United Nations
Educational, Scientific and
Cultural Organization

UNESCO
INSTITUTE
FOR
STATISTICS

Information Paper N. 34
April 2017

open
access

Age adjustment techniques in the use of 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 978-92-9189-206-8
Ref: UIS/2017/ED/TD/5/REV.1
DataLink: https://dx.doi.org/10.15220/978-92-9189-206-8-en

Table of contents

Page

List of figures

List of tables

1. Introduction

It is extraordinarily hard, in many developing countries, to obtain accurate birth data and age data for children of school age. As Pullum (2006) notes in a review of age and date of birth reporting in the Demographic and Health Surveys (DHS): "…the kinds of problems that appear in DHS surveys are largely endemic in settings where ages and dates are not important in the daily lives of most people and do not necessarily indicate faulty data collection procedures. For example, [age] heaping is repeatedly observed in all censuses and surveys in some countries, and cannot be eliminated even with special training of interviewers".

In many countries, birth certificates may be a rarity, in large part because there is no culture of documenting or attending to age. In addition, the system of registering births may be not fully functional or less available in rural or remote areas. In some cases, even if services are available, they may be expensive and difficult for families to access.

These difficulties with age data plague not only census and survey efforts, but also school registration and administrative records. In fact, in some countries, between the challenges in estimating number of children of primary school age attending either primary or secondary school, net rates are not used widely – or not used with confidence – and instead gross participation rates are used.[1] In short, age data are enormously problematic in many developing country contexts, for both surveys and administrative data.

The DHS is a health, population, nutrition and HIV survey, which collects education data primarily as a background characteristic for the estimation of fertility rates and other indicators of interest in its specified fields. Age data, in whole years, are collected in the household survey questionnaire, administered to a responsible adult household member. These age data incorporate some reporting error, without doubt, and these are the age data available for use in calculating education indicators, such as net attendance rates (NAR). In contrast, the DHS collects birth date data and undertakes an exhaustive review to ensure the congruence of age in completed years and birth date data for populations of greatest interest – adult women and men eligible for the individual interviews and for children in selected age ranges (usually up to age 5).

[1] There is error in the denominator as well, in the estimation of the numbers of children of primary school age, but it is less problematic than the numerator, in this instance, given the rather more predictable or imputable numbers of people by age in a population.

The Multiple Indicator Cluster Surveys (MICS) from 2009 onwards collect birth date data for all household members using its household questionnaire. The MICS first asks about birth date, then age in completed years, for each household member. In the household questionnaire, though, there are no instructions for interviewers in how to check and reconcile inconsistencies between birth date and age in completed years, nor are there instructions for supervisors and editors to check these data in the household questionnaire.[2] MICS faces the same challenging context for getting accurate age data as does the DHS, and perhaps more, given the effort to get birth date data from a single household respondent who may not know other household members' dates of birth. In addition, there is likely to be a high rate of inconsistency between household members' ages and birth dates.

Barakat (2016) notes that it is the norm for household surveys and census efforts to collect only age data in completed years, rather than birth date data, or both pieces of information. In International Public Use Microdata Series (IPUMS), less than one-half of the latest datasets include birth date data "and many of those contain more missing values than data entries". This point underscores the data quality issues inherent in trying to collect birth date data for all household members from one informant, and the impracticality of determining whether these valid answers actually are correct.

A related technical note delves more deeply into the reasons for discrepancies between participation rates produced using administrative and household survey data; that discussion in its entirety will not be repeated here. Instead, this note focuses on the survey timing issues that necessitate the adjustment of children's age data to allow the estimation of school participation rates.

2. Early intimations of survey timing effects on participation rates

In the late 1990s and early 2000s, the close examination of education data from multiple-topic surveys such as the DHS for the express purpose of examining children's participation in schooling was in its infancy. DHS and other similar surveys were designed primarily to collect health, population, nutrition, and HIV data, and the approach of the DHS, including length of fieldwork and timing of data collection relative to the school year, were not part of the survey calculus. The DHS and similar surveys used educational attainment mainly to examine variations in fertility, immunization, and other rates rather than as variables of interest themselves. Data on children's participation in schooling were presented in DHS reports, but by age range framed by demographic ranges – such as 6-10 years, 11-15 years and so on – rather than by the target age range for the primary and secondary levels.

