do household survey data *(see section below).* First, as with household survey data, in a context where little attention is given to birth dates and exact ages, school staff have no better information on age than do household respondents and may indeed have less information on age. And further, there may be a bias toward saying that children attending primary or lower secondary school are within the target age range for that level. In some places, there are age restrictions for grades, making it problematic for overage children to continue to attend school. These incentives may bias downward children's actual ages in the ages that schools report.

DHS and MICS

The DHS draws a sample usually using a two-stage cluster design, first with enumeration areas drawn from census files, and then households drawn from an updated household listing. The intent of the sampling frame is to capture the target population of interest, which is all women aged 15 to 49 years and children up to age 5 living in residential households (ICF International, 2012). Sometimes areas of a country that are extremely remote or insecure are excluded from the sampling frame.

There are, of course, populations of interest to education that may be excluded from the DHS sampling frame, including children living in households headed by children under age 15; children in orphanages and other institutions such as juvenile detention centers and hospitals; homeless and street children; and nomadic populations. If these populations are substantial, and school participation rates vary appreciably from those among other children, the exclusion may have a noticeable (but difficult to quantify) effect on participation rates, and on the numbers and percentages of OOS children and adolescents.

The UIS (2010) examined age data distributions for ten DHS surveys to see whether there was coherence with the age distribution from census data. While the NAR changed very little in some countries, it changed substantially for several countries.

The UIS (2010) also reviewed non-sampling and sampling errors in DHS. The former includes errors in non-response, survey frame coverage, instrument, interviewer, respondent, processing, estimation and analysis. It is difficult to assess many of these types of errors, with the exception of non-response. DHS typically has extremely high response rates, making these errors of little concern.

The UIS (2010) also found that sampling errors also were low, including for NAR, though it noted that is partly because of the high correlation between numerator and denominator. The review also noted that while there is a very low incidence of missing data for NAR values, DHS does not impute values for those missing data but rather omits those cases. Hence, there is a negligible effect on NAR from sampling error and from the missing data.

Fieldwork for both the MICS and DHS sometimes is not conducted during a single school year. With fieldwork lasting from about four months to as long as eight months, and averaging approximately five to six months, DHS surveys often are not conducted during one school year, but often go into between-school year holidays, or even into a second school year. This timing issue complicates the questions about 'current' year school attendance, particularly in conjunction with the wording of the attendance questions. MICS, in contrast, averages two to four months in the field, minimizing the magnitude of the problem.

The UIS (2010) created a 'blended' approach to estimating the NAR for these surveys that cross school years, including the between-year holidays, to combine the participation rates across two school years. More recent MICS and DHS, however, query school attendance during a specified school year, such as the 2015-2016 school year, rather than asking about 'current' attendance, obviating the need to estimate a blended NAR.

However, even with the improved question about a specified school year, household survey data can suffer from a mismatch between age data and school participation data. In many surveys, including earlier rounds of the DHS and MICS, age data are available only in completed years, as of the time the household was surveyed. A child who was 6 years old, in completed years, at the end of a school year might have been only 5 years old at the start of the reference school year, and in some countries, not eligible for school entry. Barakat (2016) provides a comprehensive review and quantification of the age-school year timing problem, quantifying the magnitude of change in the estimated NAR when age is adjusted back to the start of the school year in question.

More recently, MICS and DHS have developed approaches to collecting birth date data, which allows for a more precise estimation of age at the start of the school year. MICS collects birth date data for all household members using the household questionnaire, with a single respondent. DHS collects age in completed years in the household questionnaire, and for children of school age whose mothers are interviewed in the women's questionnaire, uses children's birth date data to calculate age as of the reference year. For children whose mothers are not interviewed in the women's questionnaire, DHS imputes a birth month.[6]

In instances where data users access published statistics either online or in hard-copy reports, caution should be used in using and interpreting age-adjusted statistics alongside unadjusted numbers. For instance, since DHS does not plan to revise attendance rate estimates retroactively, such as in its STATcompiler online, those estimates from earlier surveys for a given country will continue to reflect the age-unadjusted NAR, and recent surveys will reflect the age-adjusted NAR. Direct comparisons of these age-adjusted and unadjusted statistics over time, therefore, are inadvisable and would be misleading, since the data are not comparable. However, it is important to underscore that analysts

6 Personal communication with ICF International, July 2016.

using the DHS datasets can adjust for age and readily compare survey data over time. In those instances, of course, analysts' age-adjusted statistics will differ from published data for older, age-unadjusted datasets.

3. Phrasing of questions on school participation and on education level and grade attended

Phrasing of household survey questions on education participation

The Education Policy and Data Center (EPDC, 2008) did a thorough review of sets of education questions in 30 household surveys from 1996 to 2005 and made recommendations about core questions, including using standardised language; about assessing reading and literacy skills; and other domains of interest. Reviewed survey modules included those from DHS, MICS and selected national-level surveys. Clearly, the framing and wording of questions – and their translations into local languages – affect data quality, including response error rates.

One key issue for survey programmes such as DHS and MICS is the comparability of data over time and the extent to which time series can be used to measure change, or whether there is noise introduced into results from shifting specifications. This review does not focus in great detail on the range of approaches to collecting education data, but rather notes that changes in question wording can shape participation rates and ratios over time, particularly with the DHS.

The DHS questions on school participation have changed over time, and while overall have the same intent, have slightly different effects on estimates of education participation. The earliest DHS education questions followed this pattern: educational attainment ("highest grade completed" at a given level of schooling) among household members, followed by the question "Is (NAME) still in school?" If the answer to the latter question was yes, the person was considered to be attending school, and the current grade was estimated as the highest grade completed + 1 year.[7] It is important to note that if a person had dropped out of school during the school year surveyed, the answer to the question about 'still' in school would be 'no'. Later versions of the questionnaires, in contrast, capture participation at any point during a given school year, and hence would count the dropout in the

[7] DHS instructions clarified that 'completed' meant successful completion, so that if a person had not passed a grade, the highest grade completed would have been the previous grade that was completed successfully, from which the person was promoted. If, however, the question about completion was not understood as intended, and the person still was in school but was repeating a grade, the highest grade completed + 1 grade for the current level of schooling would be incorrect.

previous example as attending, or having attended, school.[8] As a consequence, the later versions of the questionnaires – those asking about participation at any point during a given school year – all things being equal, would produce higher attendance rates than the older versions of the questions (whether a person is "still in school"), with the differential depending on the extent of within-year dropout. Generally speaking, the differential is expected to be small, but nevertheless, remains.

Current DHS and MICS questionnaires ask whether a child/adolescents attended school at any point during a specified school year (say 2015-2016), and if the answer is yes, counts that person as attending school. MICS also asks about attendance in the previous school year (for instance, 2014-2015), allowing the estimation of dropout and repetition rates. DHS used to include those questions but dropped them some years ago.