2 *MICS Instructions for Interviewers and MICS Instructions for Supervisors and Editors* (2014). MICS does, however, have instructions on reconciling birth date data and age for children covered by the 0-5 questionnaire.

In the early 2000s, when both the DHS and the linked but separate DHS EdData Program were working in some of the same countries and both were generating NAR estimates, it became clear that survey timing and children's reported ages were producing surprising variations in participation estimates.[3] The Uganda DHS (UDHS) was in the field from 28 September 2000 through 3 March 2001, and the Uganda DHS EdData Survey (UDES) was in the field from 10 April to 22 July 2001. The UDES included the subset of UDHS households with children age 5 to 18 years at the time of the UDHS.

Both surveys used the same question about current school attendance to measure generally whether a child was in school. A child who attended occasionally, or usually went to school but was absent recently, was counted as attending school, as the intent was to determine whether the child attended at all.

The official age range for primary school in Uganda is 6 to 12 years, and both surveys used current attendance data to estimate the NAR. The UDHS found an NAR of 79%, compared with the UDES NAR of 87%. The UDHS data referred to the 2000-2001 school year, and the UDES to the 2001-2002 school year, since the school year in Uganda runs from late February to early December. Nevertheless, this 8 percentage point differential was striking.[4]

Notable are the differences in the age-specific attendance ratio (ASAR) among 6-year-olds, with the UDHS showing under 60% of 6-year-olds attending either pre-primary or primary school, and the UDES showing 78% attending (Uganda Bureau of Statistics and ORC Macro, 2002; and ORC Macro, 2004). Attendance rates are most sensitive at the margins, and in particular around the official starting age. Because many of the children who were age 6, in completed years, between September 2000 and December 2000, when that school year ended, were age 5 when the reference school year began in February 2000, the school year for which current attendance data were collected, they were not actually eligible to attend primary school that year. The last week or so of the DHS fieldwork was done

---

[3] The DHS EdData Program, 1999 – 2004, was funded by USAID and implemented by ORC Macro. DHS EdData conducted four nationally-representative education household surveys that were linked to the DHS in those same countries. Household surveys, capturing data on the demand for schooling, including the reasons for non-attendance and factors in dropout; household expenditures on schooling; parents' views on the quality of education; and other topics, were conducted in Uganda (2001), Zambia (2002), Malawi (2002), and Nigeria (2004). In Zambia and Nigeria, anthropometric (height and weight) data also were collected, allowing the examination of the relationship between stunting and wasting, and school enrolment age and performance. Reports and datasets are available through the website of the successor project, EdData II, at: https://www.eddataglobal.org/countries/index.cfm?fuseaction=showdir&pubcountry=NG&statusID=3&showtypes=0

[4] See Uganda Bureau of Statistics and ORC Macro (2002); Uganda Bureau of Statistics and ORC Macro (2002); and ORC Macro (2004).

during the new school year, so children who were listed as 6 (completed) years old at the start of that school year definitely were only 5 years old at the start of the reference school year.

Other factors clearly may have been in play as well. Despite the intent of the 'current attendance' question, it is possible that respondents interpreted current attendance not to mean attendance in general, but rather, specifically, attendance at that point in time or even on that very day. Since, depending on the timing of the household visit, it was either late in the school year, during the between-year holidays, or even early in the 2001 school year, respondents might have answered the question in a way unintended by the survey. Clarification about questions is provided only if respondents ask or seem unsure, and clarification instructions make clear what the intent of the question is. Furthermore, assuming the question was understood as intended, the likelihood of a child dropping out during the school year grows as the year continues, so the responses may have captured a higher dropout incidence than did the UDES. It is notable, though, that the UDES found a dropout rate of just 2.5% among children aged 6 to 12 years, suggesting that dropout does not explain much of the difference between the UDHS and UDES NAR estimates (Uganda Bureau of Statistics and ORC Macro, 2002).

Soon after the completion of the DHS EdData Survey in Uganda, there were discussions between ORC Macro, the implementer of both the DHS and DHS EdData Programs, and the funding agency for both programmes, USAID; and other education statistics experts, including those from the UNESCO Institute for Statistics (UIS). Survey timing and age adjustment problems were explored for several countries, and early suggestions were made about how to adjust children's ages back to the start of the reference school year. At the time, USAID decided that given the uncertainties about how best to approach the problem, the technical challenges inherent in a statistical adjustment and concerns about the degree to which these modifications would be understood and embraced by policymakers and data users, and the implications for adjusting statistics for many years' worth of DHS education data, it would not support the adjustment process at that time.