Participation in non-formal education and in religious schools

On the whole, household surveys either do not query non-formal education participation and non-academic, religious school participation, but rather ask about participation in the formal schooling system – including pre-school participation. As a consequence, participation rates capture primarily formal school participation. It cannot be ruled out, of course, that respondents interpret or misinterpret questions about schooling and include, for example, participation in a purely religious school as schooling, but it seems likely to be a rare occurrence and one that cannot be estimated without further study.

EPDC (2008) found some national-level household surveys with questions about non-formal education participation, such as apprenticeships and the like. However, internationally-comparable surveys such as the DHS and MICS typically do not collect this kind of information for children and adolescents. There is a question in a recently-fielded MICS education module that includes 'alternative education program' as a level of schooling. The question itself, though, does not ask specifically about non-formal education participation, so it is possible that respondents will not consider alternative education to be education.

This situation is similar to the one faced in Nigeria by the DHS EdData Survey in 2004, as discussed below. In order to ensure that respondents would respond to questions as intended – and to consider non-academic, religious schooling, for one question, and academic schooling in another – there was

[8] The question format following the "still in school" approach asked, instead, two questions: whether the person was "currently" attending school, and if not, whether he/she had attended at any point during the current school year. The NAR, then, was estimated by counting all those who were still in school + had been in school at some point during the current school year. The next version of the education participation questions also captures attendance at any point during the specified school year (usually given by listing the year of the school year), by asking whether the person attended school at any point during the given school year, and capturing the grade attended.

a preamble to frame each question for respondents. The same principle applies to efforts to collect data on non-formal education. It is a category that is difficult to specify across contexts, and so needs to be defined in each country.

There are some challenges to obtaining data on participation in non-academic, religious schools and in non-formal education programs. Some surveys collect information on the type of school attended, which primarily has to do with school ownership (whether a religious institution, government, government-aided, and so on; see EPDC, 2008). If a parent or guardian is asked what kind of school a child attends, one of the difficulties in getting the information sought is the problem of missing data or of response error, as respondents may not know the information or understand the question. Some surveys have asked for school names, and then matched them with Ministry of Education school identifiers, although this effort also can be fraught with difficulty. If school code information is relatively complete, and interviewers have local listings to match against, error rates can be reduced. When these data have a low error rate, and government systems use the same codes, it then is possible to merge data on school quality and other variables with household-based data to produce a rich dataset.

Some household surveys have captured participation in religious schools, as well as in formal academic schools. The 2004 Nigeria DHS EdData Survey, for example, captured participation in formal academic schooling, for all children aged 4 to 16 years and in religious education for Muslim youth. To measure the latter, a parent or guardian was asked the following question: "In a moment, I will ask you about (NAME)'s formal schooling. Before that, I would like to know whether (NAME) attends a Qur'anic school. Does (NAME) attend a school that teaches children the Qur'an, but does not teach academic subjects like mathematics or English?" The intent of this set of questions was to capture participation in both formal and religious education, ensuring that respondents were clear on what type of education was being queried. As it happened, 12% of Muslim children aged 4 to 16 years attended a formal/academic school only; 44% attended both a formal/academic school and Qur'anic school; 34% attended only a Qur'anic school; and 11% attended neither type of school (National Population Commission [Nigeria] and ORC Macro, 2004).

In terms of the estimation of participation and OOS rates among children and adolescents, the key issue is to define what types of participation 'count' and which ones do not, and to ensure that household survey designs support these intentions. Specifically, questions must be phrased clearly, to specify what kind of education/schooling is being asked about; answer categories must be clear and complete; and interviewers and supervisors must be well-trained to ensure that the survey questions are implemented as intended. At present, there is no widespread survey program that systematically includes data collection on participation in religious education.

4. Definition and interpretation of indicators of school participation

As discussed above, similar – but not the same – indicators of participation can be estimated using either administrative and population data, or household survey data. The UIS Data Centre includes estimates OOS rates, separately, using both administrative and household survey data. The remainder of this review focuses on the estimation of OOS numbers and rates at the primary and lower secondary levels, starting with the specification of the indicators.

Out-of-school rate, primary (OOSR primary): This indicator is calculated as 100% minus the adjusted NER, which is the percentage of children of primary school age enrolled in either primary or secondary school.

Out-of-school rate, lower secondary: This indicator is calculated as 100% minus the total NER, or the percentage of adolescents of lower secondary school age enrolled in either primary or secondary school.

The out-of-school rate also can be estimated using household survey data and the indicators are similar to the estimations, except that they refer to attendance data – to whether children and adolescents attended school at any point during the reference school year. The parallel indicators are:

Out-of-school rate, primary: This indicator is estimated as the percentage of children of primary school age who attended primary or secondary school during a specified school year.

Out-of-school rate, lower secondary: This indicator is estimated as the percentage of children of lower secondary school age who attended any level of schooling (primary or higher) during a specified school year.

In using household survey data, the UIS estimates children's ages at the start of the reference school year. If a household survey has birth date data, as recent rounds of MICS does, those data are used to determine children's ages at the start of the school year. If the survey does not have birth date data, the UIS either does not adjust age (if data on the majority of children were collected less than six months into the school year) or estimates all household members' ages as (age – 1) if the majority of children's households were surveyed more than six months into the school year. If ages are adjusted, both numerator and denominator include the population in the adjusted age range (UNICEF and UIS, 2015).

5. Problems inherent in combining enrolment figures with population estimates

To estimate OOS numbers and rates, UIS combines data from administrative sources with population data to produce numbers and rates. In a sense, the rates actually are ratios, given that the data come from two quite different sources and while an indicator such as the NAR is intended to be a percentage and is interpreted as one, it really is a ratio of data from two different sources on the – in theory – same population. Given, though, that administrative data and population data incorporate different types of error and face different data quality issues, it is useful, perhaps, to think of the rate as a ratio instead.

One illustration of this point is that, as discussed above, for some small states and for some countries with very high participation rates, net school participation rates sometimes exceed the theoretical maximum of 100% – suggesting that no children or adolescents are out of school, which is extremely unlikely. The UIS makes adjustments to these rates, often capping them at 100% participation.

The UIS regularly updates its estimates of OOS children and adolescents when revised population projections become available, producing rates that differ from those published earlier. Given that imprecision is inherent in the estimation effort, this process of revisiting and revising numbers is an iterative process. In the context of the Sustainable Development Goals (SDG) and with the continuing drive to get all children and adolescents into school and keep them there, it is imperative to be flexible and make adjustments as necessary.

The further from a recent census a country is the more challenging it is to estimate population overall and by individual ages. A year or two away from a recent census, it is comparatively straightforward to estimate population trends, with smaller error overall; whereas further out, in the absence of updated data, assumptions about trends have increasing influence, and the degree of inaccuracy increases substantially. In addition, where different sets of assumptions or models are used, the extent of variation across estimates – made by the UN Population Division and those by a range of groups within a country – increases as well.

The UIS (2010) investigated methodological approaches and data quality issues in estimating education participation, focusing on UIS estimates and DHS estimates. Among other things, the review pointed out the challenges in using UN Population Division data for UIS purposes, as the estimates are made for five-year groups, rather than for individual ages. Because the population of primary or lower secondary age does not conform to these groupings, it is necessary to estimate population by single ages, from the five-year groups. The UIS (2010) found that the demographic technique typically used, the Sprague Interpolation, produced estimates that differed, sometimes appreciably, from the population estimates for 5 of the 19 countries covered by the World Education Indicators (WEI) work. The UIS (2010) suggested that the WEI consider a different approach to estimation.

6. Population estimates from different sources and their effects on out-of-school rates and numbers

Population estimates are technically challenging and often politically sensitive. Even within countries, various ministries may differ in their internal estimates of population numbers, and differ with the UN Population Division estimates as well. In these contexts, determining the official numbers may not be straightforward.

The worldwide estimates of the numbers and rates of OOS children and adolescents for a given year can change substantially, depending on population estimates. With updates from the UN Population Division, these numbers can change, and have, by a few million children. For instance, just before EPDC (2013) published its review of data challenges in measuring the extent of OOS, UIS updated its estimates, changing the 2010 estimate of OOS primary-aged children from 61 to 59 million. In short, population estimates have substantial impact on OOS population estimates.

The UIS (2010) found meaningful changes in enrolment rates estimated using updated population estimates from the UN Population Division, issued in its World Population Prospects. In other words, when census estimates were updated, the participation rates changed appreciably, suggesting the sensitivity of participation rates to these changes in population estimates. The UIS (2010) recommended caution in using population estimates based on census data that are several years old or older, and suggested that these older estimates "may be a source of error, accounting for the discrepancies between adjusted NER and NAR figures."

The quandary remains: accurate data are required for the estimation of OOS numbers and rates, and yet remain notoriously difficult to obtain, particularly in insecure and conflict environments. UIS, for instance, includes notes about the limitations of its data, pertaining to both enrolment and population data, to explain that some areas are excluded or that an alternate source of data is used.

7. The advantages and disadvantages of using administrative and household survey data to estimate school participation

Both administrative and household survey data have advantages and disadvantages, and ideally, can be used to provide points of comparison for one another to identify ways to improve precision. Data from both sources can result in over- or underestimates of participation rates. As the UIS glossary of indicators describes the limitations on the OOS rate: "The administrative data used in the calculation of the rate of out-of-school children are based on enrolment at a specific date which can bias the results by either counting enrolled children who never attend school or by omitting those who enroll after the reference date for reporting enrolment data. Furthermore, children who drop out of school after the reference date are not counted as out of school. Discrepancies between enrolment and population data from different sources can also result in over- or underestimates of the rate. Lastly,

the international comparability of this indicator can be affected by the use of different concepts of enrolment and out-of-school children across countries."[9]

Similarly, household survey data typically count a child with as little as one day of participation as attending school, regardless of whether the child has dropped out; sampling and non-sampling error contribute to imprecision in the estimates; the target population often excludes some children (those not living in households); and age data are problematic, whether adjusted back to the start of the school year or from estimated birth date.

In sum, both data sources are flawed. And yet, they are the indispensable, as they provide the only widely-available data on OOS numbers and rates.

Administrative data often are available every year, providing timely information on participation and allowing for measurement of trends over time; when complete, cover the entire country without the concern of sampling error; have the great advantage of being linked to the Education Management Information System (EMIS), where data on school, teacher and student characteristics (mainly age and sex) can be linked to participation numbers; when data are of high quality, have strong country-level and international buy-in.

Administrative data in the UIS database can be disaggregated by sex but not by other characteristics. Household survey data, in contrast, provide OOS data by critical disaggregates – including not only sex, but also household wealth, educational attainment of the household head, urban/rural location, and other characteristics. Given that Hattori (2014) found that household wealth was a statistically significant determinant of OOS status in 56 of 63 countries, having this information within and across countries provides crucial insights into who and where OOS children are – and can contribute to an understanding of the ways in which to bring OOS children into, or back into, school.

Another context in which household survey data are critical is in conflict-affected environments. The UIS and GEMR (2016) estimate that for the school year ending in 2014, "areas in 32 countries affected by armed conflict were home to 21.5 million or 35% of the global number of out-of-school children," at the primary level, and that "15 million or 25% of out-of-school adolescents of lower secondary age and 26 million or 18% of all out-of-school youth of upper secondary age lived in conflict-affected areas." The UIS and GEMR (2016) also point out for countries where precise estimates are not available – including conflict-affected countries and areas, such as the Democratic Republic of Congo and Afghanistan – global estimates do use the best available figures for numbers of OOS children, adolescents, and youth in these areas.

9 See http://www.uis.unesco.org/Pages/Glossary.aspx

It is terribly difficult to estimate the numbers and rates of OOS in conflict-affected environments, where administrative statistics are unavailable, outdated or incomplete. In these contexts, while they also have limitations, household survey data provide an invaluable source of information on OOS rates.

Having data from both the school level and household level on comparable, but not the same, indicators of participation, provides a fuller, richer set of data that can be used to improve understanding of OOS numbers and rates. Each data source can be used to help improve the other, as well, pointing out where there are anomalous differences that bear further investigation.

8. Discrepancies in school participation estimates from different data sources

Sources of difference

There are myriad reasons for discrepancies in school participation and OOS numbers and rates between administrative and household survey data. Various reviews (see UIS, 2004; UIS, 2015; EPDC, 2013; Barakat, 2016) have attributed differences to:

- conceptual differences between enrolment and attendance, so in effect administrative and household surveys measure different things;
- the definition of target groups and who is excluded (primarily non-government school data, for administrative data; and those children not living in households, in survey data);
- the timeliness of data (such as outdated census data; and infrequent household survey data collection) and the extent of missing data, which complicate household survey and administrative data comparisons and limit the years for which data align sufficiently;
- the magnitude of error in population estimates, and in administrative data (especially on student ages); and
- the definition of being in school (the inclusion of NFE and of pre-primary, in household surveys; and their exclusion in UIS estimates; although it should be noted that the former groups can be excluded in estimating participation rates, when using datasets rather than looking just at published statistics).

More recently, there has been a strong critique of the quality of administrative data and the suggestion that not only are household survey data of better quality, on the whole, than are administrative data, but that the latter are systematically and deliberately skewed. Sandefur and Glassman (2014) compare administrative and household survey data from a group of African countries, focusing on health (vaccination rates) and education (primary school participation rates) data. Sandefur and Glassman argue that the there is a systemic “misrepresentation of national statistics” as a result of “the incentives of data producers to overstate development progress”.