Looking beyond those early discussions at ORC Macro, to the timing of the DHS surveys in the country, is instructive. All surveys collected education participation data, albeit with differently-worded questions.[5] **Table 1** shows the timing of each DHS survey over the years:

---

5 In 1995, household respondents were asked about the highest grade completed, then asked "Is (NAME) still in school?" In 2001, household respondents were asked the same questions about completion, then asked about NAME's current attendance, as discussed above. In 2006 and 2011, household respondents were asked whether (NAME) attended school at any time during the 2006 school year, and if so, what grade was attended.

Table 1. Timing and duration of Uganda DHS surveys

Survey	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec
1995			X	X	X	X	X	X				
2000-01*	X	X	X						X	X	X	X
2006					X	X	X	X	X	X		
2011						X	X	X	X	X	X	X
School year		X	X	X	X	X	X	X	X	X	X	X

* In 2000-01, fieldwork started in September 2000 and ended March 2001. The other surveys were done during a single calendar year.

Apart from demonstrating that each UDHS has had a different starting and ending month, Table 1 also shows that surveys take five to six months, on average, to administer. Even the survey most closely aligned with the start of the school year, the 1995 UDHS, requires age adjustment back to the start of the school year, given how many months it takes to do fieldwork. And clearly, it is inadvisable to compare participation rates across multiple surveys in a given country without making age adjustments, given the variability and extent of the age adjustment problem. Whether there has been change in participation rates, and if so, in which direction and in what magnitude, cannot be addressed using unadjusted age data.

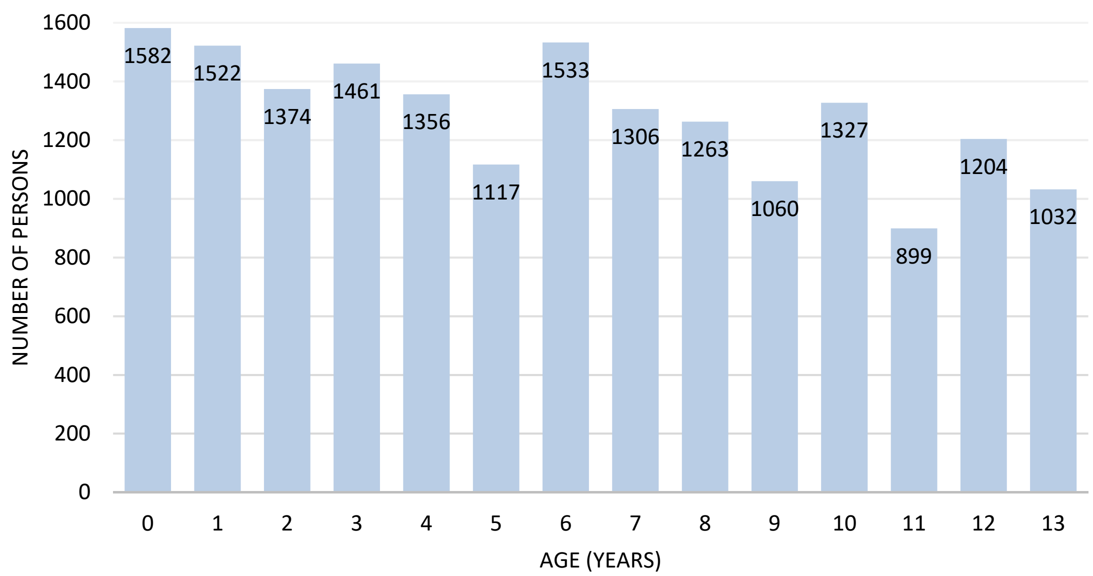
MICS surveys are in the field for about two to four months, which suggests that the survey timing effect may be less problematic than for a survey that takes longer to collect data. In addition, as discussed below, from about 2009 onward, MICS has collected birth date data for household members and when possible, uses these data to determine children's ages at the start of the reference school year.

3. Age displacement and heaping

Another age problem has to do with individual age distribution, rather than with survey timing. The UDHS data are instructive on the issues of age heaping and displacement. The DHS Program monitors age distributions and is aware of the tendency to give a 'rounded' age, say one ending in 0 or 5, rather than another age, for household members in the information provided by a single respondent to the household questionnaire. Also, in the case of the DHS, 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.

In the case of the UDHS household age distribution *(see **Figure 1**)*, there is a substantial difference in the total numbers of children listed as age 5 and those listed as age 6, from 1,117 to 1,533, an increase of 416 children or over 37% (Uganda Bureau of Statistics and ORC Macro, 2002). Pullum (2006), in a review of DHS age data, notes that it is common for children's ages to be given as even, rather than odd numbers, which may play a part in the distributional shift. The heaping at age 10 is consistent with Pullum's point. For purposes of the estimation of the NAR, looking at participation among children aged 6 to 12 years, the combination of interviewers wanting to avoid additional questions for these children that would be asked if they were age 0 to 5 years, and the natural tendency for respondents to give even rather than odd numbers, may have inflated the numbers of household members said to be age 6. This shift is particularly influential since participation rates at the age of official entry to primary school – 6 years in Uganda – are sensitive to timing, and the 2000-2001 DHS was done primarily very late in the school year and over the between-year break.

Figure 1. Uganda DHS household age distribution, 2000-2001



4. Approaches to adjusting ages from household surveys

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) and proposes adjustment approaches that both build on and depart from measures used to date. 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.

Over the years, a variety of age adjustment approaches has been discussed and developed. For some time, though, it was common not to address the mismatch between age data and the timing of the school year, with participation rates and ratios estimated and presented without age adjustment; or to note the problem and not adjust the data (Barakat, 2016). Doubtless the issues associated with the need to adjust data and statistics already in use and published, combined with the conviction that differences between adjusted and unadjusted data were small, were factors in not addressing adjustment.

Barakat (2016) classifies adjustment approaches to date as proportional and in discrete one-year adjustments. This review will use the same terms and reprise Barakat's discussion. One approach used to date was developed by the Education Policy and Data Center (EPDC; 2009), which adjusts the proportion of children who have had a birthday since the start of the school year according to the percentage of the school year that elapsed before fieldwork was done. Another approach has been what Barakat (2016) calls binary, which either adjusts age according to the whether the median observation is in the first or second half of the school year or rounds off the proportion of the year that has passed to either 0 (no change in age in completed years) or 1 (age in completed years – 1).

For most of the recent MICS surveys, there are data on both household members' ages and dates of birth. MICS uses the latter to estimate age at the beginning of the school year. If only year of birth or age data are available, MICS adjusts age based on the timing of the survey, either subtracting 1 from age in completed years (if the survey was done six months or longer after the school year began), or using age as it is (if the survey was done less than six months after the school year began).[6]

As of 2016, the DHS estimates the NAR for all children listed in the household schedule, estimating age as of the start of the reference school year (or country-specific cut-off date). DHS uses the date of birth for children of women who are covered by the individual interview and for children whose mothers are not interviewed, imputes a month of birth[7].

5. The imperative of age adjustment

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

[6] Personal communication from UNICEF, 12 July 2016.

[7] Personal communication from ICF International, 15 July 2016.

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.

Below is an examination of the 21 survey pairings Sandefur and Glassman (2014) used to support the argument that administrative data are biased upward deliberately. **Figure 2** shows that in 12 of the 21 survey pairings, 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 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 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.

There are complexities in interpreting the direction that age adjustment would move the estimated participation rates, based on survey timing, but the categories below capture general expected trends. Each group is organized according to whether the changes in the NARs, once adjusted for survey timing and age, likely would diminish or increase the relative change in NAR between the administrative data and household survey data. Sandefur and Glassman's (2014) find a gap of between 21 percentage points, and negative 10 points (higher NAR using survey data).

NAR likely would increase more for first survey than second survey, which might support authors' argument about higher rates of change in administrative data

The five survey comparisons include:

1) Cameroon 1991 versus 2001: The fieldwork for the second DHS started earlier in the school year than did the first DHS (January versus April), but it is worth noting that the second DHS was in the field eight months, through August, while the first DHS was in the field only six months, through September. While it is possible that an age-adjusted NAR for the first survey would

increase more than that for the second DHS, it is not definite, given the fieldwork duration and considerable overlap between the two fieldwork periods.