Sandefur and Glassman describe the DHS as the "gold standard" for social statistics. While acknowledging that household survey data also have "imperfections," the authors argue that their argument requires only that household survey data errors are "independent of, not necessarily smaller than, the errors in administrative data systems" (p. 1). This is a curious argument to make if the intent of the paper is to address the possible approaches to improving data quality and accuracy, working with the admittedly imperfect data sources available.

There is evidence that one of the main sources of large discrepancies has been the failure to adjust children's ages from household surveys. More recent MICS and DHS household surveys obtain data on birth dates so that children's ages can be estimated at the start of the relevant school year. In the past, however, there was a mismatch between age (in completed years, at the time the household was surveyed) with the start of the school year in question. Children who actually were too young, or too old, to be attending a given level of schooling at the start of the school year were included in the numerator and denominator of the estimated NAR, biasing participation rates downward.

Barakat (2016) sets out a persuasive argument for necessity of adjusting children's age data from household surveys that record age in completed years (rather than having data on birth month and year). As Barakat points out, the magnitude of change in participation rates from unadjusted to adjusted ages is at a level that is meaningful for policy and must be addressed, despite the frequent assertion that the differences between survey-timing adjusted and unadjusted participation rates are minimal. Barakat correctly observes that "speculative explanations have been proposed for what is actually a measurement artefact," and proposes solutions to the problem.

Barakat also examines data that Sandefur and Glassman (2014) use to make their points about administrative and household survey data. Sandefur and Glassman (2014) argue that the comparison of household survey data and administrative data on participation reflect a systematic and deliberate upward bias in administrative participation rates, in response to incentives attached to showing progress under free primary education. The authors compare administrative and survey data from before and after the declaration of free primary education and find a considerably higher rate of change in administrative data than in household survey data. For instance, the authors found that between 2003 and 2008, Kenyan administrative data showed almost 8 percentage points change in NER, whereas during the same timeframe, the DHS found no change in the NAR.

Barakat (2016) deftly points out that once survey data are age-adjusted, the differences between administrative and household survey data on participation that Sandefur and Glassman (2014) identified actually disappear (Kenya) or are reduced considerably (Rwanda). Interestingly, in both those countries, the second DHS was conducted quite a bit later in the school year than was the first DHS. This difference in timing of the DHS relative to each school year increased the age problem and hence depressed the participation rates at the second point in time compared with the first point in

time. Once adjustments were made, the gap in rates of change between administrative and survey data disappeared.

Fair (2016) examines the other household survey data cited by Sandefur and Glassman (2014), and finds that in 12 of the 21 survey pairings used to support the argument that administrative data are biased upward deliberately, the second DHS was conducted later in the school year than was the first DHS; in four instances, there effectively was no difference in timing; and that in five cases, the first survey was done later in the school year than was the second survey. It is clear that participation rates are very sensitive to age data, and that where the age data for the second DHS come from later in a given school year than they do for the first DHS, it is expected that once age data are adjusted for each DHS survey pair, the NAR for the second survey will rise at a higher rate. In the majority of instances that Sandefur and Glassman (2014) use to critique the quality of administrative data and suggest deliberate inflation of participation rates, it is quite possible that instead, what the authors captured was merely the effects of survey timing and age-unadjusted data.

In sum, a host of factors contributes to differences – sometimes substantial ones – between participation estimates from administrative and household survey data.

Country data comparisons: Primary OOS numbers and rates

In many instances, there is congruence between OOS estimates using administrative and population data, and household survey data *(see **Table 1**).*[10] In Indonesia, for example, using household survey data and administrative data, the UIS estimates the same OOS primary rate of 5% in 2003 and 4% in 2008.

In other cases, there is a substantial differential between administrative and household survey data. For example, in Ghana in 2015 there is a 22 percentage point difference in the estimated OOS rate, which contrasts strongly with the near-congruence of the 2009 estimate (only 1 percentage point difference; see Table 1). The Ghana OOSCI country study (UIS and UNICEF, 2012a) examined differences between country EMIS and DHS data from 2008-2009. Using age-unadjusted data, the country report found quite different participation rates than those presented in Table 1: the primary NER was estimated at 89% (EMIS data) while the primary NAR was estimated at 74% – a 15 percentage point difference in favor of administrative data. The GER was estimated at 95% (EMIS) and the GAR at 111% (DHS) – a difference of 16 points, in favor of household survey estimates.

[10] The UIS household survey estimates use age-adjusted data, where necessary.

Table 1. UIS database primary out-of-school rate and number estimates, household survey and administrative data

Country	Year	Household survey OOS rate (%)	Administrative survey OOS rate (%)	Percentage point difference (surveys–admin)	Number of OOS children based on administrative data
Bangladesh	2006	19	4	15	680,000
	2011	13	5 (2010)	8	907,000 (2010)
Ethiopia	2000	64	60	4	6.4 million
	2005	50	39	11	4.9 million
	2011	38	22	16	3.3 million
Ghana	2009	25	24	1	815,000
	2015	30	8	22	319,000
Indonesia	2003	5	5	0	1.3 million
	2008	4	4	0	1.1 million
Kenya	2003	25	25	0	1.3 million
	2008	13	17	4	1 million
Malawi	2004	13	2	11	40,000
	2010	9	2 (2009)	7	48,000 (2009)
Mali	2006	54	43	11	900,000
	2013	46	37	9	1 million
Nigeria	2007	43	29	14	6.6 million
	2008	31	35	4	8.4 million
	2011	30	34 (2010)	-4	8.7 million
Pakistan	2007	29	29	0	5.7 million
	2012	36	29	7	5.7 million
Yemen	2006	30	24 (2005)	6	885,000 (2005)

Notes: Data downloaded on 19 July 2016. Data presented for years in which both household survey and administrative data are available. Percentages are rounded to nearest percentage; numbers are rounded. Age data from household surveys have been adjusted by UIS to start of school year.

A comparison of these differentials illustrates some of the common reasons for difference in administrative and household survey data estimates. First, both the administrative and household survey have 'bad' age data. The household survey data were not adjusted back to the start of the school year, so many of the children who were 6 years old at the time the household was surveyed were only 5 years old at the start of the school year – and so should have been excluded from the estimation. On the other end, many children who were 12 years old at the time of fieldwork were 11 at the start of the school year, and so should have been included in the estimation of NAR. The fact that age was not adjusted results in an underestimate of the NAR. The EMIS data may reflect inaccurate age data, as well. There may be an inherent bias toward estimating ages for children in primary school within the 6 to 11 age range, which results in an overestimate of NAR.