2) Niger 1992 and 2006: The school year runs from October to July, and the first survey was in the field from March to June, compared with January to May for the second survey.

3) Ethiopia 2005 versus 2011: The school year runs from September to July, and the first survey was in the field from April to August, compared with December to June for the second survey. As a consequence, it is expected that the age-adjusted NAR would increase by a larger percentage for the earlier survey than for the later survey. At the same time, the second survey was in the field longer, seven versus five months, and there are overlaps in fieldwork timing, so the differential increase may not be large.

4) Guinea 1999 versus 2005: The first survey was in the field during months 8 and 9 of the school year, and in the first month of the between-year holiday. The second survey was done earlier, in months 5 to 9 of the school year.

5) Tanzania 1992 versus 1996: The second survey was fielded earlier in the school year than was the first survey, so it is likely that the 3 percentage point gap Sandefur and Glassman (2014) found in favor of the rate of change using surveys would be diminished or disappear. Either way, there would not be a case to be made, with this comparison, for deliberate manipulation of the administrative data.

It is notable that for several of these countries, the gap Sandefur and Glassman (2014) estimated between changes in administrative statistics and changes in survey data is very small: about 2 percentage points in Niger, Ethiopia and Guinea. Even if this gap widens slightly with the age-adjusted NARs, it would remain a modest differential. The age-unadjusted gap is larger in Cameroon (9 percentage points), meriting further exploration.

NAR likely would increase more for second survey than first survey: likely to counter authors' argument about bias in administrative statistics

These 12 survey comparisons include:

1) Kenya 1998 versus 2003: The first DHS started early in the school year (in February, for a school year starting in January), while the second DHS started in April. Once age is adjusted to the start of the school year, it would be expected that the NAR for the 2003 survey would increase more than that for the 1998 survey. Barakat (2016) found when age was adjusted, both the differential between the DHS NARs and the DHS NARs and the administrative NERs disappeared.

2) Rwanda 2004-2005 versus 2011: The timing of the school year changed between the first and second surveys, complicating the comparison. The 2004-5 DHS started seven months into the school year, and fieldwork lasted five months. In comparison, the 2011 DHS started fieldwork nine months into one school year, then continued over the between-year break, and ended in March, three months into the following school year. Without doubt, adjusting the NAR for age at the start of the reference school year will affect the 2011 NAR profoundly.

3) Ethiopia 2000 versus 2005: The 2000 DHS started six months into the school year, compared with nine months into the school year for the 2005 survey. Furthermore, the first survey was in the field four months, compared with five months for the second survey. The NAR for the second survey, when age adjusted, is expected to rise considerably more than an age-adjusted NAR for the first survey.

4) Kenya 2003 and 2008: The 2003 survey was in the field from months 4 to 9 of the school year, compared with the 2008 survey, in the field from month 11 of one school year, over the between-year holiday, and into month 2 of the subsequent school year. As with the Rwanda case, the age-adjusted NAR clearly will rise at a greater rate for the second survey than for the first.

5) Benin 1996 versus 2006: The school year runs from October to July. The 1996 survey was in the field from months 9 and 10 of the school year, and in the first month of the between-year holidays (August). The 2006 survey was in the field only during the holidays following the school year and into the start of the next school year, from August to November, with the reference school year being the one that was completed before the households were surveyed.

6) Eritrea 1995 versus 2002: The first survey was in the field from months 1 to 5 of the school year, and the second from months 7 to 10, and then in the first month of the between-year holidays. The age-adjusted NAR will rise at a higher rate for the second survey than for the first.

7) Namibia 1992 and 2000: The school year runs from January to December, and the first survey was in the field from July to November, while the second was fielded from September to December. Given that the gap Sandefur and Glassman (2014) identified between changes in participation rates between administrative and household survey data was less than 1 percentage point, it is likely that this ‘gap’ will be eliminated or reversed.

8) Burkina Faso 1999 versus 2003: The first survey was fielded between months 2 and 6 of the school year, while the second survey was done from month 9 in one school year, over the between-year break and through month 2 of the new school year. The age-adjusted NAR would be expected to change considerably more for the second survey than for the first, and given that the ‘gap’ between the change in administrative rates for that period and the change in

household survey-based NAR is only 0.2 percentage points, it is expected that the gap in rates of change would be reversed.

9) Tanzania 1999 versus 2004: The first survey was in the field months 9 to 11 of one school year, compared with the second survey, fielded during from month 10 of one school year, into the between-year holidays, and ended in month 2 of the following year. Sandefur and Glassman (2014) found a higher rate of change in household survey data than in administrative data; once data are age adjusted, this differential rate of change in favor of surveys likely would increase.

10) Nigeria 1999 versus 2003: Both surveys started in March, but the second survey was in the field twice as long (six months instead of three), and further into the school year. Hence it would be expected that the rate of change in the age-adjusted NAR would be higher for the second survey than for the first.

11) Nigeria 2003 versus 2008: The first survey was in the field during months 7 to 11 of the school year, plus the first month of holidays; the second survey was fielded during months 10 and 11 of the school year, over the between-year holidays, and two months into the next school year.

12) Tanzania 1992 versus 1996: The first survey was in the field during months 7 to 11 of the school year, the second survey was fielded during months 9 to 11 of the school year.

For these 12 comparisons, the rate of increase in NAR is expected to be considerably higher for the second survey than the first, diminishing, eliminating, or even reversing the gap, as Barakat (2016) estimated for one of the Kenya and the Rwanda survey comparisons.

Fieldwork timing substantially similar for both surveys: no expected effect on gap between changes in NAR using survey data versus NER using administrative data

These four survey comparisons include:

1) Burkina Faso 1993 versus 1999; and 2) 1999 versus 2003: For both comparisons, the fieldwork period is very nearly the same for both surveys, with the second DHS starting one month before the first survey did.

3) Senegal 2005 versus 2010: The first survey was in the field during months 5 to 9 of the school year, and the second survey during the first 7 months of the school year. Given the difference in duration of the fieldwork, and considerable overlap of the months, it is not clear the extent to which age-adjusted NAR will change differentially for the two surveys.

4) Lesotho 2004 versus 2009: The school year runs from March to December, and there was little difference in survey timing (the first survey was fielded from September to January and the second from October to January). It is worth noting, though that Sandefur and Glassman (2014) found that there was a higher rate of change for household survey data than for administrative data, so even before adjusting for age, there is no gap suggesting mishandling of administrative data.

In addition, even before adjusting age, in six of the paired comparisons Sandefur and Glassman (2014) made between administrative and household survey data, the percentage point change was greater for the survey than for the administrative data. Adding that understanding to the need to adjust for age bolsters the strong recommendation that in order to avoid regrettable policy conclusions, such as those made by Sandefur and Glassman (2014), it is crucial to be aware of survey timing and to adjust age data accordingly. Not to adjust age data is to embrace the vagaries of chance survey timing, and to mis-estimate participation ratios, as well as to estimate incorrectly the extent of the out-of-school population.

Figure 2. Timing of DHS surveys relative to the school year, in response to Sandefur and Glassman (2014)

Country	School year or survey	Month																
		9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1
Kenya	School year					January - December												
	DHS 1998						February - July											
	DHS 2003								April - September									
Rwanda	School year 2004-05	September - June																
	DHS 2005							March - July										
	School year 2011					January - November												
	DHS 2010-2011	September - March																
Ethiopia	School year	September - July																
	DHS 2000						February - May											
	DHS 2005								April - August									
Cameroon	School year	September - June																
	DHS 1991								April - September									
	DHS 2011					January - August												
Burkina Faso	School year		October - June															
	DHS 1993				December - March													
	DHS 1999			November - March														
Kenya	School year					January - December												
	DHS 2003								April - September									
	DHS 2008			November - February														
Benin	School year		October - July															
	DHS 1996										June - August							
	DHS 2006												August - November					
Burkina Faso	School year		October - June															
	DHS 2003										June - November							
	DHS 2010									May - December								
Eritrea	School year	September - June																
	DHS 1995	September - January																
	DHS 2002							March - July										
Niger	School year		October - July															
	DHS 1992							March - June										
	DHS 2006					January - May												