In some countries, over the years, the percentage point gap between survey and administrative data estimates narrow, while in others, they increase. In Ethiopia, the percentage point gap – with a higher OOS rate estimated from household survey data – increases from just 4 points (2000) to 16 points (2011).

Data discrepancies between administrative and household estimates in a handful of populous countries (such as India and Nigeria) and how those are handled, have substantial effects on both country and worldwide OOS rates.

India

Estimating school participation in India has been an enduring challenge over the years. EPDC (2013) found that data from the 2006 DHS estimated that 17% of children aged 6 to 10 years were out of school, compared with 5%, the UIS figure at the time for the same year and 2% from official statistics. The difference in these rates amounts to 14.6 million more OOS children. Omoeva et al. (2013) note that the official starting age for primary school is 5 years in 21 states and 6 years in 19 states, and that the number of years in the primary cycle ranges from four to five. As a consequence, the official age ranges for primary range from 5-8, 5-9, 6-9 to 6-10 years. Hence, any age range applied to the whole of the country – whether in surveys or in administrative data – invariably will be a mismatch with the system in a number of states.

Omoeva et al. (2013) explore age adjustment to align children's ages with the start of the reference school year, noting that the 2006 India National Family Health Survey (NFHS) collected data over a nine-month period (December 2005–August 2006) and spanned two school years. The survey did ask about children's attendance during a reference year, 2005-2006, ensuring that all data were for a given school year rather than for two, but children's ages were reported in the household questionnaire at the time the household was surveyed. As a consequence, many children reported to be age 6, for instance, at the time of the survey would have been only 5 years old at the start of the school year in question. In addition, children reported to be 11 years old at the time of the survey would have been 10 years old and within the age range for primary education. The EPDC report adjusted children's ages downward for this NFHS dataset in India, producing an overall non-attendance rate among children aged 6 to 10 years that was 7 percentage points lower than the non-adjusted rate using the same dataset.

Furthermore, the timing of the school year varies across the country, complicating efforts to adjust age data from household surveys back to the start of the reference school year. The UIS and UNICEF (2016) note that across India's 29 states and 7 union territories, the start of the school year in India ranges from January to November, with four starting in January, four in February, 11 in April, one in May, ten in June, four in July, and two in November. Administrative data use the start of the school year in each state as the reference date for children's ages, while the NFHS surveys use the first day

of April, across households regardless of the official beginning of the school year in a given state. The authors suggest that the preferred approach would be to adjust each child's age to completed years as of the time of the start of the school year in each state, to avoid error at the margins – the inclusion of those children officially too young to have started primary school, and the exclusion of those already outside the official age range.

The UIS and UNICEF (2016) also note a common issue with administrative statistics – the completeness of data on all schools. There is variability in the schools covered by official statistics, from government schools, private aided schools, private unaided schools and private unrecognised schools.

Sigdel (2014; cited in UIS and UNICEF, 2015) attributes India's interest in participating in the Global Initiative on Out-of-School Children, at least in part, to the desire to document the factors behind its widely variable participation and OOS numbers and rates. Two main constellations of factors emerged from a country team review of the evidence: the sample design and timing of the survey; and definitions of OOS children. Despite using the same sampling frame, two surveys found quite different OOS rates. The Social and Rural Research Institute of IMRB International (SRI-IMRB) survey in 2009-2010 estimated an OOSR of 4% (primary education) and 5% (lower-secondary education); while the National Sample Survey Office (NSSO) surveys in 2007-2008 estimated the rates at 11% and 8%, respectively. As UIS (2015) notes, in a populous country such as India, the greatest possible precision is critical where a percentage point difference in the OOS rate can produce a difference of about 1 million OOS children and adolescents.

In sum, a host of issues shapes participation and OOS estimates in India, and continues to produce wide variation in both numbers and rates of OOS. The UIS and UNICEF (2015) note that the government of India is working on a clear, national definition of dropout, which will address some of the discrepancies in estimates. Remaining factors, though, include for household surveys: sampling frames; survey timing and the issues surrounding age (adjusting age to the start of the school year; and fitting the age range for a given state rather than a national average); and for administrative statistics: completeness of data on all school types, and the perennial challenges of population estimates – although the recent 2011 census has improved estimates considerably.

Both census and survey data suffer from the problem of age heaping. There is a well-known tendency to tendency for a household respondent to give a 'rounded' age, say one ending in 0, 5, 10, or 15, rather than another nearby age. In addition, in some household surveys, such as the DHS and MICS, there are incentives to interviewers to shift, or displace, some potential respondents and children out of the age ranges that trigger the use of the women's questionnaire, the men's questionnaire, and the collection of a considerable amount of data on children aged 0 to 5 years. The extent to which heaping and displacement occur varies from country to country and survey to survey, and census to census (see Pullum, 2006). Whatever the extent of the age heaping, age data can be smoothed out, or 'unheaped,' to reflect better the actual distribution of ages in a population. However, various

approaches produce varying results and affect indicators to differing degrees, and in the case of household survey data, interact with efforts to adjust age back to the start of the school year.

Nigeria

The UIS administrative and household survey data estimates show a notable gap between administrative and household survey data estimates of primary OOS rates in 2007 (14 percentage points), but a narrower gap the following year (4 percentage points), and then a gap in the other direction (-4 percentage points) in 2010-2011. Interestingly, the distance between estimates narrows both because the OOS percentage rose, in the administrative estimate, and declined, in the household survey estimate. The 4 percentage point gap found in 2008 is relatively small in terms of percentage but substantial in absolute numbers of OOS children.

The OOSCI Nigeria Country Study (UIS and UNICEF, 2012b) relied on DHS data from 2003 and 2008, in its study of the OOS phenomenon in Nigeria. This study's age-adjusted estimate of the primary OOS rate is comparable to the DHS-based estimate in the UIS database (30% and 31%, respectively). The report estimates that 7.3 million children of primary school age were out of school in 2008, compared with the UIS administrative data-based calculation of 8.4 million children (based on a 35% OOS rate). While these 2008 percentage estimates are close, a 4 percentage point difference produces a difference of 1.1 million children OOS – a non-trivial differential.

The study also found, using DHS age-adjusted data, that in 2008 26% of lower secondary school-age adolescents were out of school (comparable to the UIS household survey estimate). There is no point of comparison from administrative data, as UIS does not have lower secondary OOS estimates.

As in the case of India, Nigeria illustrates how in a populous country, small differences in OOS rates can produce meaningful differences in estimated OOS populations. These differences, aggregated up across regions and to a worldwide total, have substantial effects on worldwide estimates of OOS numbers and rates.

Country data comparisons: Lower secondary OOS numbers and rates

A comparison of the patterns for primary and lower secondary OOS rates shows one main difference *(see **Table 2**).* For the ten countries profiled where there are differentials at the primary level, the household survey OOS rate estimates are higher than administrative rates. In contrast, at the lower secondary level, the reverse is true – administrative data produce higher OOS rates in 11 of the 18 comparisons.