Country	School year or survey	Month																
		9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1
Ethiopia	School year	September - July																
	DHS 2005								April - August									
	DHS 2011				December - June													
Guinea	School year		October - June															
	DHS 1999									May - July								
	DHS 2005						February - June											
Senegal	School year		October - July															
	DHS 2005						February - June											
	DHS 2010		October - April															
Namibia	School year					January - December												
	DHS 1992											July - November						
	DHS 2000													Sept - December				
Burkina Faso	School year		October - June															
	DHS 1999			November - March														
	DHS 2003										June - November							
Tanzania	School year					January - December												
	DHS 1999													Sept - November				
	DHS 2004		October - February															
Tanzania	School year					January - December												
	DHS 1992		October - March															
	DHS 1996											July - November						
Nigeria	School year	September - July																
	DHS 1999							March - May										
	DHS 2003							March - August										
Nigeria	School year	September - July																
	DHS 2003							March - August										
	DHS 2008										June - October							
Tanzania	School year					January - December												
	DHS 1992											July - November						
	DHS 1996													Sept - November				
Lesotho	School year							March - December										
	DHS 2004													September - January				
	DHS 2009														October - January			

6. Recommended adjustment methods

While it is clear that in using household survey data on participation, age adjustment is necessary, the development of approaches to do so has been halting and incomplete on the whole. As Barakat (2016) notes, many of the early efforts to identify a workable approach acknowledged the limitations of the proposed approaches – as with the binary adjustment. The UIS was an early proponent of identifying and rectifying survey timing issues and proposing solutions, and has been instrumental in identifying the limitations of various approaches and pressing for new methodologies (see UIS, 2004; UIS, 2005; and UIS, 2010, for instance.)

Moving forward, questions remain as to how to approach an age adjustment fix. Is there one best approach? Should there be one preferred method to be adopted by all, or a range of methods, depending on context, need, and effectiveness? What complexities are introduced if, in addition to adjusting age back to the start of the reference school year, age heaping and displacement also need to be addressed? How replicable, user-friendly and understandable should the approach be, given that it will be used by statisticians and must be supported by decisionmakers? And what happens to the overall undertaking if there is not widespread buy-in from influential consumers and users of these statistics? What happens if there are competing sets of numbers in widespread use, from age-unadjusted to a range of age-adjusted estimates?

Barakat (2016) offers several possible approaches to age adjustment, and tests their effectiveness. These approaches are summarised below.

1. Random/proportional age assignment: This approach is a variation on the mean binary adjustment discussed earlier, and that is to “calculate a weighted average of the NAR with and without shifting all ages, weighted by the elapsed fraction of the school year.” This approach has the advantage of making adjustments according to the interview date, for each child, relative to the start of the school year.
Assessment: Barakat himself argues this approach likely would underestimate NAR, based on the expectation that children who were of age to start school are more likely than younger children to be in school, such that random assignment is not appropriate (see below).

2. Age assignment ordered by attendance status: This approach is a refinement of the random proportional age assignment, as it proposes to adjust the likelihood of adjusting age up or downward according to schooling status – based on the argument that at the start of primary school, too-young children are less likely to be in school than children of school age, and too-old children (at the upper end of the primary grades) are less likely than younger children to be in school. Barakat (2016) suggests that the NAR then would be estimated using adjusted ages, and “the mid-point of the interval between the estimated bounds” would be a good estimate for overall NAR.

Assessment: This approach is understandable to most consumers of data, but involves considerable judgement calls to be made, and would result in some variation across statisticians. How to determine how much to adjust NAR by age would present challenges.

3. Regression on interview dates: Barakat describes a maximum likelihood estimation of the ASAR at the beginning of the school reference year, though he notes the risk that these ASARs, aggregated into an NAR, might be creating an NAR "out of thin air." Further, Barakat notes that the regression estimates would be biased if data collection dates correlate with determinants of attendance – such as location. And further, Barakat notes that the sampling weights, linked to location and other characteristics, also complicate the effort.

Assessment: Barakat suggests that regression not be the standard adjustment approach, as it may fail to capture the relationship between the variables. Indeed, to do a regression on interview dates would require having all the necessary variables, correctly specified, and would open up the process to a wide range of interpretations and estimations. It also should be noted that this approach would be not only very difficult to do consistently, but also challenging to explain and justify to policymakers. It would seem that this approach is more suited to research efforts into the effectiveness of whatever 'best' method is developed for age adjustment, than to being used as the adjustment approach across the board.

4. Truncated NAR: This approach would leave out the youngest and oldest ages, and estimate NAR for a slightly-revised group. If the primary age range is 6 to 12 years, the truncated NAR would be estimated for those aged 7 to 11, which eliminates the problem of age adjustment for children at the margins of the target age range for the grade.

Assessment: This approach, as Barakat (2016) notes, would be unaffected (or nearly unaffected) by survey timing, and has the advantage of being readily understood and implemented, and of not requiring modeling to implement. It could, of course, also be complemented by the ASAR for the ages left out of the truncated age range, to offer insights into what is happening at the margins – with the understanding that those ages are not adjusted. This approach has much to recommend it, though it might be a reason for survey sceptics to discount use of the data alongside administrative data, arguing they are not comparable. However, these objections could be overcome, supported by a discussion of the mismatch between age data and the start of the school year and the need for some adjustment to be made.

Barakat (2016) examines the effectiveness of the modeling approaches in adjusting ages in 9 household surveys across four countries. He is able to check the consistency of results across models, and to compare results using data with only age in completed years against those with birth date data. For the two countries (Nigeria and Indonesia) with birth date data, which he uses to determine the 'exact' NAR, he finds that the unadjusted NAR underestimates the exact NAR by 4 to 6 percentage points; that the median binary approach is more effective in Indonesia (within about 1 percentage point of the exact NAR) than in Nigeria; that the individual binary adjustment matches the exact NAR in Indonesia but not in Nigeria; that the random assignment is within 1 percentage point of the exact NAR in Indonesia and within 2 percentage points in Nigeria; that the ordered assignment overestimates NAR in both countries by 1 to nearly 4 percentage points; and that the regressions Barakat ran produced estimates indistinguishable from the exact NAR in Indonesia and within 2 percentage points in Nigeria.

It would be inadvisable to draw conclusions about the effectiveness and advisability of each approach with only a couple of examples with an estimated 'exact' NAR. However, further comparisons of these approaches would be quite useful, especially where both age in completed years and birth date data are available. Before doing so, however, it would be useful to consider the overall appetite for these proposed adjustments, which vary in conceptual complexity, ease of implementation, potential for variation and error, and understandability.

On the whole, it would be best to identify and use one approach for age adjustment, rather than a host of approaches, and to select a method that improves data quality sufficiently, in terms of policy implications. A fuller sense of how exact estimates can be, in the context of the noise and error already known to exist in the raw data, should guide these explorations. Once a method is agreed upon, a collective effort to revise participation estimates over time should be undertaken – in the same way that UIS has revised estimates of out-of-school populations as better or revised data become available.

This review of the history of age adjustment of household survey participation data also suggests two general recommendations on data use and the way forward.

- Because age-specific attendance rates (ASAR) are considerably more sensitive to age adjustments than are NARs by level, the education statistics community should estimate and use ASARs with great caution. In fact, it might be advisable not to use the ASAR at all, except in instances of comparing the participation rates at the margins of a level of schooling, in the situation of – for instance – using truncated NAR by level.

- 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.

References

Barakat, B. (2016). "Sorry I forgot your birthday!": Adjusting apparent school participation for survey timing when age is measured in whole years, 2016. *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.

ICF International (2013). *Tabulation Plan for DHS Final Report,* Demographic and Health Surveys Methodology.

ORC Macro (2004). *DHS EdData Education Country Profile: Uganda.* DHS EdData Country Profiles. Calverton, MD, USA: ORC Macro.

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*. http://dx.doi.org/10.7910/DVN/26712

Uganda Bureau of Statistics (UBOS) and ORC Macro (2001). *Uganda Demographic and Health Survey 2000-2001.* Calverton, Maryland, USA: UBOS and ORC Macro.

Uganda Bureau of Statistics and ORC Macro (2001). *Uganda DHS EdData Survey 2001: Education Data for Decision-making.* Calverton, Maryland U.S.A.: Uganda Bureau of Statistics and ORC Macro.

UNESCO Institute for Statistics (UIS) (2004). *Guide to the Analysis and Use of Household Survey and Census Education Data.* 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.* 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, UNESCO Institute for Statistics. http://www.uis.unesco.org/Library/Documents/tp4-en.pdf

UNICEF (2014). *MICS Instructions for Interviewers*, 20 March, 2014.

UNICEF (2014). *MICS Instructions for Supervisors and Editors*, 21 March, 2014.