As was the case with the data at the primary level, there are many instances where there is little or no difference between the administrative and household survey OOS rates (Bangladesh, Indonesia and Mali). It is notable that for Indonesia, the primary OOS rate estimates were very similar as well.

Ghana data show the same pattern at the lower secondary level as at the primary level – with a small gap in OOS rates in 2009 and a large difference in 2015.

In Ethiopia, the gap between administrative and survey data estimates from 2000 and 2005 (-7 and -8 percentage points), increases to -12 percentage points in 2011. The pattern of a growing gap is the same as in the primary OOS data, but interestingly the differential goes the opposite direction. At the primary level, the household survey estimates of OOS children are higher than the administrative data, going from 4 to 11 to 16 percentage points at those three points in time. It seems plausible that this reversing pattern has something to do with the age data, particularly at the margins – whether the differential is attributable to age data from the survey, administrative, or population data sources.

Most striking, though, are the differentials in Pakistan. While at the primary level there were minimal differences in the 2007 and 2012 OOS estimates (0 and 7 percentage points, in favor of the survey data), the gap between administrative and survey data at the lower secondary level is massive: -27 percentage points in 2007 and -25 points in 2012, with the differential in favor of the administrative data.

The substantial gap raises questions about the data sources, although it is unexpected that estimates from the same data sources might be problematic for one level of schooling and not another. The UIS estimates use the Pakistan DHS from 2006-07 and 2012-13. UNICEF (2013) includes among its sources the Pakistan Social and Living Standards Measurement Survey (PSLM) 2007–2008 and its Household Integrated Economic Survey 4 (HIES), conducted in the four largest provinces of Pakistan – Punjab, Sindh, Balochistan and Khyber Pakhtunkhwa (KP); and the 2007-2008 MICS conducted in Punjab. Not included in the sample for any of these surveys were Gilgit-Baltistan, Pakistani Administered Kashmir, Islamabad Capital Territory and the Federally Administered Tribal Areas. Using these data, UNICEF (2013) found lower secondary OOS estimates using household survey data that were more comparable to the administrative data – 65% OOS, versus the administrative data estimate of 60% OOS – than to the survey data estimates in the UIS Data Centre – at just 29%.

It is not uncommon for surveys in Pakistan to exclude certain insecure areas. As documented in Omoeva et al. (2013; citing NIPS and Macro International, 2008), the 2006-2007 DHS in Pakistan excluded the Federally Administered Tribal Areas, Federally Administered Northern Areas, and Azad Jammu and Kashmir areas that made up about six percent of the total population.

It would be well worth an investigation into factors in these substantially different estimates – whether related to national representativeness or other factors. Indeed, for Pakistan, if survey estimates of the OOS rate at lower secondary are more accurate than are administrative data, there are substantial implications for OOS numbers for the country.

Table 2. UIS database lower secondary out-of-school rate and number estimates, household survey and administrative data

Country	Year	Household survey OOS rate (%)	Administrative survey OOS rate (%)	Percentage point difference (surveys – admin)	Number of OOS children and adolescents based on administrative data
Bangladesh	2006	32	32	0	3.1 million
	2011	19	23 (2010)	-4	2.2 m (2010)
Ethiopia	2000	59	66	-7	4 million
	2005	41	49	-8	3.4 million
	2011	30	42	-12	3.7 million
Ghana	2009	12	18	-6	297,000
	2015	21	7	14	121,000
Indonesia	2003	23	23	0	3 million
	2008	20	21	-1	2.8 million
Kenya	2003	9	3	6	51,000
	2008	4	1 (2006)	3	13,000 (2006)
Malawi	2004	19	23	-4	293,000
	2010	16	22 (2009)	-6	319,000 (2009)
Mali	2006	59	54 (2009)	5	534,000 (2009)
	2013	52	50	2	562,000
Nigeria	2007	32	----	----	----
	2008	27	----	----	----
Pakistan	2007	29	56	-27	6.4 million
	2012	28	52	-25	6.1 million
Yemen	2006	34	39 (1999)	-5	516,000 (1999)

Notes: Data downloaded on 19 July 2016. Data presented for years in which both household survey and administrative data are available. Percentages are rounded to nearest percentage; numbers are rounded. Age data from household surveys have been adjusted by UIS to start of school year.

9. Recommendations on improving the quality and comparability of education statistics from different data sources and how to consolidate estimates

Approaches to analysing household survey data, alongside administrative data, have evolved considerably over the last decade. One critical point in the process was when UNESCO and UNICEF (UIS, 2005) collaborated to develop a joint estimate of OOS children, as described here: "The number of out-of-school children reported by UNESCO and UNICEF has often differed substantially. In 2004, UNESCO published the figure of 104 million out-of-school children in its Education for All Global Monitoring Report. In the same year, UNICEF published the figure of 121million in its flagship report, The State of the World's Children. The new joint estimate presented in this report is different – 115 million children are out of school."

Advancements documented in that report include the adjusted enrolment/attendance rate, which ensured that, for instance, children of primary school age who were attending secondary school were not counted as out of school; and the combination of administrative and household data to estimate more accurate OOS numbers. The first change reduced the OOS estimate by 11 million children; the second increased the OOS estimate by over 20 million children (UIS, 2005).

At this juncture, a decade on from the UIS report, seems an appropriate time to revisit estimation approaches, and to open access to information, consistent with the UN call for a data revolution. The UIS Data Centre documentation addresses approaches used for estimations only in general terms (EPDC, 2013), constraining review and comment by other data producers and users. Without doubt, the particular approaches used in any given context – especially with the data quality assurance process the UIS employs in an iterative way with countries and their ministries of education – can be sensitive. The manipulation and use of household survey data are less sensitive, in some respects, and yet not under any one group's control. The DHS and MICS programmes confer, and sometimes approaches diverge. While there are extensive consultations on the estimation of various indicators, these survey programs determine their approaches, in the end.

Efforts to continue to improve the accuracy of estimates of participation, toward the end of achieving universal access, would benefit immensely from consensus on how to handle and manage both administrative and household data. To come to consensus, though, requires considerably more sharing of approaches and their limitations across actors – data producers and users alike. There must be transparency in how estimations are done: how definitions are used in data specifications; how missing data are handled; how data quality issues are assessed and addressed.

In addition to the data openness suggested above, several specific recommendations follow from this review:

1. Household survey age data must be adjusted back to the start of the reference school year. As others have noted, the non-adjustment for age explains a non-negligible part of discrepancies between participation rates estimated using administrative/population and household survey data (Barakat, 2016). The age adjustment problem is less pressing for recent, and hopefully future, household surveys – at least for MICS and DHS – since they obtain and use birth date data. In the use of older survey data, however, it is strongly recommended that children's ages be adjusted and that the adjustment be well documented in data files, in online and print reports, and so on. Not doing age adjustment, across the board, results in underestimates of participation rates. Furthermore, it is critical to adjust ages before looking at changes in participation rates over time. When age data are not adjusted, as was the case in the Sandefur and Glassman (2013) comparison of administrative and household survey data, erroneous conclusions about administrative and/or household survey data may be drawn, with serious policy implications (Fair, 2016). Many approaches to age adjustment have

been suggested and used (see Barakat, 2016 for a thorough review), but further technical and operational review is needed to determine the most expedient, workable approach to adjusting for age in surveys lacking birth date data. One promising approach is to use a truncated age range, leaving out the 'problematic' ages at the margins. For instance, if the target age range for primary school is 6 to 12 years, estimate in-school and out-of-school rates for children aged 7 to 11 years, as even with household survey timing issues, the narrow age range covers children who were of primary-age at the start of the reference school year. One drawback to this approach is the question of comparability with enrolment rates and ratios. Other considerations include whether disaggregation will become problematic with a reduced sample size; and the question of how age heaping interacts with the truncated age range. Further investigation is needed to explore the advisability of using a truncated age range.

2. Investigate and document the accuracy and consistency of age and birth date data. Undertake a systematic review of MICS household survey data that include both age in completed years and birth date data to determine the extent of missing data and the consistency between the data sources. It would be useful to examine patterns of missing data and the extent to which using age in completed years for some school-age children, and birth date data for others, introduces error or bias into NAR estimates. Especially if complemented by a qualitative examination of field-based experience and lessons about the ease or difficulty of these kinds of questions, such a review would offer insights into the feasibility and advisability of collecting birth date data in single-respondent household questionnaires.

It is important to remember that while it is appealing to call for birth date data to be collected in all household surveys, to allow the estimation of age at the start of the school year, getting good or even 'good enough' data on birth dates is extremely difficult. Household surveys such as the DHS go to extraordinary lengths to obtain age (in completed months and/or years) and birth date data that are consistent. Interviewers are trained to ask respondents to collect birth certificates, immunization and other records for themselves and their children. Interviewers ask first about a woman's date of birth, and then age, and then must reconcile the responses so that they are consistent. A respondent having trouble with age may be asked to think about her age at the time of marriage or giving birth for the first time, or to relate her age to someone else in the household whose age is known more precisely, or to say how old she was at the time of a momentous event such as an earthquake, flood, or change in political regime. The interviewer works with the respondent to estimate her age, then must reconcile it with birth date. The same process is undertaken with children up to age 5, because of the precision required for anthropometry, immunization, and so on.

In addition, attention must be given to the extent of age heaping and to approaches to redistributing ages to smooth out the peaks and valleys, to approximate the distribution of 'true' age in the population. Multiple-topic surveys, such as the DHS, have focused to date more on age heaping and displacement among women aged 15 to 49 years and children up to age 5, than on among school-age

children and youth. Yet age heaping can have a substantial effect on attendance rate estimations, particularly at the margins of age ranges for a given level of schooling.

Greater scrutiny is needed with administrative age data, for the same types of error that bedevil household surveys, and to try to improve the accuracy of reported age data for students. Age heaping, for instance, also occurs in administrative age estimates. In contexts where there are regulations governing age for a particular grade, this will be especially challenging, as children who are over age for the grade may be moved to accelerated learning programmes or face sanctions.

3. In countries with both large populations and considerable variation in the age range for primary school, such as India, consider estimating participation and non-participation rates and numbers according to each state/region's system rather than for one age range. Aggregating these rates across states to the national level will produce a considerably more precise estimate of the extent of the OOS problem in these populous states, in their regions, and worldwide.

4. Develop and implement consensus on what it means to be 'in school,' and clearly document departures from the definition, in reports and datasets. The UIS excludes pre-school and non-formal education participation from its estimates of participation, while many household surveys include pre-school and/or non-formal education in their estimates of school participation. As the UIS (2015) found, two surveys in India used the same sampling frame, but different definitions of OOS and produced quite different estimates of OOS numbers and rates. As the UIS notes, in a populous country such as India, where a percentage point difference in the OOS rate can produce a difference of 1 million children in the OOS numbers.

5. Consider publishing an upper and lower bound, or confidence interval, for OOS populations by level. There is great appeal to and power in having a single point estimate of OOS numbers and rates, even if these numbers have to be revised at intervals, when improved population data become available or when a better approach to adjusting age is developed, and so on. Yet, as EPDC (2013) phrased it, it is important to "understand the bounds of uncertainty" around measures of exclusion. Improving single point estimates is a laudable and necessary goal, but in the meantime, between multiple sources of conflicting data points and the error inherent in any of these estimates, some sense of the magnitude of uncertainty, or the confidence interval, is important.

10. Problems in defining, measuring and interpreting data on exclusion from upper secondary education

The interaction between the right to education and the right to work is a dynamic one, when it comes to youth. UN declarations include:

- The right to free, compulsory primary education for all, an obligation to develop secondary education accessible to all, in particular by the progressive introduction of free secondary

education, as well as an obligation to develop equitable access to higher education, ideally by the progressive introduction of free higher education.

- Everyone has the right to work, to free choice of employment, to just and favorable conditions of work and to protection against unemployment.

The latter right is circumscribed, somewhat, by age and by constraints on child labour. The Convention No. 138 on the Minimum Age for Admission to Employment, 1973 sets out a framework that defines work and sets out age limitations, as follows:

	The minimum age at which children can start work	**Possible exceptions for developing countries**
Hazardous work Any work which is likely to jeopardise children's health, safety or morals should not be done by anyone under the age of 18	18 (16 under strict conditions)	18 (16 under strict conditions)
Basic minimum age The minimum age for work should not be below the age for finishing compulsory schooling, which is generally 15	15	14
Light work Children between the ages of 13 and 15 years old may do light work, as long as it does not threaten their health and safety, or hinder their education or vocational orientation and training.	13-15	12-14

Clearly, there are overlaps between the age groups covered by the right to education and the right to work, and these make it difficult to measure and discuss OOS numbers and rates at the upper secondary level of schooling. These constraints are not at issue with younger adolescents, those typically of lower secondary age, given that their right to do light work, as defined above, is set out as acceptable as long as it does not harm them or interfere with education or training; and the basic minimum age "should not be below the age for finishing compulsory schooling," which typically goes through lower secondary.

It is possible to collect data on participation at the upper secondary level, and the UIS presents statistics on OOS numbers and rates for this level, but the collection and interpretation of these data are problematic (UIS and GEMR, 2016). ISCED defines the upper secondary level of education as programmes forming the second/final stage of secondary education; these programmes may be

either general or vocational, and some allow direct access to ISCED level 4 (post-secondary non-tertiary education) and/or level 5, 6 or 7 (short-cycle tertiary education to Master's degree or equivalent level). In short, there is a tremendous diversity of education that is included at the upper secondary level. There often are different providers of education at the upper secondary level, including the ministry of education and others, and programs may include general academic schools, the vocational-technical schools, and other specialised schools. Hence, information on participation is spread across a number of government entities, complicating the effort to collect complete administrative data.

Adding to the complexities is the SDG goal of access to technical, vocational and tertiary education, which spans upper secondary and tertiary programmes. As the current discussion of indicators notes, getting data on a participation rate in technical-vocational education programmes for youth aged 15 to 24 years is a complicated matter.

The UIS database has missing data on upper secondary OOS numbers and rates for a substantial number of countries, perhaps reflecting the difficulty countries have in reporting on the indicator. Also, the UIS database does not include OOS estimates for upper secondary education using household survey data, so there is no secondary source of data on participation rates at this level.

Household survey data could be used to provide insight into what youth of upper secondary age are doing – whether attending school or a vocational-technical training programme, working or being married and raising a family. The DHS and MICS surveys collect education and employment data for women aged 15 to 49 years and men in a similar age range, which covers the upper secondary age group. Census data also can provide data on education participation and work. In contexts where countries have difficulty estimating participation rates, household survey data offer much-needed insight.

The critical question, though, is how to interpret and use the data. The OOS numbers and rates for upper secondary-age youth do not have the same clear, understandable interpretation as do these statistics at the primary and lower secondary levels. It is possible to collect and report on rates of participation in various formal programs (though still challenging to get complete information across a wide range of programs and providers), on participation in the workforce (again, with the complexities those estimations entail), and on not being engaged in either pursuit (perhaps by marital status), but it is not clear how the rights to education and to work would be interpreted. If one rate goes up, or down, is that a desirable thing? Perhaps the central question is what youth of upper secondary age want to be doing, and whether they are able to pursue their education, work, and family goals – a question difficult to address with participation rate data alone.

References

Barakat, B. (2016). "Sorry I forgot your birthday!": Adjusting apparent school participation for survey timing when age is measured in whole years. *International Journal of Educational Development*, 49, 300-313.

Education Policy and Data Center (EPDC) (2008). "How (well) is education measured in household surveys? A comparative analysis of the education modules in 30 household surveys from 1996-2005". IHSN Working Paper 002. International Household Survey Network.

Fair, K. (2016). "Age Adjustment Techniques in the Use of Household Survey Data." UIS Information Paper No. 34. Montreal: UNESCO Institute for Statistics. http://dx.doi.org/10.15220/978-92-9189-206-8-en

ICF International (2012). *Demographic and Health Survey Sampling and Household Listing Manual.* MEASURE DHS, Calverton, Maryland, U.S.A.: ICF International.

ICF International (2015). *Demographic and Health Survey Interviewer's Manual.* Rockville, Maryland, U.S.A.: ICF International.

National Institute of Population Studies [NIPS] Pakistan & Macro International (2008). *Pakistan Demographic and Health Survey 2006-07*. Islamabad: NIPS Pakistan and Macro International Inc.

National Population Commission [Nigeria] and ORC Macro (2004). *Nigeria DHS EdData Survey 2004: Education Data for Decision-Making*. Calverton, Maryland, USA: National Population Commission [Nigeria] and ORC Macro.

Omoeva, C., B. Sylla, R. Hatch and C. Gale (2013). *Out-of-School Children: Data Challenges in Measuring Access to Education.* fhi360. Education Policy and Data Center (EPDC), Washington, DC.

Pullum, T.W. (2006). "An assessment of age and date reporting in the DHS surveys, 1985-2003". DHS Methodological Reports 5. Macro International, Inc., Calverton, MD.

Sandefur, J. and A. Glassman (2014). *The Political Economy of Bad Data: Evidence from African Survey and Administrative Statistics*. Access at: http://dx.doi.org/10.7910/DVN/26712

Sigdel, S. (2014). "Process behind selecting and calculating out-of-school children estimates in India". Background paper prepared for *Global Report on Out-of-school Children*. Montreal: UNESCO Institute for Statistics.

UNESCO Institute for Statistics (UIS) (2004). *Guide to the Analysis and Use of Household Survey and Census Education Data.* Montreal: UNESCO Institute for Statistics. http://www.uis.unesco.org/Library/Documents/hhsguide04-en.pdf

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

UNESCO Institute for Statistics (UIS) (2010). "Measuring Educational Participation: Analysis of Data Quality and Methodology Based on Ten Studies". UIS Technical Paper No. 4, Montreal: UNESCO Institute for Statistics. http://www.uis.unesco.org/Library/Documents/tp4-en.pdf

UNESCO Institute for Statistics (UIS) and Global Education Monitoring Report (GEMR) (2016). "Leaving No One behind: How Far on the Way to Universal Primary and Secondary Education?" Policy paper 27/Fact sheet 37. Montreal and Paris: UIS and GEMR. Accessed at: http://unesdoc.unesco.org/images/0024/002452/245238E.pdf

UNESCO Institute for Statistics (UIS) and UNICEF (2012a). "Ghana Country Study for the Global Initiative for Out-of-school Children". http://www.uis.unesco.org/Education/Documents/OOSCI%20Reports/ghana-oosci-report-2012.pdf

UNESCO Institute for Statistics (UIS) and UNICEF (2012b). "Nigeria Country Study for the Global Initiative for Out-of-school Children". http://www.uis.unesco.org/Education/Documents/OOSCI%20Reports/nigeria-oosci-report-2012.pdf

UNESCO Institute for Statistics (UIS) and UNICEF (2013). *Out-of-School Children in the Balochistan, Khyber Pakhtunkhwa, Punjab and Sindh Provinces of Pakistan.* New York: UNICEF. http://www.uis.unesco.org/Education/Documents/OOSCI%20Reports/pakistan-oosc-report-2013-en.pdf

UNESCO Institute for Statistics (UIS) and UNICEF (2015). *Fixing the Broken Promise of Education for All: Findings from the Global Initiative on Out-of-School Children*. Montreal: UNESCO Institute for Statistics. http://unesdoc.unesco.org/images/0023/002315/231511e.pdf

UNESCO Institute for Statistics (UIS) and UNICEF (2016). *Estimating the Number of Out-of-School Children: Methodological Problems and Alternative Approaches - India Case Study*. Montreal: UNESCO Institute for Statistics. http://www.uis.unesco.org/Education/Documents/oosc-statistical-report-india-case-study.pdf

UNICEF and UNESCO Institute for Statistics (UIS) (2015). *Global Out-of-School Children Initiative Operational Manual.* New York: UNICEF. http://allinschool.org/wp-content/uploads/2015/12/F_UNICEF1017_OOSCI_manual-web.pdf